

PERSPECTIVE TAKING: BUILDING A NEUROCOGNITIVE FRAMEWORK FOR INTEGRATING THE “SOCIAL” AND THE “SPATIAL”

EDITED BY : Klaus Kessler, Sarah H. Creem-Regehr and Antonia Hamilton
PUBLISHED IN : Frontiers in Human Neuroscience



frontiers

Frontiers Copyright Statement

© Copyright 2007-2015 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88919-417-9

DOI 10.3389/978-2-88919-417-9

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

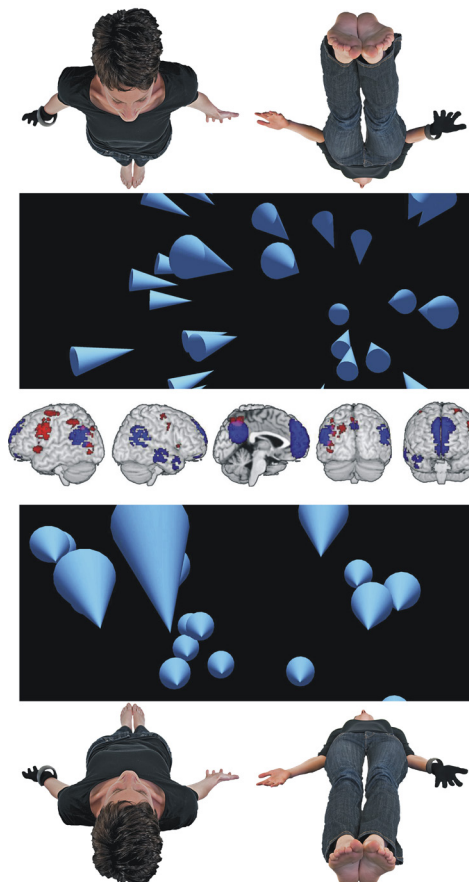
PERSPECTIVE TAKING: BUILDING A NEUROCOGNITIVE FRAMEWORK FOR INTEGRATING THE “SOCIAL” AND THE “SPATIAL”

Topic Editors:

Klaus Kessler, Aston University, United Kingdom

Sarah H. Creem-Regehr, University of Utah, USA

Antonia Hamilton, University of Nottingham, United Kingdom



The image shows a composite image adapted from three contributions in this eBook. The middle row displays images of brain activity adapted from Schurz et al. (Common brain areas engaged in false belief reasoning and visual perspective taking: a meta-analysis of functional brain imaging studies), the spatial layouts of conic shapes are adapted from Takahashi et al. (Psychological influences on distance estimation in a virtual reality environment), and the body images in various orientations are adapted from Braithwaite et al. (Fractionating the unitary notion of dissociation: disembodied but not embodied dissociative experiences are associated with exocentric perspective-taking).

Background: Interacting with other people involves spatial awareness of one's own body and the other's body and viewpoint. In the past, social cognition has focused largely on belief reasoning, which is abstracted away from spatial and bodily representations, while there is a strong tradition of work on spatial and object representation which does not consider social interactions. These two domains have flourished independently. A small but growing body of research examines how awareness of space and body relates to the ability to interpret and interact with others. This also builds on the growing awareness that many cognitive processes are embodied, which could be of relevance for the integration of the social and spatial domains: Online mental transformations of spatial representations have been shown to rely on simulated body movements and various aspects of social interaction have been related to the simulation of a conspecific's behaviour within the observer's bodily repertoire.

Both dimensions of embodied transformations or mappings seem to serve the purpose of establishing alignment between the observer and a target. In spatial cognition research the target is spatially defined as a particular viewpoint or frame of reference (FOR), yet, in social interaction research another viewpoint is occupied by another's mind, which crucially requires perspective taking in the sense of considering what another person experiences from a different viewpoint. Perspective taking has been studied in different ways within developmental psychology, cognitive psychology, psycholinguistics, neuropsychology and cognitive neuroscience over the last few decades, yet, integrative approaches for channelling all information into a unified account of perspective taking and viewpoint transformations have not been presented so far.

Aims: This Research Topic aims to bring together the social and the spatial, and to highlight findings and methods which can unify research across areas. In particular, the topic aims to advance our current theories and set the stage for future developments of the field by clarifying and linking theoretical concepts across disciplines.

Scope. The focus of this Research Topic is on the SPATIAL and the SOCIAL, and we anticipate that all submissions will touch on both aspects and will explicitly attempt to bridge conceptual gaps. Social questions could include questions of how people judge another person's viewpoint or spatial capacities, or how they imagine themselves from different points of view. Spatial questions could include consideration of different physical configurations of the body and the arrangement of different viewpoints, including mental rotation of objects or viewpoints that have social relevance. Questions could also relate to how individual differences (in personality, sex, development, culture, species etc.) influence or determine social and spatial perspective judgements. Many different methods can be used to explore perspective taking, including mental chronometry, behavioural tasks, EEG/MEG and fMRI, child development, neuropsychological patients, virtual reality and more. Bringing together results and approaches from these different domains is a key aim of this Research Topic. We welcome submissions of experimental papers, reviews and theory papers which cover these topics.

Citation: Kessler, K., Creem-Regehr, S. H., Hamilton, A., eds. (2015). Perspective Taking: Building a Neurocognitive Framework for Integrating the "Social" and the "Spatial". Lausanne: Frontiers Media. doi: 10.3389/978-2-88919-417-9

Table of Contents

- 06** *Perspective Taking: Building a Neurocognitive Framework for Integrating the “Social” and the “Spatial”*
Antonia F. de C. Hamilton, Klaus Kessler and Sarah H. Creem-Regehr
- 09** *The Use of Embodied Self-Rotation for Visual and Spatial Perspective-Taking*
Andrew Surtees, Ian Apperly and Dana Samson
- 21** *Fractionating the Unitary Notion of Dissociation: Disembodied but not Embodied Dissociative Experiences are Associated with Exocentric Perspective-Taking*
Jason J. Braithwaite, Kelly James, Hayley Dewe, Nick Medford, Chie Takahashi and Klaus Kessler
- 33** *Strategy Modulates Spatial Perspective-Taking: Evidence for Dissociable Disembodied and Embodied Routes*
Mark Gardner, Mark Brazier, Caroline J. Edmonds and Petra Gronholm
- 41** *Relating Spatial Perspective Taking to the Perception of Other’s Affordances: Providing a Foundation for Predicting the Future Behavior of Others*
Sarah H. Creem-Regehr, Kyle T. Gagnon, Michael N. Geuss and Jeanine K. Stefanucci
- 55** *Does that Look Heavy to You? Perceived Weight Judgement in Lifting Actions in Younger and Older Adults*
Corrina Maguinness, Annalisa Setti, Eugenie Roudaia and Rose Anne Kenny
- 66** *Shared Action Spaces: A Basis Function Framework for Social Re-Calibration of Sensorimotor Representations Supporting Joint Action*
Giovanni Pezzulo, Pierpaolo Iodice, Stefano Ferraina and Klaus Kessler
- 82** *Through Your Eyes: Incongruence of Gaze and Action Increases Spontaneous Perspective Taking*
Tiziano Furlanetto, Andrea Cavallo, Valeria Manera, Barbara Tversky and Cristina Becchio
- 87** *Psychological Influences on Distance Estimation in a Virtual Reality Environment*
Kohske Takahashi, Tobias Meilinger, Katsumi Watanabe and Heinrich H. Bühlhoff
- 94** *The Role of Potential Agents in Making Spatial Perspective Taking Social*
Amy M. Clements-Stephens, Katarina Vasiljevic, Alexandra J. Murray and Amy L. Shelton
- 105** *The Influence of Visual Perspective on the Somatosensory Steady-State Response During Pain Observation*
Dora L. Canizales, Julien I. A. Voisin, Pierre-Emmanuel Michon, Marc-André Roy and Philip L. Jackson
- 113** *Perspective Taking in Language: Integrating the Spatial and Action Domains*
Madeleine E. L. Beveridge and Martin J. Pickering

- 124 Reference Frame Selection in Dialogue: Priming or Preference?**
Katrin Johannsen and Jan P. De Ruiter
- 134 What Drives Successful Verbal Communication?**
Miriam De Boer, Ivan Toni and Roel M. Willems
- 142 Flexible Spatial Perspective-Taking: Conversational Partners Weigh Multiple Cues in Collaborative Tasks**
Alexia Galati and Marios N. Avraamides
- 158 How Culture Influences Perspective Taking: Differences in Correction, not Integration**
Shali Wu, Dale J. Barr, Timothy M. and Gann, Boaz Keysar
- 165 Developmental Changes in Mental Rotation Ability and Visual Perspective-Taking in Children and Adults with Williams Syndrome**
Masahiro Hirai, Yukako Muramatsu, Seiji Mizuno, Naoko Kurahashi, Hirokazu Kurahashi, and Miho Nakamura
- 178 A Review of Visual Perspective Taking in Autism Spectrum Disorder**
Amy Pearson, Danielle Ropar and Antonia F. de C. Hamilton
- 188 The Primacy of Social Over Visual Perspective-Taking**
Henrike Moll and Derya Kadipasaoglu
- 197 Old Cortex, New Contexts: Re-Purposing Spatial Perception for Social Cognition**
Carolyn Parkinson and Thalia Wheatley
- 204 Common Brain Areas Engaged in False Belief Reasoning and Visual Perspective Taking: A Meta-Analysis of Functional Brain Imaging Studies**
Matthias Schurz, Markus Aichhorn, Anna Martin and Josef Perner
- 218 The Economy of Social Resources and its Influence on Spatial Perceptions**
Elizabeth B. Gross and Dennis Proffitt
- 223 Insight into Others' Minds: Spatio-Temporal Representations by Intrinsic Frame of Reference**
Yanlong Sun and Hongbin Wang
- 234 Visual Perspective Taking and Laterality Decisions: Problems and Possible Solutions**
Mark May and Mike Wendt
- 241 Minimal Self-Models and the Free Energy Principle**
Jakub Limanowski and Felix Blankenburg



Perspective taking: building a neurocognitive framework for integrating the “social” and the “spatial”

Antonia F. de C. Hamilton^{1*}, Klaus Kessler² and Sarah H. Creem-Regehr³

¹ Institute of Cognitive Neuroscience, University College London, London, UK

² Aston Brain Centre, School of Life and Health Sciences, Aston University, Birmingham, UK

³ Department of Psychology, University of Utah, Salt Lake City, UT, USA

*Correspondence: a.hamilton@ucl.ac.uk

Edited and reviewed by:

Hauke R. Heekeren, Freie Universität Berlin, Germany

Keywords: social cognition, spatial cognition, social neuroscience, perspective taking, self-other distinction

From carrying a table to pointing at the moon, interacting with other people involves spatial awareness of one's own body and the other's body and viewpoint. In the past, social cognition has often focused on tasks like belief reasoning, which is abstracted away from spatial and bodily representations. There is also a strong tradition of work on spatial and object representation which does not consider social interactions. The 24 papers in this research topic represent the growing body of work which links the spatial and the social. The diversity of methods and approaches used here reveal that this is a vibrant and growing research area which can tell us more than the study of either topic in isolation.

Online mental transformations of spatial representations are often believed to rely on action simulation and other “embodied” processing and three papers in the current research topic provide new evidence for this process. Surtees and colleagues reveal that embodied egocentric transformations are used for visual as well as for spatial perspective taking, extending the generality of the embodied processing principle (Surtees et al., 2013). Braithwaite et al.'s contribution distinguishes between embodied and disembodied body-related hallucinations, showing that only the latter speeds up perspective taking (Braithwaite et al., 2013). Gardner and colleagues also highlight distinct processing routes towards perspective taking outcomes, where some individuals use embodied- while others use abstract (unembodied) calculation strategies (Gardner et al., 2013).

Several of the papers in this research topic have a focus on action systems in perspective taking. Creem-Regehr et al. analyze the literature on human judgments of other's affordances and how this relates to spatial perspective taking, concluding that these are complementary processes that work to inform understanding of another's behavior (Creem-Regehr et al., 2013). Maguinness et al. look at how observing another's action of lifting influences the discrimination of the weight of the objects lifted, and how this is modulated by age (Maguinness et al., 2013). Pezzulo et al. propose that that sensorimotor representations are recalibrated in social contexts to create shared action spaces serving joint action or more generally, social interaction (Pezzulo et al., 2013). Furlanetto et al. present a study examining the role of both gaze and action on perspective taking, finding the intriguing result that when gaze and action intention conflict, spontaneous perspective taking is increased (Furlanetto et al., 2013). Together, these papers suggest that perception, action

and spatial processing all interact with and contribute to social cognition.

Direct interactions between spatial factors and social factors can be seen in a variety of domains, including emotional stimuli such as threat and pain. Takahashi et al. use virtual reality to show that potentially threatening objects are perceived as closer to the participant (Takahashi et al., 2013). Clements-Stephens et al. investigate the influence of the presence of an agent and the role of social skills on spatial perspective taking, finding a complex relationship among tasks, targets, and context (Clements-Stephens et al., 2013). Finally, the impact of perspective taking on observation of other's pain is examined by Canizales et al., finding both subjective evaluation and neural somatosensory responses are modulated by the perspective taken (Canizales et al., 2013).

The relevance of social and visuospatial perspective taking for successful communication is emphasized in five contributions in this research topic. Focusing on the integration of action- and spatial- perspective taking, Beveridge and Pickering propose that alignment of spatial perspectives may serve as a prerequisite for action language simulations (Beveridge and Pickering, 2013), in which language users adopt a particular action-perspective or frame-of-reference (FOR). Johannsen and De Ruiter show that priming of a relative FOR can dominate an a priori preference for an intrinsic FOR in communication, while communicative success is predicted by the amount to which interlocutors adapt to each other's strategies—whatever these are (Johannsen and Ruiter, 2013). De Boer and colleagues approach the question of communicative success from the angle of individual traits and report that motivational as well general-purpose cognitive abilities play a crucial role (De Boer et al., 2013). The flexibility of perspective taking in communication is further highlighted by Galati and Avraamides who show that people weigh multiple cues (including social ones) to consider the relative difficulty of perspective-taking for each partner, and adapt behavior to minimize collective effort (Galati and Avraamides, 2013). In this context cultural background could make a difference. Wu and colleagues report that Westerners and East-Asians differ in their strategies of controlling ego- vs. other-centred perspective taking outcomes but are similar in their immediate (egocentric) integration of communication context (Wu et al., 2013).

Developmental and neuroscientific approaches are also important in understanding perspective taking. New data from Hirai and colleagues shows that people with Williams syndrome find it hard to perform a level 2 visual perspective taking (VPT2) task, and this may be due to difficulties in spatial processing of body postures (Hirai et al., 2013). These data complement the review from Pearson et al. which shows that children with autism also find these VPT2 tasks hard (Pearson et al., 2013). Though Williams syndrome and autism are sometimes considered to have opposite effects on social cognition, here the intersection of spatial and social processing seems to be difficult for both populations. Moll et al. argue against the traditional view that VPT is simpler than cognitive perspective taking (theory of mind) and suggest that social coordination and communication occurs developmentally prior to full VPT abilities (Moll and Kadipasaoglu, 2013). This view contrasts with the paper from Wheatley and colleagues which suggests that in human evolution, brain systems for spatial processing have been repurposed for social cognition (Parkinson and Wheatley, 2013). Finally, Schurz and colleagues report a meta-analysis of fMRI data showing that perspective taking and theory of mind engage overlapping brain regions (Schurz et al., 2013). Together, these studies show clear links between spatial and social processing, and the question of which is “primary” may become an important debate in the future.

Finally, advances in our experimental data need to be interpreted in a solid theoretical framework. Several rival theories are available. Gross and Proffitt make the claim that social connections can modulate participant’s perception of space (Gross and Proffitt, 2013). Sun and Wang consider how both spatial and social problems can be conceptualized in terms of different frames of reference, and can be broken down to similar low-level components (Sun and Wang, 2014). May and Wendt evaluate theoretical accounts of perspective taking with a focus on two different tasks that require laterality judgments (May and Wendt, 2013). Limanowski and Blankenburg take a very different approach, providing an account of the experience of “self” in terms of the free energy principle that a brain functions to minimize surprise (Limanowski and Blankenburg, 2013).

Overall, the variety of papers in this research topic reflect the diversity and dynamism of the field. Recognition of the importance of studying spatial and social information processing in the same framework has come from many angles. Future studies can examine how these different types of task can scaffold each other and interact, possibly in an embodied fashion, to enable humans to cooperate and engage in a social space.

REFERENCES

- Beveridge, M. E. L., and Pickering, M. J. (2013). Perspective taking in language: integrating the spatial and action domains. *Front. Hum. Neurosci.* 7:577. doi: 10.3389/fnhum.2013.00577
- Braithwaite, J. J., James, K., Dewe, H., Medford, N., Takahashi, C., and Kessler, K. (2013). Fractionating the unitary notion of dissociation: disembodied but not embodied dissociative experiences are associated with exocentric perspective-taking. *Front. Hum. Neurosci.* 7:719. doi: 10.3389/fnhum.2013.00719
- Canizales, D. L., Voisin, J. I. A., Michon, P.-E., Roy, M.-A., and Jackson, P. L. (2013). The influence of visual perspective on the somatosensory steady-state response during pain observation. *Front. Hum. Neurosci.* 7:849. doi: 10.3389/fnhum.2013.00849
- Clements-Stephens, A. M., Vasiljevic, K., Murray, A. J., and Shelton, A. L. (2013). The role of potential agents in making spatial perspective taking social. *Front. Hum. Neurosci.* 7:497. doi: 10.3389/fnhum.2013.00497
- Creem-Regehr, S. H., Gagnon, K. T., Geuss, M. N., and Stefanucci, J. K. (2013). Relating spatial perspective taking to the perception of other’s affordances: providing a foundation for predicting the future behavior of others. *Front. Hum. Neurosci.* 7:596. doi: 10.3389/fnhum.2013.00596
- De Boer, M., Toni, I., and Willems, R. M. (2013). What drives successful verbal communication? *Front. Hum. Neurosci.* 7:622. doi: 10.3389/fnhum.2013.00622
- Furlanetto, T., Cavallo, A., Manera, V., Tversky, B., and Becchio, C. (2013). Through your eyes: incongruence of gaze and action increases spontaneous perspective taking. *Front. Hum. Neurosci.* 7:455. doi: 10.3389/fnhum.2013.00455
- Galati, A., and Avraamides, M. N. (2013). Flexible spatial perspective-taking: conversational partners weigh multiple cues in collaborative tasks. *Front. Hum. Neurosci.* 7:618. doi: 10.3389/fnhum.2013.00618
- Gardner, M. R., Brazier, M., Edmonds, C. J., and Gronholm, P. C. (2013). Strategy modulates spatial perspective-taking: evidence for dissociable disembodied and embodied routes. *Front. Hum. Neurosci.* 7:457. doi: 10.3389/fnhum.2013.00457
- Gross, E. B., and Proffitt, D. (2013). The economy of social resources and its influence on spatial perceptions. *Front. Hum. Neurosci.* 7:772. doi: 10.3389/fnhum.2013.00772
- Hirai, M., Muramatsu, Y., Mizuno, S., Kurahashi, N., Kurahashi, H., and Nakamura, M. (2013). Developmental changes in mental rotation ability and visual perspective-taking in children and adults with Williams syndrome. *Front. Hum. Neurosci.* 7:856. doi: 10.3389/fnhum.2013.00856
- Johannesen, K., and Ruiter, J. P. De. (2013). Reference frame selection in dialog: priming or preference? *Front. Hum. Neurosci.* 7:667. doi: 10.3389/fnhum.2013.00667
- Limanowski, J., and Blankenburg, F. (2013). Minimal self-models and the free energy principle. *Front. Hum. Neurosci.* 7:547. doi: 10.3389/fnhum.2013.00547
- Maguinness, C., Setti, A., Roudaia, E., and Kenny, R. A. (2013). Does that look heavy to you? Perceived weight judgment in lifting actions in younger and older adults. *Front. Hum. Neurosci.* 7:795. doi: 10.3389/fnhum.2013.00795
- May, M., and Wendt, M. (2013). Visual perspective taking and laterality decisions: Problems and possible solutions. *Front. Hum. Neurosci.* 7:549. doi: 10.3389/fnhum.2013.00549
- Moll, H., and Kadipasaoglu, D. (2013). The primacy of social over visual perspective-taking. *Front. Hum. Neurosci.* 7:558. doi: 10.3389/fnhum.2013.00558
- Parkinson, C., and Wheatley, T. (2013). Old cortex, new contexts: re-purposing spatial perception for social cognition. *Front. Hum. Neurosci.* 7:645. doi: 10.3389/fnhum.2013.00645
- Pearson, A., Ropar, D., and Hamilton, A. F. D. C. (2013). A review of visual perspective taking in autism spectrum disorder. *Front. Hum. Neurosci.* 7:652. doi: 10.3389/fnhum.2013.00652
- Pezzulo, G., Iodice, P., Ferraina, S., and Kessler, K. (2013). Shared action spaces: a basis function framework for social re-calibration of sensorimotor representations supporting joint action. *Front. Hum. Neurosci.* 7:800. doi: 10.3389/fnhum.2013.00800
- Schurz, M., Aichhorn, M., Martin, A., and Perner, J. (2013). Common brain areas engaged in false belief reasoning and visual perspective taking: a meta-analysis of functional brain imaging studies. *Front. Hum. Neurosci.* 7:712. doi: 10.3389/fnhum.2013.00712
- Sun, Y., and Wang, H. (2014). Insight into others’ minds: spatio-temporal representations by intrinsic frame of reference. *Front. Hum. Neurosci.* 8:58. doi: 10.3389/fnhum.2014.00058
- Surtees, A., Apperly, I., and Samson, D. (2013). The use of embodied self-rotation for visual and spatial perspective-taking. *Front. Hum. Neurosci.* 7:698. doi: 10.3389/fnhum.2013.00698
- Takahashi, K., Meilinger, T., Watanabe, K., and Bühlhoff, H. H. (2013). Psychological influences on distance estimation in a virtual reality

environment. *Front. Hum. Neurosci.* 7:580. doi: 10.3389/fnhum.2013.00580

Wu, S., Barr, D. J., Gann, T. M., and Keysar, B. (2013). How culture influences perspective taking: differences in correction, not integration. *Front. Hum. Neurosci.* 7:822. doi: 10.3389/fnhum.2013.00822

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 26 March 2014; accepted: 20 May 2014; published online: 11 June 2014.

Citation: Hamilton AFC, Kessler K and Creem-Regehr SH (2014) Perspective taking: building a neurocognitive framework for integrating the “social” and the “spatial”. *Front. Hum. Neurosci.* 8:403. doi: 10.3389/fnhum.2014.00403

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2014 Hamilton, Kessler and Creem-Regehr. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The use of embodied self-rotation for visual and spatial perspective-taking

Andrew Surtees^{1,2*}, Ian Apperly² and Dana Samson¹

¹ Faculté de Psychologie et des Sciences de l'Éducation, Institut de Recherche en Sciences Psychologiques, University Catholique de Louvain, Louvain-la-Neuve, Belgium

² School of Psychology, University of Birmingham, Birmingham, UK

Edited by:

Klaus Kessler, University of Glasgow, UK

Reviewed by:

Sarah H. Creem-Regehr, University of Utah, USA

Jeanine Stefanucci, University of Utah, USA

*Correspondence:

Andrew Surtees, School of Psychology, University of Birmingham, Birmingham B15 2TT, UK
e-mail: andrew.surtees@gmail.com

Previous research has shown that calculating if something is to someone's left or right involves a simulative process recruiting representations of our own body in imagining ourselves in the position of the other person (Kessler and Rutherford, 2010). We compared left and right judgements from another's spatial position (spatial perspective judgements) to judgements of how a numeral appeared from another's point of view (visual perspective judgements). Experiment 1 confirmed that these visual and spatial perspective judgements involved a process of rotation as they became more difficult with angular disparity between the self and other. There was evidence of some difference between the two, but both showed a linear pattern. Experiment 2 went a step further in showing that these judgements used embodied self rotations, as their difficulty was also dependent on the current position of the self within the world. This effect was significantly stronger in spatial perspective-taking, but was present in both cases. We conclude that embodied self-rotations, through which we actively imagine ourselves assuming someone else's position in the world can subserve not only reasoning about where objects are in relation to someone else but *also* how the objects in their environment appear to them.

Keywords: visual perspective-taking, spatial perspective-taking, embodied self rotation, theory of mind, level-2 perspective-taking, perspective-taking

INTRODUCTION

Human beings operate in complex social and spatial environments. In order to be successful, we must navigate our way around this complex world, in which other people are particularly important. Cooperation and competition are thought to have played a vital role in our evolution (Tomasello, 2008). In order to cooperate with and compete against others we often need to represent their perspectives. A minimal definition of a perspective is that it is someone's relationship with objects and/or other people in their environment (Surtees et al., 2013). A perspective can be related to the visual experiences of an individual; famously in developmental psychology, Piaget and Inhelder (1956) asked children to judge how the experimenter *saw* an array of three mountains. Equally, a perspective can be related to the spatial location of an object; work on frames of reference has focused on people's sensitivity to whether an object is located above or below, or to the left or the right of someone (Carlson-Radvansky and Jiang, 1998; Levinson, 1996). It is clear that a mature system for visual perspective-taking at times necessitates processing beyond the spatial relations between a person and the objects within their environment. Take for example a woman who hands her elderly husband his glasses to examine a passage in a book that, while it looks perfectly clear to her, she knows will appear blurry to him. In contrast to these special cases, however, there are a multitude of everyday social situations where rapid decision-making about approximations to other people's visual experiences can be made simply on the basis of spatial relations and orientations. In this paper, we build on recent

work comparing visual and spatial perspective-taking judgements (Kessler and Thomson, 2010; Kessler and Rutherford, 2010; Michelon and Zacks, 2006; Surtees et al., 2013) and examine the role for embodiment and rotation in visual and spatial perspective judgements.

VISUAL PERSPECTIVE-TAKING

Since Piaget's early description of children as egocentric (Piaget and Inhelder, 1956), a lot of focus has been placed on the age at which children first begin to understand that others do not share their own view (Flavell et al., 1981) or a good view (Light and Nix, 1983) of the world. Such judgements are thought to require children to have a Theory of Mind (Hamilton et al., 2009), that is to understand that other people are independent actors and that their behavior is dependent upon their own mental states (Premack and Woodruff, 1978) as well as the particular, current, state of the world. Whilst much of the focus in the literature on Theory of Mind has been on children's ability to reason about beliefs in general, and False Beliefs in particular (Wimmer and Perner, 1983), successful reasoning based on the visual perceptions of others similarly requires us to be sensitive to their mental states and to overcome our own, egocentric biases (Surtees and Apperly, 2012).

Research with children and non-human animals suggests that perspective-taking is not a unitary ability (Masangkay et al., 1974; Flavell, 2000; Call and Tomasello, 2008). Flavell and colleagues (Masangkay et al., 1974; Flavell et al., 1981) make a distinction between level-1 and level-2 perspective-taking. Level-1

perspective-taking requires understanding of *what* can be seen, simply knowing which objects in the world are visually accessible to another person. Masangkay et al. (1974) showed children as young as 3 to be able to successfully report that an adult could see a dog pictured on the reverse of a card when they themselves saw a cat on its obverse. Children of this age were, however, unable to report that a picture of a turtle on a flat-lying card would look upside down to the adult when it looked the right way up to them. This latter task reflects level-2 perspective-taking, judging *how* someone sees the world, specifically judging that a single object can be represented differently by two different people based on their viewpoint in the world. The emergence of level-2 perspective-taking has been associated with other Theory of Mind developments that also occur around the age of four (Perner, 1991), such as False Belief reasoning (Wimmer and Perner, 1983), reasoning about the difference between appearance and reality (Flavell et al., 1983). Similarly, a number of other cognitive abilities significantly progress at this age, such as counterfactual thinking (Riggs et al., 1998), early reasoning about regret (Weisberg and Beck, 2010) and also executive functioning (Espy, 1997; Kirkham et al., 2003). The distinction between level-1 and level-2 perspective-taking appears not to be merely linked to children's development, with many non-human animals, such as chimpanzees (Tomasello et al., 2003), goats (Kaminski et al., 2001), dogs (Hare and Tomasello, 2005) and Western Scrub-Jays (Emery and Clayton, 2004) showing level-1, but as yet no evidence of level-2 abilities. Similarly, infants (Song and Baillargeon, 2008) and adults (Samson et al., 2010) seem to be spontaneously sensitive to whether or not someone sees a given object, but again there is no such evidence for level-2 perspective-taking (Surtees et al., 2012a). It is this convergence of evidence that has led Apperly and Butterfill (2009) to suggest that the distinction between level-1 and level-2 perspective-taking may demarcate a signature limit on efficient theory of mind, such that level-2 judgements are always demanding of cognitive resources. In the current paper, we examine level-2 type judgements. Specifically judgements of how a numeral looks to someone else.

SPATIAL PERSPECTIVE-TAKING

For Spatial perspective-taking here we mean the ability to understand the spatial relationship between an individual and the objects in their environment (sometimes spatial perspective-taking is used to refer to mentally occupying another's position in space; Kessler and Wang, 2012). Unlike for visual perspectives, the content of Spatial perspectives is non-mental. A Spatial perspective is solely and definitively prescribed by the exact spatial relationship between a person and objects around them, rather than what they think about those objects. Whilst we may use our understanding of how others perceive the world around them to inform our judgements of where items are located in space relative to them, it is not necessary to do so. A book may remain to the front and the left of someone, regardless of whether they have perfect vision, suffer from short-sightedness or are blind. Similarly, if someone were to close their eyes, we should understand that they no longer have a visual perspective on the world, but maintain their spatial perspective. For this reason, spatial perspectives are not necessarily linked

to individual people (Surtees et al., 2012b). A book can be located to the front and left of a chair in the very way in which it can be located to the front and the left of a person. Consequently, spatial perspectives have been most commonly considered in terms of frames of reference. A frame of reference is a set of axes upon which to consider the location of objects (Levinson, 1996, 2003). These axes can be absolute, defined by an unchanging element of the environment- Birmingham is located to the North of Brussels, regardless of where we are. They can be relative, defined by the position of objects in relation to the viewer- you cannot see the Manneken Pis if you stand on the Grand Place in Brussels because the Hotel de Ville is in front of it. Or they can be intrinsic, defined by one of the objects we are reasoning about- I can move the Palais Royal from being behind me to being in front of me by the simple expedient of turning myself around. It is these *intrinsic* frames of reference that incur spatial perspectives. Calculating an intrinsic reference frame requires understanding the relationship between an individual person or object and their environment. Spatial frames of reference are calculated automatically following the use of prepositions (Carlson-Radvansky and Jiang, 1998), requiring inhibition to choose the most appropriate frame. Both adults (Carlson-Radvansky and Irwin, 1993) and children (Surtees et al., 2012b) are known to be concurrently sensitive to multiple frames of reference. Like for visual perspective-taking, there is evidence that children do not necessarily use all aspects of a frame of reference at the same age (Hands, 1972; Harris and Strommen, 1972; Cox, 1981; Bialystok and Codd, 1987). They show a preference for the intrinsic frame of reference in early childhood and also learn the spatial referents "in front" and "behind" (Harris and Strommen, 1972; Cox, 1981; Bialystok and Codd, 1987), before the referents "left of" and "right of" (Hands, 1972). Interestingly, adults seem spontaneously sensitive to other people's spatial perspectives. Tversky and Hard (2009) found that adults described objects as being to the left or right of a person even though the task only asked them to describe the location of an object.

PROCESSES FOR VISUAL AND SPATIAL PERSPECTIVE-TAKING

Whilst much of the focus in the visual perspective-taking literature has been on the conceptual demands of understanding other people's minds, it is clearly of importance to understand the cognitive architecture that allows us to represent others' point of view (Kessler et al., under review). Such processing must take into account complex relationships between individuals and objects within their spatial environment. Recently, a number of studies have looked to identify the different processes for visuo-spatial perspective-taking. Michelon and Zacks (2006) proposed 2 kinds of visuo-spatial perspective-taking processes. The first of these, equivalent to level-1 visual perspective-taking, was used when adults had to judge if an object could be seen or not. This process was sensitive only to the distance between the target other and the object about which the perspective was taken. It was concluded that this process involved tracing the line of sight of the avatar. A second process was sensitive to the angular disparity between the participant and the other person in the scene and was used when participants had to judge if a specified object was to the left or right from the avatar's position. This second process was concluded to

require mental self rotation to align one's own perspective with that of another. Whilst the exact question was if the other saw the object as on its left or right, it is clear that this second judgment is primarily spatial in nature and equivalent to a purely spatial judgment of whether the object was to the other's left or right. Michelon and Zacks's (2006) findings are in line with previous evidence of the effect of angular disparity on spatial judgements (Huttenlocher and Presson, 1973; Kozhevnikov and Hegarty, 2001; Keehner et al., 2006) and the identification of both visuo-spatial perspective-taking processes has since been replicated by Kessler and colleagues (Kessler and Rutherford, 2010; Kessler and Thomson, 2010; Kessler and Wang, 2012). In a recent study, we (Surtees et al., 2013) looked to further delineate these processes, and in particular examined whether the differences found by Michelon and Zacks (2006), Kessler and Thomson (2010) were primarily caused by judgements being of a visual vs. spatial nature, or whether they were primarily caused by judgements being of an early developing kind or a later developing kind. We found that spatial perspective judgements of an object as being in front of, or behind, like visual perspective judgements of whether something was visible, were not dependent on the angular disparity between the self and other. The difficulty of visual judgements of how a numeral appeared, on the other hand, like spatial judgements of something as being to the left or to the right for someone, were dependent on this angular disparity. We concluded that the selection of processing strategy was not determined by the nature of the content, as mental or non-mental, but rather by the specific task requirements and the degree to which simple features could be used. A rotational mechanism seemed to be the default method for only two kinds of judgements; level-2 visual perspective judgements of *how* something appeared to someone else and spatial judgements on the left to right dimension of an intrinsic frame of reference.

EMBODIED SELF ROTATION VS. VIEWPOINT ROTATION

Difficulty based on angular disparity could be indicative of three different types of rotational strategies. The first is a mental self rotation, which uses an embodied representation of the self that is then rotated to the current bodily position of the target perspective (Kessler et al., under review). Such a process uses motor representations to imagine transporting ourselves to another's position (Kessler and Thomson, 2010) and then simulates a self perspective from that new position. The second is a mental object rotation, through which we rotate the world from the angle of the target perspective to our own current position (Kessler et al., under review). Finally, the third is a mental viewpoint rotation (Kessler et al., under review), through which we use visuo-spatial cues to calculate a viewpoint in a given position without occupying that point of view in an embodied way. Only the first of the three strategies would require embodied self representations. Kessler and Thomson (2010) and Kessler and Rutherford (2010) used an innovative method to investigate whether mental self rotation was used for left and right judgements. Varying the angle of participants' own bodies in relation to the screen, whilst keeping head position fixed, they reasoned, would only affect performance if mental self rotation was employed. They found (Kessler and Rutherford, 2010) that even though the visual impression remained the same (because

the head position was fixed) across conditions, participants' performance varied as a function of their own body angle, with better performance when their own body posture more closely matched that of the avatar. They concluded from this that judging if an object was to the left or the right of someone else involves an embodied process of self rotation to align our perspective with theirs. They found no impact of their body rotation manipulation on judgements of whether an object was visible to the avatar or not. This is perhaps not surprising as these judgements are not affected by angle at all.

THE CURRENT STUDY

The aim of the current study was to test whether embodiment is also used in visual perspective-taking. In the current study, we adapted Kessler and colleagues' (Kessler and Rutherford, 2010; Kessler and Thomson, 2010) body posture manipulation to compare its effect on two kinds of perspective-taking task. When participants judged if an object was to an avatar's left or right (spatial perspective judgement), we predicted that they would use embodied self rotation. We expected that their performance would be affected not only by the angle of disparity between the avatar's position and participants' own position, but also by participants' own body posture- with better performance when their own body posture was more similar to that of the avatar (as found by Kessler and Rutherford, 2010; Kessler and Thomson, 2010). When participants judged how a number looked to the avatar (visual perspective judgement), we predicted that judgements would again be affected by angular disparity (as found in Surtees et al., 2013). Our central research question was whether this process was embodied or not. If these judgements were also affected by body posture, it would indicate a common embodied self rotation process implicated in both visual and spatial perspective-taking. If these judgements were independent of body posture consistency, this would suggest that these judgements involved non-embodied viewpoint rotation.

EXPERIMENT 1

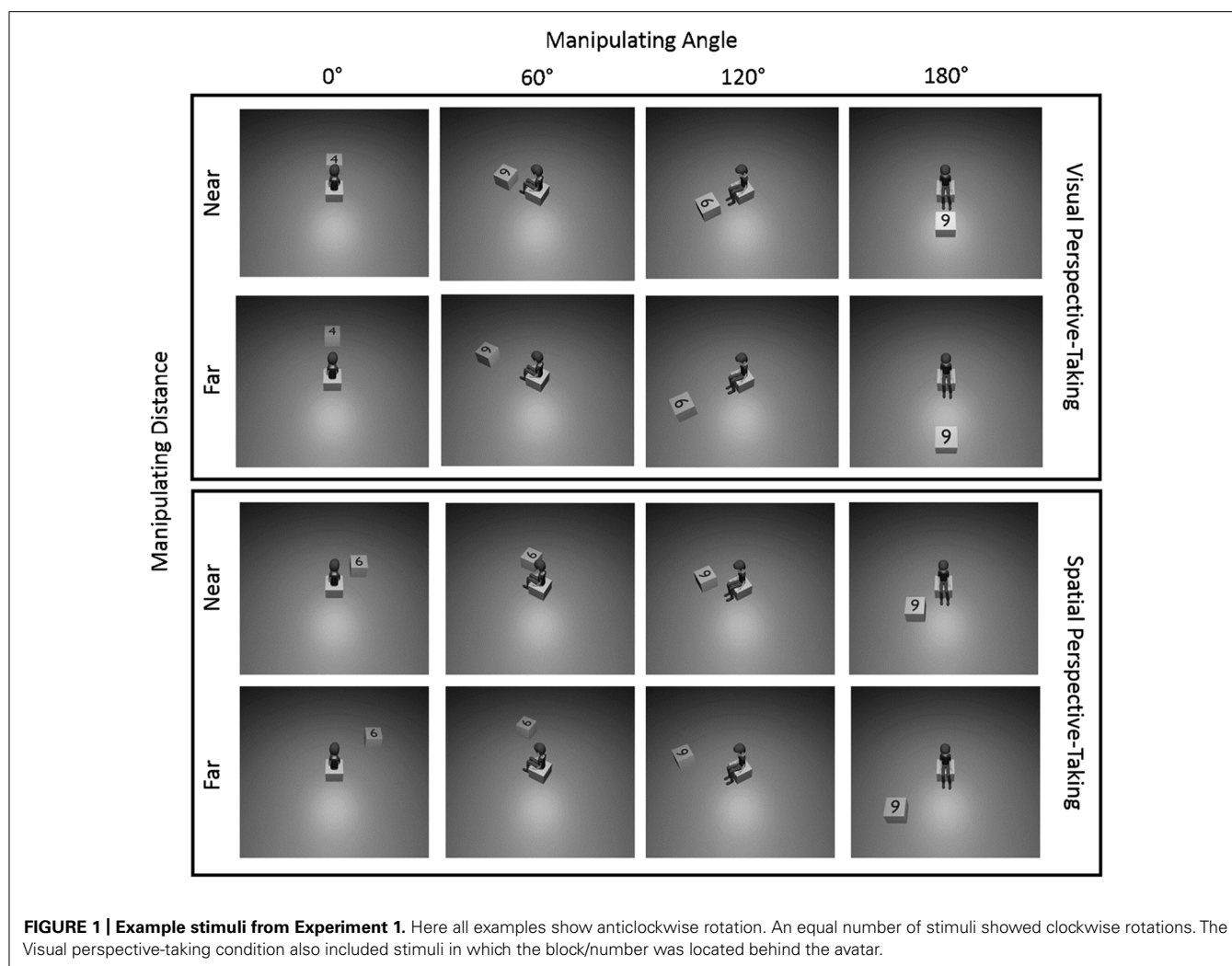
MATERIALS AND METHODS

Participants

Participants were 40 undergraduate students (11 male) from the University catholique de Louvain, Belgium. They all participated in the study in exchange of course credit or a small honorarium of 8 Euros. Participants had an average age of 20.77 years (range 18–25). One participant was not included in the final sample on the basis of performing below chance.

Stimuli

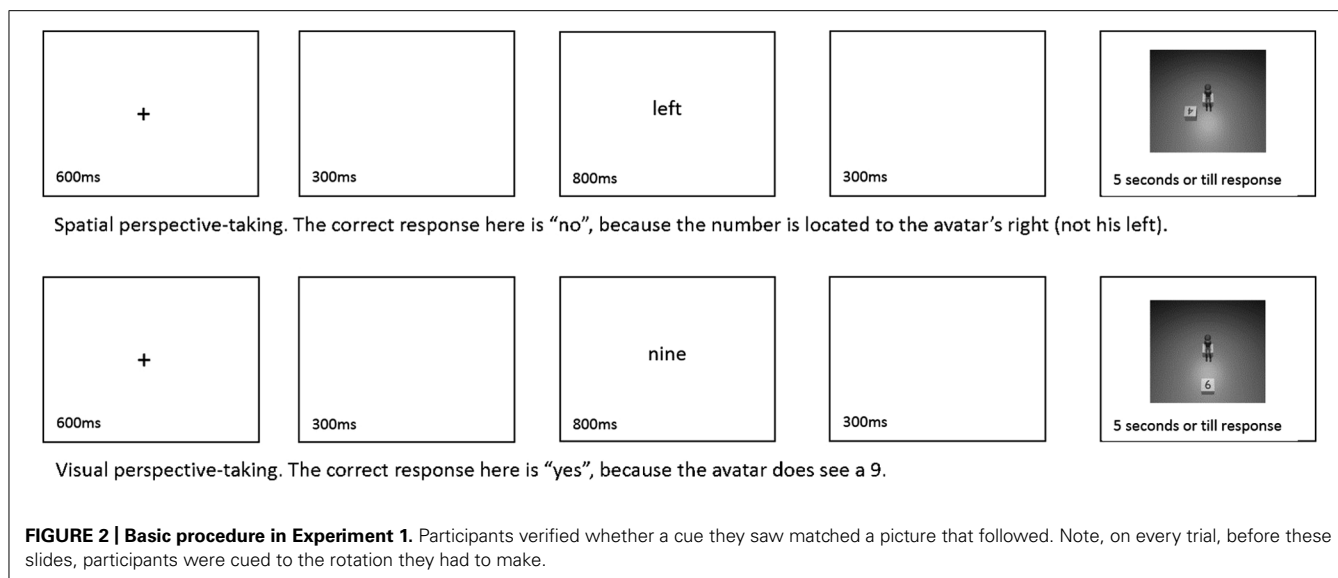
In all of the pictures that participants saw, an avatar was placed in the center of a featureless room (see **Figure 1**). The stimuli were created using Blender (www.blender.org). The room also contained a single cube, with a numeral written on its top-most face (4, 6, 7 or 9). Within each stimulus, we varied two features orthogonally. Angular disparity between the participant and the avatar was varied through the positioning of the virtual camera in relation to the avatar, creating angles of four different magnitudes: 0°, 60°, 120°, 180°. For angular disparity of both 60° and 120°, separate stimuli were created showing the avatar in clockwise



and anticlockwise variants, crucial for evaluating our embodiment hypothesis. We also varied distance, by placing the block at one of two distances from the avatar: “Near” and “Far,” where the Far condition was placed at a distance that was twice as far within the virtual world as the Near condition. In the spatial condition, an equal number of stimuli placed the block/number to the left and to the right of the avatar, at an angle of 45° from the avatar and always in front of him. In the visual condition, stimuli showed the block/number to be directly in front, or directly behind the avatar.

Participants were randomly divided into two groups, Visual or Spatial. Before the experiment, all participants were given the same basic information, that they would be performing a perspective-taking task. Participants were sat in a rotating chair, with a red rectangle attached to the floor at approximately 60° angle to their right and a blue rectangle at approximately 60° to their left. They placed their chin on a chin-rest (located 50 cm from the screen) on every trial. After further instruction, giving example procedures and the correct answer, all participants completed 16 practice trials without rotation, 8 practice trials with rotation and finally 256 experimental trials divided into four blocks. All trials followed the same basic procedure (see **Figure 2**). Participants were

first of all cued with a picture showing a red or blue square with a schematic illustration of a person (adapted from Kessler and Thomson, 2010). Participants had been instructed that the red picture meant they should rotate their body to the left/anticlockwise and place their feet on the red rectangle on the floor, they were instructed to keep the mouse on their lap (see **Figure 3**). The blue picture conveyed the same instruction, but to the right/clockwise. These rotations meant that participants’ own body orientation varied from approximately 60° clockwise to approximately 60° anticlockwise in relation to the screen for every trial. Importantly, though, by keeping their chins on the chin rest, participants’ visual impression did not change (beyond the variations in the stimuli type presented on the screen). Following the rotation cue, participants saw a further screen, asking whether they had made the rotation, this required a mouse click to progress. The experimenter observed a sample of these rotations and saw no cases in which participants made errors in their rotations (this included at least 20 consecutive trials for each participant). Following this stimulus, the standard trial sequence (Surtees et al., 2013) was presented (**Figure 2**). A fixation cross was followed by a cue (for spatial, left or right; for visual, four, six, seven, or nine). This cue was



followed by the picture itself. In response to the picture, participants pressed the left mouse key to indicate that it matched the cue and the right mouse key to indicate that it did not. Participants received feedback during practice, but not during the experiment itself.

For the Spatial condition, trials were equally and orthogonally divided on four experimental factors. There were an equal number of trials in which the cue did and did not match the picture (Match/Mismatch). An equal number of trials of each of the Angles (0° , 60° , 120° , 180°), of which the angles 60° and 120° were equally often clockwise or anticlockwise rotations. An equal number of stimuli showed each Distance (Near, Far). Finally, an equal number of trials required a left (red) or a right (blue) rotation. Note, that our variable of particular interest- the consistency between avatar and participant rotation- was varied through a combination of Angle and Rotation. Consistent trials occur with a left rotation of the participant and an anticlockwise

rotation of the avatar *or* with a right rotation of the participant and a clockwise rotation of the avatar. This means that the factor Consistency was independent of stimulus and independent of participant rotation. Stimuli were also varied on whether the cue/object was left or right and whether the number was a 4, 6, 7 or 9.

For the Visual condition, trials were again equally divided between Match and Mismatch trials. For Match trials, stimuli were varied exactly as above, save for the fact that the block/number was always directly in front of the avatar. For half of mismatch trials, “number mismatch trials,” the same stimuli were presented, with the block/number directly in front of the avatar, but the preceding cue being a different number to that seen by the avatar in the picture. In the other half, “location mismatch trials,” the cue would have been correct had the avatar been looking at the number, however, it was placed directly behind him. This manipulation meant that participants had to take into account the avatar’s view and

could not use the rotation of the number alone as a cue to the correct answer.

In summary, the Visual and the Spatial conditions for Match trials (those to be analyzed) were identical other than features necessary for the specific judgements. Both conditions included a range of angles from 0° to 180°, to confirm that rotation was being used for the task at hand. By manipulating the position of the participant in relation to the screen, and the positioning of the avatar on the screen, we varied consistency between body postures. This was independent of Angle, Distance, Rotation, Task Content (Visual, Spatial), Number, Cue and Direction, so that any influence could only be the result of the congruency between the embodied state of the participant and the avatar (see **Figure 3**).

RESULTS

Only Match trials- those in which the cue matched the picture- were included in the final analysis. Outliers were excluded from the analysis of response times on the basis of being more than 2.5 standard deviations away from the mean response time (2.9% for visual, 2.7% for spatial), as were incorrect responses.

Our first analyses investigated the effects of Angle and Distance on perspective-taking. Particularly important here are the effects of Angle and any interaction between Angle and Content. A linear effect of Angle would be representative of participants using some form of rotation to complete the task. This analysis does not investigate the embodied nature of the process.

A $4 \times 2 \times 2$ ANOVA with Response Time as a dependent variable, Angle (0°, 60°, 120°, 180°) and Distance (Near, Far) as within subjects factors and content (Visual, Spatial) as a between subjects factor revealed a main effect of distance, $F(1, 38) = 10.46$, $p = 0.003$, $\eta^2 = 0.216$, with shorter avatar-object distances processed more quickly¹. There was also a main effect of Angle, $F(3, 114) = 37.71$, $p < 0.001$, $\eta^2 = 0.498$, which represented a linear trend. There was also a main effect of content, with Visual judgements being responded to more quickly, $F(1, 38) = 12.89$, $p = 0.001$, $\eta^2 = 0.253$.

An interaction between Angle and Content, $F(3, 114) = 7.43$, $p = 0.004$, $\eta^2 = 0.163$, revealed a different relationship with Angle for each Content. For both Visual, $F(1, 19) = 41.36$, $p < 0.001$, and Spatial, $F(1, 19) = 31.83$, $p < 0.001$, perspective-taking the relationship with Angle fitted a linear trend. We investigated this relationship further, by computing separate *t*-tests for adjacent angles for each content. For Spatial perspective-taking, the strongest effect was for participants being slower at 120° than 60°, $t(19) = 5.67$, $p < 0.001$, with a less strong, but still significant effect of 180° being slower still $t(19) = 5.67$, $p = 0.003$. Though responses at 0° were the slowest, these were not significantly slower than at 60°, $t(19) = 1.361$, $p = 0.190$. Visual perspective judgements showed a different pattern of performance (see **Figure 4**). Here difference was greatest for judgements at 180° being slower than at 120°, $t(19) = 4.75$, $p < 0.001$. There was a trend for an effect of faster judgements at 60° than 0°, $t(19) = 1.934$, $p = 0.068$ and no significant effect between 60° and 120°, $t(19) = 1.341$, $p = 0.196$, though again the larger angle produced a numerically

longer response time. The interaction between Distance and Content, $F(1, 38) = 7.00$, $p = 0.012$, $\eta^2 = 0.156$, illustrated that there was an effect of Distance on Visual, $F(1, 19) = 20.85$, $p < 0.001$, $\eta^2 = 0.543$, but not Spatial, $F(1, 19) = 1.47$, $p = 0.705$, $\eta^2 = 0.008$, judgements.

Error rates across conditions were generally low and did not contradict the findings from response time (see **Table 1**).

Trials in which angular disparity was either 60° or 120° could be either Consistent or Inconsistent on the basis of whether participants have rotated their body to the left or to the right. Analysing this subset of trials with Consistency as an additional factor can test the role of embodiment. A $2 \times 2 \times 2$ ANOVA was completed with Content as a between subjects factor and Consistency (Consistent, Inconsistent) and Angle (60°, 120°) as within subjects factors. A main effect of Angle, $F(1, 38) = 25.55$, $p < 0.001$, $\eta^2 = 0.402$, was moderated by an interaction between Angle and Content, $F(1, 38) = 10.63$, $p = 0.002$, $\eta^2 = 0.219$. Over this smaller range of angles, the effect was only significant in the Spatial domain, $F(1, 19) = 33.07$, $p < 0.001$, $\eta^2 = 0.635$, not the Visual domain, $F(3, 19) = 1.69$, $p = 0.210$, $\eta^2 = 0.081$. There was no significant effect of Consistency, $F(1, 38) = 1.30$, $p = 0.261$, $\eta^2 = 0.033$, but there was a trend for an interaction between Consistency and Content, $F(1, 38) = 3.63$, $p = 0.064$, $\eta^2 = 0.087$. This illustrated a trend for Consistent trials being easier than Inconsistent, but only in the Spatial condition, $F(1, 19) = 3.041$, $p = 0.097$, $\eta^2 = 0.138$ (see **Figure 4**).

In Experiment 1, Visual perspective-taking, but not spatial perspective-taking showed an effect of distance. This was surprising and had not been evidenced in our previous study (Surtees et al., 2013), in which we found a significant effect of distance that did not differ across conditions. One possibility is that having fewer conditions here (two rather than four) has given a greater power to identify a difference. This is supported by the fact that in the spatial condition of Surtees et al. (2013), at two angles (0° and 120°), judgements at shorter distances were actually more difficult than at longer distances. This was never the case in the visual condition, where further distance always conferred greater difficulty. Both Visual and Spatial perspective-taking showed a strong and linear effect of angular disparity between the participant and the avatar on the screen in front of them, replicating the findings of Surtees et al. (2013) and suggesting a rotational process was employed. That is not to say, however, that this relationship was identical. For Spatial perspective-taking, the strongest effect was between the two mid-range angles, 60° and 120°. For Visual perspective-taking, this difference was not significant, instead it was the difference between 120° and 180° that was most strongly significant. This is in some ways surprising, as this difference was not found by Surtees et al. (2013) who used the very same stimuli. One possibility is that the physical rotations (regardless of direction) had a different effect on the rotational processes of Visual and Spatial perspective. Specifying exactly how is very speculative at this stage, but one possibility is that for Spatial perspective taking, the 60° condition was made artificially easy because here the character's basic body posture matched the participant's.

Experiment 1 also showed a trend for an interaction between Consistency and Content suggesting that visual and spatial perspective-taking may be embodied to a different degree. Spatial

¹In both experiments, all statistics are Greenhouse-Geisser corrected to guard against violations of sphericity.

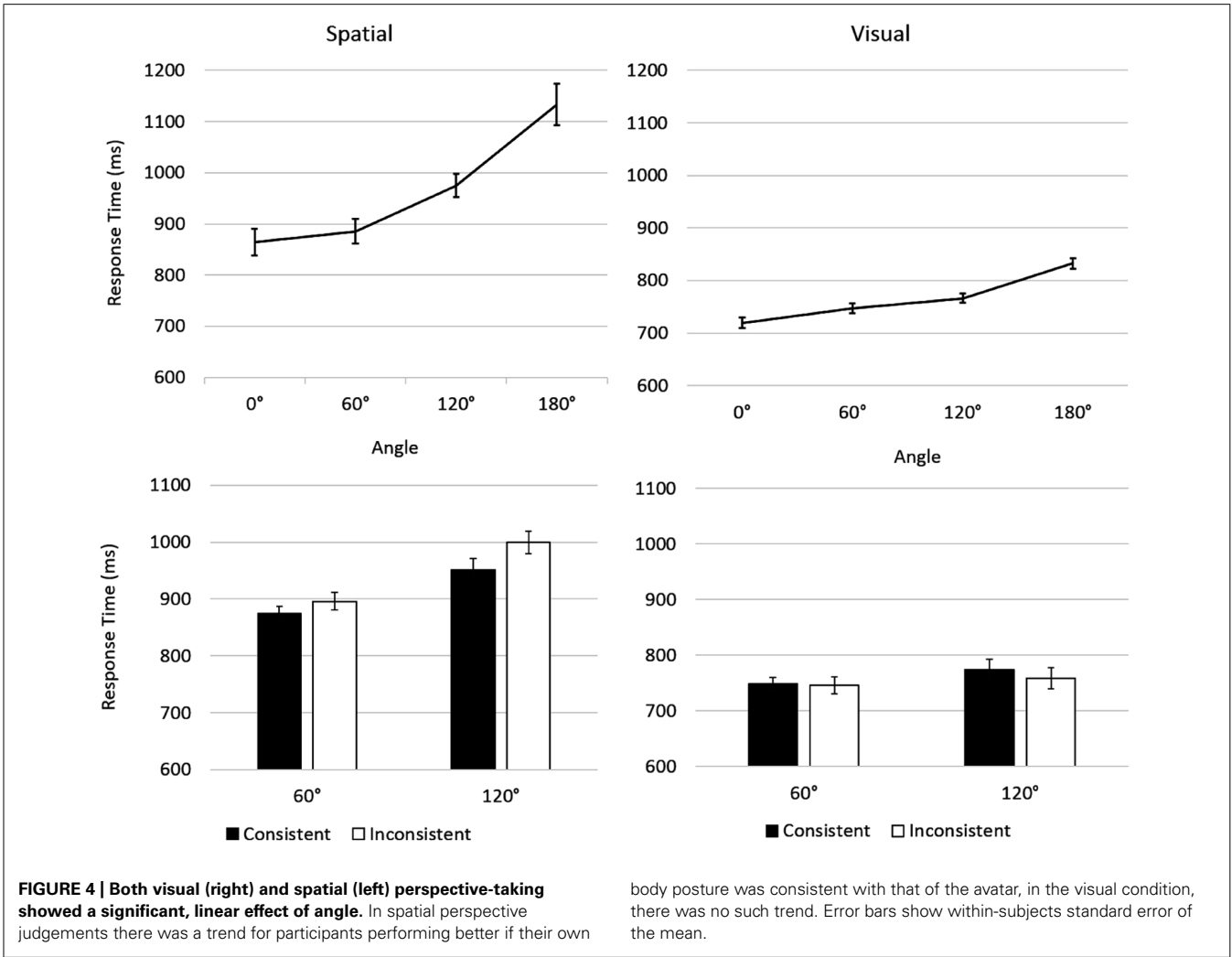
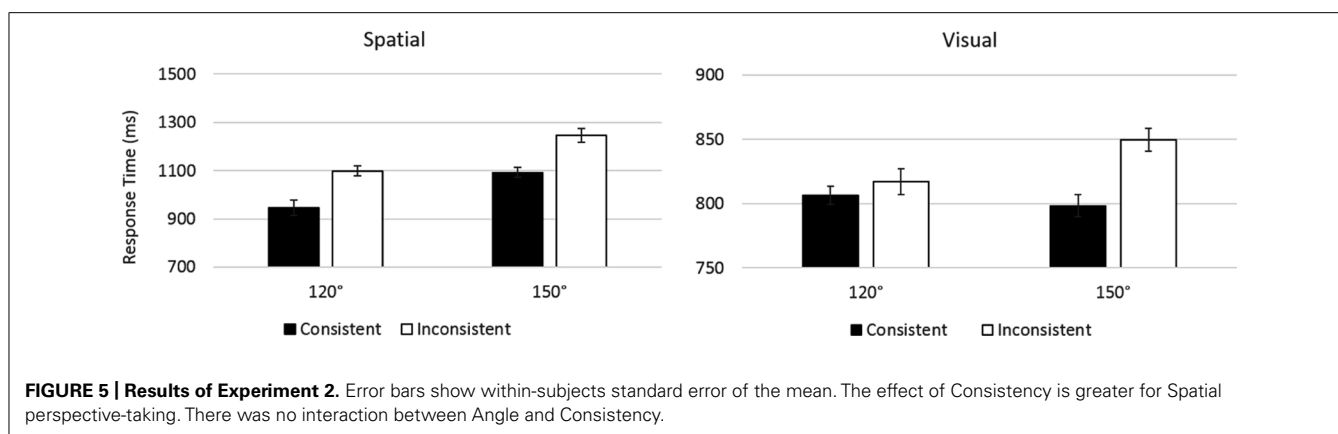


Table 1 | Percentage error rates (standard deviations) from Experiments 1 and 2.

Experiment 1					
	0°	60°	120°	180°	
Spatial	5 (5)	3.59 (4)	4.84 (5)	7.34 (7)	Consistent
Visual	1.87 (3)	3.75 (5)	1.41 (3)	3.28 (3)	Inconsistent
Experiment 2					
	120° Consistent	120° Inconsistent	150° Consistent	150° Inconsistent	
Spatial	3.65 (4)	7.03 (9)	2.60 (3)	5.99 (7)	
Visual	3.65 (5)	5.99 (5)	3.13 (3)	3.65 (6)	

perspective-taking showed a trend for an effect of Consistency. Participants trended toward performing better when their own position was aligned with that of the avatar. From this, we tentatively concluded that spatial perspective-taking recruited an embodied self rotation process, while visual perspective-taking recruited a (non-embodied) viewpoint rotation process. However, as the trends were non-significant, it also remains possible that our test was insensitive to differing embodied effects (for we should expect an effect of body posture consistency at least in the spatial condition to replicate the findings of Kessler and Rutherford, 2010; Kessler and Thomson, 2010). Also, our first experiment investigates one specific circumstance when we have to confirm a pre-defined proposition for the other’s perspective (our task required a verification, yes/no, judgement). It is possible



that actively calculating another person's perspective uses embodiment to a different degree. In Experiment 2, we addressed both the concern over lack of sensitivity and over perspective confirmation vs. calculation by using a forced choice methodology. As well as removing the verification aspect of the procedure, this method has the advantage of increasing the power (as all responses are permissible in the final analysis). Also to increase power, we removed the stimuli showing the avatar at 0° and 180° (as these only tested for rotation, not embodiment *per se*) and tested only at angles higher than 90°, those which previous studies have found to show clearest embodiment effects (Kessler and Thomson, 2010; Kessler and Rutherford, 2010).

EXPERIMENT 2

MATERIALS AND METHODS

Participants

Participants were 32 undergraduate students (9 male) from the University catholique de Louvain, Belgium. They all participated in the study in exchange for a small honorarium of 8 Euros. Participants had an average age of 21.93 years (range 18–26).

Design and procedure

The design of Experiment 2 was identical to that of Experiment 1, other than the following details. Instead of making responses to a preceding cue, here participants made a forced choice response. For Spatial perspective-taking, this meant pressing the left button on the mouse when the number was located to the left of the avatar and the right when it was to his right (note here that effects of spatial compatibility, Simon, 1969, were controlled across body posture consistency). For Visual perspective-taking, it meant pressing the left button when the number the avatar saw was a number six and the right when he saw a number nine. In this case, only stimuli where the avatar saw a six or nine were included and they were always placed in front of him (and displaced to the left or right in the spatial condition). After completing 24 practice trials (as in Experiment 1, 16 with the task alone, 8 with rotation), participants completed 96 experiment trials. Fewer trials were needed here as all trials were included in the final analysis (the analysis of Consistency has more power). New stimuli were created that had the avatar placed at either 120° or 150° angle from the participant. Again, participants were

cued before each trial to rotate to the left or right, again placing their feet on the mat at an angle of approximately 60° to the screen.

RESULTS

Again, trials in which participants made incorrect responses were excluded (see **Table 1**), as were trials in which response time was more than 2.5 standard deviations away from the mean (3.2% for visual, 3.1% for spatial).

A $2 \times 2 \times 2$ ANOVA with Content (Visual, Spatial) as a between subjects factor, and Angle (120°, 150°) and Consistency (Consistent, Inconsistent) as within subjects factors revealed an effect of Angle, $F(1, 30) = 19.88$, $p < 0.001$, $\eta^2 = 0.399$, $120^\circ < 150^\circ$ (**Figure 5**). This effect was moderated by an interaction with Content, $F(1, 30) = 14.20$, $p = 0.001$, $\eta^2 = 0.321$, showing the effect of Angle was significant for Spatial judgements, $F(1, 15) = 18.95$, $p = 0.001$, $\eta^2 = 0.558$, but not for Visual judgements, $F(1, 15) = 1.11$, $p = 0.308$, $\eta^2 = 0.069$. There was no interaction between Angle and Consistency, or Angle, Consistency and Content, $F_s < 1.10$, $p_s > 0.307$, $\eta^2 < 0.035$.

Crucially, there was an effect of Consistency, $F(1, 30) = 25.84$, $p < 0.001$, $\eta^2 = 0.463$. This was moderated by a significant interaction, with Content, $F(1, 30) = 11.42$, $p = 0.002$, $\eta^2 = 0.276$. Investigating this interaction showed that while the size of this effect was numerically greater in the spatial condition, $F(1, 15) = 18.79$, $p = 0.001$, $\eta^2 = 0.566$, it was also significant in the Visual condition, $F(1, 15) = 15.31$, $p = 0.001$, $\eta^2 = 0.505$. As in Experiment 1, there was also an effect of Content, $F(1, 30) = 9.79$, $p = 0.004$, $\eta^2 = 0.246$, such that Visual perspectives were processed more quickly.

DISCUSSION

Across two experiments, we investigated the degree to which perspective-taking required mental rotation and the degree to which that rotation was embodied. We tested this for two very different kinds of perspective-taking. We found further evidence that an explicitly spatial task recruited mental rotation. When participants judged whether an object was to the left or the right of an avatar it became increasingly more difficult as the angle of the avatar's body became increasingly more different from the participant's position. In addition to this, we found evidence that this

rotational process was an embodied self rotation, as has previously been shown by Kessler and Thomson (2010), and Kessler and Rutherford (2010). Participants found it easier (statistical trend in Experiment 1 and significant effect in Experiment 2) to make spatial judgements when their own body posture more closely matched that of the avatar- even though this was manipulated independently of the visual impression of the scene. When participants completed a visual perspective-taking task, we also found evidence of rotation. Experiment 1 showed that it is harder to judge how a number appeared to an avatar whose angular viewpoint differed from one's own to a greater degree. Perhaps most surprisingly, in Experiment 2, we showed that this process could also involve an embodied self rotation. In sum, findings for spatial perspective-taking suggested consistent use of embodied mental self rotation. For visual perspective-taking we evidenced the same process, but the strength of the effect was neither as strong nor as consistent as for spatial perspective-taking. The embodiment of this process was only evidenced in Experiment 2 and even here was not as strongly significant as for spatial perspective-taking.

VISUAL PERSPECTIVE-TAKING

Judging how a numeral looks to someone else who does not view it from the same angle as us is a clear example of level-2 visual perspective-taking, knowing that a single object can make a different visual impression on two people who view it from different angles (Flavell et al., 1981). This process is known to be difficult both for children (Masangkay et al., 1974) and for adults (Surtees et al., 2012a). Developmentalists have tended to focus on the conceptual difficulties posed by holding two conflicting relationships on a single object (Perner, 1991) or on the demands in inhibiting a salient self perspective (Surtees et al., 2012a). Here we present evidence that one source of difficulty in these tasks is rotation. Whilst it is clear that in Flavell's classic "turtle" task we have to understand that another person can represent the same turtle differently *and* inhibit a salient self-view of a turtle happily upstanding or disarmingly prostrate, we also need to mentally align how we see the world with how it is seen by the person with whom we are interacting. We replicate findings from our previous study (Surtees et al., 2013) that level-2 visual perspective judgements become more difficult as our angle becomes more different from that of the person whose perspective we take. In Experiment 1, we showed a linear effect of Angle on speed of responses.

Kessler and Rutherford (2010), Kessler and Thomson (2010) showed that judging that one object was on the left or the right from someone else's point of view was affected by the participant's current body angle in the world. Here we show that the same applies to judgements of visual perspectives. In Experiment 2, participants' own body angle affected their ability to judge if a number looked like a six or a nine to an avatar on the screen. This is the first finding showing that a judgment of a purely mental state can also require us to align our bodies with that of someone else in the world. This suggests that, at least in some cases, to think of how someone else sees the world requires us really "putting ourselves into their shoes." The effect of body posture consistency was however only significant in Experiment 2 and not

in Experiment 1. There are two possible explanations to account for these discrepant results. One possibility is that Experiment 1 simply was not sensitive enough to demonstrate this effect. A second possibility is that the difference reflects the employment of different processes determined by surface demands of the situation. Experiment 1, in which participants have to hold in mind a cue (e.g., "nine" meaning that they have to verify if the object looked like a 9 to the avatar), may promote a different strategy from Experiment 2 in which participants' judgements are solely based on the picture stimulus (here participants have to decide whether the object looks like a "6" or a "9" to the avatar when presented with the picture). In Experiment 1, participants could have used the cue to create a mental image of an expected stimulus and then used a geometrical comparison between this and the final picture. This would result in the observed effect of angle and the absence of effect of body posture consistency. In Experiment 2, on the other hand, the effect of body posture consistency seems to rule out that such geometrical comparison was used consistently across trials and participants. It is also possible that some participants used conditional rules to calculate visual perspectives (e.g., If he faces toward me then he does not see the same number as me), but our significant findings, of angular disparity in Experiment 1 and embodiment in Experiment 2 suggest this was not widely applied. We propose that level-2 visual perspective-taking requires flexible processing (Apperly and Butterfill, 2009) and its challenges may be met in a number of ways (Kessler et al., under review), and may be dependent on the precise requirements of a problem and even individual differences (Kessler and Wang, 2012). Further studies may look to experimentally manipulate strategy use through systematically priming the use of conditional rules, geometrical comparison and embodied self rotation or through using a dual-task situation to occupy resources for language, imagined spatial manipulation or proprioception.

SPATIAL PERSPECTIVE-TAKING

Evidence that we use spatial alignments of perspectives to calculate the perceptions of others suggests perspective-taking is reliant on an understanding of the relationships between people and objects in space. There are, of course, many judgements that explicitly require us to use such relationships, with no pretext of mental state use whatsoever. When I ask a colleague to pass the coffee cup that's to her left, I'm using my understanding of her intrinsic frame of reference- her spatial perspective. Interestingly, some cultures do not use these spatial perspectives for these kinds of judgements, preferring the absolute reference frame- pass the cup that is nearer to the river than you are (Levinson, 1996; Bowerman and Choi, 2003). We show, here and across two experiments, that these judgements that something is to someone's left or right require embodied self rotation. Like judgements of how a numeral looks, they are sensitive to both the angular disparity between us and the person whose perspective we take *and* to the consistency of our current body position (replicating the findings of Kessler and Thomson, 2010; and Kessler and Rutherford, 2010). It seems that to judge that a coffee cup is to someone's right involves us imagining that we are where they are and then judging if the coffee cup would be to our left or

right. Quite noticeably, the effects of angular disparity and body posture consistency were stronger in the spatial than visual perspective judgment conditions, suggesting that the use of embodied self mental rotation was a strategy more widely used across trials and participants. One possibility is that this is the result of us using our own body representations as a cue to remember the locations of left and right (in England and in Belgium, for example, a common strategy is to remind school children that your “right is the one you write with”). An interesting further question addresses whether such judgements of spatial perspectives are principally for or exclusive to reasoning about human others. There is good evidence that human and non-human spatial transformation do not necessarily use the same cognitive (Zacks et al., 2000) or neural (Zacks and Michelon, 2005) processes. On the other hand, no studies have examined the processes used for locating objects relative to other people or other objects (rather they have focused on identifying the left or right arm of a person vs. the left or right side that a handle of a cup is on). Similarly, we may predict a role for strategy (Kessler and Wang, 2012) and for specific expertise- such as a tennis fan who can quickly judge a ball as being to Novak Djokovic’s forehand (right) side or Rafael Nadal’s forehand (left) side or a naval officer who can quickly conclude that a shoal of dolphins is to the port (left) of the HMS Ark Royal.

COMPARING VISUAL AND SPATIAL PERSPECTIVE-TAKING

Similarities. On the basis of the findings of Experiment 2 in which body posture consistency effects were found on both types of perspective-taking judgements, we have concluded that both Spatial perspective judgements of an object as being to the left or right of someone and visual perspective judgements of how something looks to them recruit processes including embodied self rotation. Under tightly controlled experimental conditions, in which participants take another’s perspective on multiple occasions, both sets of judgment are sensitive to the angular disparity between the target other person and the self viewpoint *and* to the current orientation of the self body. We suggest that an important step for each problem is to imagine ourselves in the position of the other.

Differences. It is clear that further processing beyond an embodied rotation is required to solve these problems and that this processing necessarily differs for each task. Mature visual perspective-taking must take into account individual characteristics of their target: blindfolds, blurred vision or a lack of attention can significantly change how we judge another’s visual perspective in a way that is not required for spatial perspective-taking. These extra demands of visual perspective-taking may be in part responsible for the fact that our embodiment effect was less reliable for visual than for spatial perspective-taking. In Experiment 1, there was no evidence of an effect of consistency of body posture for visual perspective-taking and the effect in Experiment 2 was significantly stronger for spatial perspective-taking. We follow Kessler and Wang (2012) in promoting the idea that in these effortful perspective-taking tasks strategies may differ between individuals and on the basis of specific task demands. Our experiments suggest that variable strategy use was more prevalent for visual than for spatial perspective-taking. We also found evidence that spatial perspective-taking was substantially more difficult than

visual perspective-taking in both experiments. We believe the most parsimonious explanation of this is that we use embodied self rotation and then simulate the perspective from that position. As judging objects as being to one’s own left or right is likely to be more difficult than judging how a number looks (a simple, automatized reading process) this would explain the overall difference.

Visual and spatial perspective-taking also differed in the nature of their relationship with Angle. We concluded that both processes required rotation, based on their linear relationship with Angle in Experiment 1. There was, however an interaction between Angle and Content. Following up this interaction showed that the precise pattern of added difficulty gained with increasing angle was not identical between the two kinds of perspective-taking. Most notably, while the difficulty of taking spatial perspectives grew *most* substantially between 60° and 120°, for visual perspective-taking, this comparison did not reach significance. Similarly, while Experiment 2 showed a robust effect of Angle in spatial perspective-taking, this was not the case in visual perspective-taking. These findings differ somewhat from the findings of Surtees et al. (2013), in which we used a similar method without the physical act of rotating the body, although importantly, both studies show a basic linear relationship between angular disparity and the difficulty of visual perspective-taking. We suggest that this rotation may have made some difference to both the exact nature of processing difficulty at different angles and to the variability in responses. Exactly explaining these specific differences may require further study, but the matter of key importance is that minor experimental changes affected spatial and visual perspective-taking differently, further suggesting that though they adopt similar processes, there are still clear differences in the instantiation of these processes.

THE DEVELOPMENT OF VISUAL AND SPATIAL PERSPECTIVE-TAKING

Identification of similar strategies for spatial judgements of left/right and visual judgements of *how* something looks to someone else is consistent with the developmental profile of these abilities. The ability to make left/right judgements (Hands, 1972) develops after the ability to make front/back judgements (Harris and Strommen, 1972; Cox, 1981; Bialystok and Codd, 1987). Similarly, judgements of *how* something looks (level-2 visual perspective-taking) are achieved after judgements of whether or not someone can see something (Flavell et al., 1981; Moll and Tomasello, 2006; Moll and Meltzoff, 2011). Our current findings imply one possible explanation for this. That the most common and robust method for achieving both of these processes requires embodied mental self rotation, suggests that it may be difficulties with this embodied rotation, rather than with perspective-taking *per se* that is evidenced in developmental studies. There is much debate and conflicting evidence regarding children’s abilities in object rotation (Perrucci et al., 2008), even after the age they pass standard perspective-taking tasks. To our knowledge, however, there has been no systematic investigation of their abilities at mental self rotation.

That success on level-2 visual perspective-taking tasks may be dependent on embodied self-rotation allows for one of two broad alternative explanations. Firstly, children may have the basic

conceptual apparatus to succeed in level-2 perspective-taking situations, even before they pass, but this conceptual knowledge may be obscured by lacking the domain general ability to imagine rotating their position in the world. This alternative is supported by findings of precocious performance on a level-2 type task employing color filters with 3-year olds (Moll and Meltzoff, 2011), rather than angular differences in perspective. Secondly, embodied self rotation may play a causal role in children learning the abstract, non-spatial notion of perspective. This idea, that increased processing flexibility may play a crucial role in children's development of complex concepts, has been suggested by Russell (1996) in relation to children's agency helping them to learn about the world. Investigation of the relative development of rotation and effortful perspective-taking should tell us whether rotation is necessary for learning perspective concepts, necessary for achieving perspective transformations in young children or co-opted once adults have developed a range of perspective-taking strategies and have substantial executive resources.

CONCLUSION

When we interact in complex social environments we undertake complex visuo-spatial reasoning which may or may not involve thinking about the mental states of other people. Taxing judgements of how the world appears to someone else and what things

are located to the left or the right of them seem to involve a comparable process of embodied self rotation. We imagine ourselves in the position of a target other. To do this we take as a starting point the current position of our own body as well as the visual input of a scene in front of us. Embodied perspective-taking processes are robust processes effective in generating visual perspectives of anyone whose basic perceptual apparatus is the same as ours and generating spatial perspectives of anyone who shares our basic anatomy. That is not to say that these processes are the same *in toto*, but rather that they share common processing features and strategy use. These processes are relatively costly and solve problems that are beyond the abilities of very young children. Further studies may look to consider what in these processes responds solely to target human others (as opposed to objects), how we deal with special cases in which other's perceptual access is compromised and how experts overcome the costly nature of this perspective-taking process.

ACKNOWLEDGMENTS

This research was funded by a FSR (Fonds spéciaux de recherche)/Marie Curie Actions of the European Commission post-doctoral fellowship awarded by the Université catholique de Louvain to Andrew Surtees. The authors would like to thank Jessica Wang, who appears in **Figure 3** and provided helpful discussion.

REFERENCES

- Apperly, I. A., and Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychol. Rev.* 116, 953–970. doi: 10.1037/a0016923
- Bialystok, E., and Codd, J. (1987). Children's interpretations of ambiguous spatial descriptions. *Br. J. Dev. Psychol.* 5, 205–211. doi: 10.1111/j.2044-835X.1987.tb01055.x
- Bowerman, M., and Choi, S. (2003). "Space under construction: language-specific spatial categorization in first language acquisition," in *Language in Mind* eds D. Gentner and S. Goldin-Meadow (Cambridge, MA: MIT Press), 387–428.
- Call, J., and Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends Cogn. Sci.* 12, 187–192. doi: 10.1016/j.tics.2008.02.010
- Carlson-Radvansky, L. A., and Irwin, D. E. (1993). Frames of reference in vision and language – where is above. *Cognition* 46, 223–244. doi: 10.1016/0010-0277(93)90011-J
- Carlson-Radvansky, L. A., and Jiang, Y. H. (1998). Inhibition accompanies reference-frame selection. *Psychol. Sci.* 9, 386–391. doi: 10.1111/1467-9280.00072
- Cox, M. V. (1981). Interpretation of the spatial prepositions in front of and behind. *Int. J. Behav. Dev.* 4, 359–368. doi: 10.1177/016502548100400304
- Emery, N. J., and Clayton, N. S. (2004). The mentality of crows: convergent evolution of intelligence in corvids and apes. *Science* 306, 1903–1907. doi: 10.1126/science.1098410
- Espy, K. A. (1997). The shape school: assessing executive function in preschool children. *Dev. Neuropsych.* 13, 495–499. doi: 10.1080/87565649709540690
- Flavell, J. H. (2000). Development of children's knowledge about the mental world. *Int. J. Behav. Dev.* 24, 15–23. doi: 10.1080/016502500383421
- Flavell, J. H., Everett, B. A., Croft, K., and Flavell, E. R. (1981). Young children's knowledge about visual-perception – further evidence for the level 1-level 2 distinction. *Dev. Psychol.* 17, 99–103. doi: 10.1037/0012-1649.17.1.99
- Flavell, J. H., Flavell, E. R., and Green, F. L. (1983). Development of the appearance reality distinction. *Cogn. Psychol.* 15, 95–120. doi: 10.1016/0010-0285(83)90005-1
- Hamilton, A. F. D., Brindley, R., and Frith, U. (2009). Visual perspective-taking impairment in children with autistic spectrum disorder. *Cognition* 113, 37–44. doi: 10.1016/j.cognition.2009.07.007
- Hands, L. J. (1972). Discrimination of left and right and the development of the logic of relations. *Merrill Palmer Q.* 18, 307–320.
- Hare, B., and Tomasello, M. (2005). Human-like social skills in dogs? *Trends Cogn. Sci.* 9, 439–444. doi: 10.1016/j.tics.2005.07.003
- Harris, L. J., and Strommen, E. A. (1972). Role of front-back features in children's front, back, and beside placements of objects. *Merrill Palmer Q. Behav. Dev.* 18, 259–271.
- Huttenlocher, J., and Presson, C. C. (1973). Mental rotation and perspective problem. *Cogn. Psychol.* 4, 277–299. doi: 10.1016/0010-0285(73)90015-7
- Kaminski, J., Riedel, J., Call, J., and Tomasello, M. (2001). Domestic goats (*Capra hircus*) follow gaze direction and use some social cues in an object choice task. *Anim. Behav.* 69, 11–18. doi: 10.1016/j.anbehav.2004.05.008
- Keehner, M., Guerin, S., Miller, M., Turk, D., and Hegarty, M. (2006). Modulation of neural activity by angle of rotation during imagined spatial transformations. *NeuroImage* 33, 391–398. doi: 10.1016/j.neuroimage.2006.06.043
- Kessler, K., Apperly, I. A., and Butterfill, S. A. (under review). Cross-disciplinary views on visual perspective-taking: where and how to look for the embodied origins of mentalizing.
- Kessler, K., and Rutherford, H. (2010). The two forms of visuo-spatial perspective-taking are differently embodied and subserve different spatial prepositions. *Front. Psychol.* 1, 1–12.
- Kessler, K., and Thomson, L. A. (2010). The embodied nature of spatial perspective-taking: embodied transformation versus sensorimotor interference. *Cognition* 114, 72–88. doi: 10.1016/j.cognition.2009.08.015
- Kessler, K., and Wang, H. (2012). Spatial perspective-taking is an embodied process, but not for everyone in the same way: differences for gender and social skills score. *Spat. Cogn. Comput.* 12, 133–158. doi: 10.1080/13875868.2011.634533
- Kirkham, N. Z., Cruess, L., and Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimension-switching task. *Dev. Sci.* 6, 449–467. doi: 10.1111/1467-7687.00300
- Kozhevnikov, M., and Hegarty, M. (2001). A dissociation between object manipulation spatial ability and spatial orientation ability. *Mem. Cognit.* 29, 745–756. doi: 10.3758/BF03200477
- Levinson, S. C. (1996). Language and space. *Annu. Rev. Anthropol.* 25, 353–382. doi: 10.1146/annurev.anthro.25.1.353
- Levinson, S. C. (2003). *Space in Language and Cognition: Explorations in Cognitive Diversity*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511613609
- Light, P., and Nix, C. (1983). Own view versus good view in a perspective-

- taking task. *Child Dev.* 54, 480–483. doi: 10.2307/1129709
- Masangkay, Z. S., McCluske, K. A., McIntyre, C. W., Simsknig, J., Vaughn, B. E., and Flavell, J. H. (1974). Early development of inferences about visual percepts of others. *Child Dev.* 45, 357–366. doi: 10.2307/1127956
- Michelon, P., and Zacks, J. M. (2006). Two kinds of visual perspective-taking. *Percept. Psychophys.* 68, 327–337. doi: 10.3758/BF03193680
- Moll, H., and Meltzoff, A. N. (2011). How does it look? Level 2 perspective-taking at 36 months of age. *Child Dev.* 82, 661–673. doi: 10.1111/j.1467-8624.2010.01571.x
- Moll, H., and Tomasello, M. (2006). Level 1 perspective-taking at 24 months of age. *Br. J. Dev. Psychol.* 24, 603–613. doi: 10.1348/026151005X55370
- Perner, J. (1991). *Understanding the Representational Mind*. Cambridge, MA: MIT Press.
- Perrucci, V., Agnoli, F., and Albiero, P. (2008). Children's performance in mental rotation tasks: orientation-free features flatten the slope. *Dev. Sci.* 11, 732–742. doi: 10.1111/j.1467-7687.2008.00723.x
- Piaget, J., and Inhelder, B. (1956). *The Child's Conception of Space*. London: Routledge & Kegan Paul.
- Premack, D., and Woodruff, G. (1978). Does the chimpanzee have a theory of mind. *Behav. Brain Sci.* 1, 515–526. doi: 10.1017/S0140525X00076512
- Riggs, K. J., Peterson, D. M., Robinson, E. J., and Mitchell, P. (1998). Are errors in false belief tasks symptomatic of a broader difficulty with counterfactuality? *Cogn. Dev.* 13, 73–91. doi: 10.1016/S0885-2014(98)90021-1
- Russell, J. (1996). *Agency: Its Role in Mental Development*. Hove: Erlbaum.
- Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., and Scott, S. E. B. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people see. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 1255–1266. doi: 10.1037/a0018729
- Simon, J. R. (1969). Reactions towards the source of stimulation. *J. Exp. Psychol.* 81, 174–176. doi: 10.1037/h0027448
- Song, H. J., and Baillargeon, R. (2008). Infants' reasoning about others' false perceptions. *Dev. Psychol.* 44, 1789–1795. doi: 10.1037/a0013774
- Surtees, A. D. R., and Apperly, I. A. (2012). Egocentrism and automatic perspective-taking in children and adults. *Child Dev.* 83, 452–460. doi: 10.1111/j.1467-8624.2011.01730.x
- Surtees, A. D. R., Apperly, I. A., and Samson, D. (2013). Similarities and differences in visual and spatial perspective-taking processes. *Cognition* 129, 426–438. doi: 10.1016/j.cognition.2013.06.008
- Surtees, A. D. R., Butterfill, S. A., and Apperly, I. A. (2012a). Direct and indirect measures of Level-2 perspective-taking in children and adults. *Br. J. Dev. Psychol.* 30, 75–86. doi: 10.1111/j.2044-835X.2011.02063.x
- Surtees, A. D. R., Noordzij, M. L., and Apperly, I. A. (2012b). How to lose yourself in space: children and adults' use of frames of reference and perspectives. *Dev. Psychol.* 48, 185, 191.
- Tomasello, M. (2008). *Origins of Human Communication*. Cambridge: MIT Press.
- Tomasello, M., Call, J., and Hare, B. (2003). Chimpanzees understand psychological states – the question is which ones and to what extent. *Trends Cogn. Sci.* 7, 153–156. doi: 10.1016/S1364-6613(03)00035-4
- Tversky, B., and Hard, B. M. (2009). Embodied and disembodied cognition: spatial perspective-taking. *Cognition* 110, 124–129. doi: 10.1016/j.cognition.2008.10.008
- Weisberg, D. P., and Beck, S. R. (2010). Children's thinking about their own and others' regret and relief. *J. Exp. Child Psychol.* 106, 184–191. doi: 10.1016/j.jecp.2010.02.005
- Wimmer, H., and Perner, J. (1983). Beliefs about beliefs – representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13, 103–128. doi: 10.1016/0010-0277(83)90004-5
- Zacks, J., Mires, J., Tversky, B., and Hazeltine, E. (2000). Mental spatial transformations of objects and perspective. *Spat. Cogn. Comput.* 2, 315–332. doi: 10.1023/A:1015584100204
- Zacks, J. M., and Michelon, P. (2005). Transformations of visuospatial images. *Behav. Cogn. Neurosci. Rev.* 4, 96–118. doi: 10.1177/1534582305281085

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 01 July 2013; accepted: 03 October 2013; published online: 05 November 2013.

Citation: Surtees A, Apperly I and Samson D (2013) The use of embodied self-rotation for visual and spatial perspective-taking. *Front. Hum. Neurosci.* 7:698. doi: 10.3389/fnhum.2013.00698

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2013 Surtees, Apperly and Samson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Fractionating the unitary notion of dissociation: disembodied but not embodied dissociative experiences are associated with exocentric perspective-taking

Jason J. Braithwaite^{1*}, Kelly James¹, Hayley Dewe¹, Nick Medford², Chie Takahashi¹ and Klaus Kessler³

¹ Behavioural Brain Sciences Centre, School of Psychology, University of Birmingham, Birmingham, UK

² Sackler Centre for Consciousness Science, University of Sussex, Brighton, East Sussex, UK

³ Aston Brain Centre, School of Life and Health Sciences, Aston University, Birmingham, UK

Edited by:

Sarah H. Creem-Regehr, University of Utah, USA

Reviewed by:

Christopher Charles French, Goldsmiths, University of London, UK
Kyle Timothy Gagnon, University of Utah, USA

*Correspondence:

Jason J. Braithwaite, Behavioural Brain Sciences Centre, School of Psychology, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK
e-mail: j.j.braithwaite@bham.ac.uk

It has been argued that hallucinations which appear to involve shifts in egocentric perspective (e.g., the out-of-body experience, OBE) reflect specific biases in exocentric perspective-taking processes. Via a newly devised perspective-taking task, we examined whether such biases in perspective-taking were present in relation to specific dissociative anomalous body experiences (ABE) – namely the OBE. Participants also completed the Cambridge Depersonalization Scale (CDS; Sierra and Berrios, 2000) which provided measures of additional embodied ABE (unreality of self) and measures of derealization (unreality of surroundings). There were no reliable differences in the level of ABE, emotional numbing, and anomalies in sensory recall reported between the OBE and control group as measured by the corresponding CDS subscales. In contrast, the OBE group did provide significantly elevated measures of derealization (“alienation from surroundings” CDS subscale) relative to the control group. At the same time we also found that the OBE group was significantly more efficient at completing all aspects of the perspective-taking task relative to controls. Collectively, the current findings support fractionating the typically unitary notion of dissociation by proposing a distinction between *embodied dissociative experiences* and *disembodied dissociative experiences* – with only the latter being associated with exocentric perspective-taking mechanisms. Our findings – obtained with an ecologically valid task and a homogeneous OBE group – also call for a re-evaluation of the relationship between OBEs and perspective-taking in terms of facilitated disembodied experiences.

Keywords: perspective-taking, anomalous bodily experiences, out-of-body experience, dissociation, depersonalization

INTRODUCTION

Stable self-consciousness, which supports appropriate behavior and experience, is dependent on a legion of multi-sensory co-ordinated processes acting in concert to maintain a coherent sense of the embodied *self* over space and time. These underlying processes include the multi-sensory spatial coding of both one's own body, the environment, and the constant interactions between body and environment. However, this typically stable process can break down in certain circumstances, leading to striking distortions in body-image and dissociative anomalous body experiences (ABE). One such hallucination that has received growing interest in recent years is the out-of-body experience (OBE).

The OBE can be defined as an experience where the individual “perceives his/her environment from a perspective outside of their physical body.” Therefore, a fundamental core aspect to the OBE is the overwhelming sense that one is experiencing the world from an external, exocentric perspective (Eastman, 1962; Green, 1968; Palmer, 1978; Blackmore, 1982; Irwin, 1985). In this sense OBE has been discussed in relation to deliberate processes of egocentric transformation and perspective-taking (e.g., Blanke et al., 2005; Braithwaite and Dent, 2011).

The current and dominant view is that the OBE occurs due to a temporary disruption in multi-sensory integration processes, where stable egocentric processing has become impaired to such an extent that it can no longer represent a coherent sense of embodied “self” (see Blanke and Arzy, 2005; Blanke and Mohr, 2005; Blanke and Metzinger, 2009 for reviews). Although it is not entirely clear how such transient disruptions occur (even more so in non-clinical samples), other independent findings have shown that OBE groups can display; (i) elevated scores on measures of anomalous experience related to disruptions in temporal-lobe processing; (ii) biases in body-transformation/perspective-taking processes; and (iii) elevated signs of visual cortical hyperexcitability – which were absent from both control groups and non-visual hallucination groups (Braithwaite et al., 2011, 2013a,b).

In addition, behavioral studies have argued that the brain processes involved in the mental transformation of one's own body may be the same as those implicated in the computation of an exocentric perspective (for review, see Kessler and Rutherford, 2010; Kessler and Thomson, 2010; Kessler and Wang, 2012; Popescu and Wexler, 2012; van Elk and Blanke, 2013) and particularly in the OBE (Cook and Irwin, 1983; Blackmore, 1987; Brugger, 2002;

Blanke and Arzy, 2005; Blanke et al., 2005; Arzy et al., 2006; Mohr et al., 2006; Easton et al., 2009; Overney et al., 2009; Braithwaite et al., 2011). Most of the latter studies used performance at the “own-body-transformation” (OBT) task to explore perspective-taking and have implicated the temporal–parietal junction in the mental transformation of one’s own body and perspective (see Blanke et al., 2005). However, only a handful of these studies actually explored performance on this task in direct relation to samples reporting OBEs – and these have produced diverse results (Easton et al., 2009; Braithwaite et al., 2011).

Interestingly, impairments and not benefits, at OBT tasks have been shown for participants who scored positively on a measure of perceptual aberrations related to schizotypy (Mohr et al., 2006) and more recently for those specifically reporting OBEs (Braithwaite et al., 2011; though see also Easton et al., 2009). These tasks present observers with a schematic figure which is either facing the observer or facing away from the observer. Participants are instructed to try to adopt the perspective of the figure and hence engage perspective-taking processes and decide on what hand (left/right) the figure wearing a distinctive glove and bracelet.

Although these tasks were originally thought to measure similar perspective-taking mechanisms to that implicated in the out-of-body perspective, findings where schizotypes and OBE groups were impaired at the task, appear to go against the intuitive idea that those reporting dissociative experience should be better at exocentric perspective-taking. Whether the typical OBT task truly is an exocentric perspective-taking task has now been questioned on the grounds that with only two exemplar avatars, other rule-based contingency strategies may be impacting more on performance rather than exocentric perspective-taking (Braithwaite and Dent, 2011; Gardner and Potts, 2011; Gronholm et al., 2012; Kessler and Wang, 2012; May and Wendt, 2012; see also Pezzulo et al., 2013).

Collectively, the evidence for clear benefits in perspective-taking, for those individuals prone to anomalous disembodied and dissociative experiences, is currently unclear, contentious, and awaiting clarification. This is likely due, in part, to; (i) diverse methodologies used to examine such processes; (ii) not all previous studies claiming to explore the mechanisms of OBEs have actually used OBE samples and; (iii) the use of other distinct groups of hallucinators (e.g., schizotypes) that may themselves reflect quite different underlying mechanisms that do not include exocentric hallucinations. These different mechanisms may well be masked as they currently exist under the generic umbrella concept of “*dissociative experience*” not all of which would conceivably index exocentric processes. As a consequence it becomes important to examine the OBE not just in its own right, but alongside other similar though distinct dissociative experiences.

Shedding light on this currently ambiguous situation will also help our understanding of the embodied processes involved in more deliberate forms of perspective-taking, where the social and/or spatial goals might be conscious and deliberately chosen, yet, where the actual mechanism for transforming the “ego” into an exocentric perspective seems to be strongly embodied (Kessler and Thomson, 2010) and compulsory rather than deliberately chosen (Kessler and Wang, 2012), and might therefore strongly resemble the spontaneous OBT underlying OBE.

DEPERSONALIZATION, DEREALIZATION, AND THE OBE

Early accounts for the OBE came from psychiatry, where it was cast as a specific instance of depersonalization (Noyes and Kletti, 1976, 1977). Depersonalization disorder (DPD) is a syndrome which reflects a severe disruption in self-awareness that can include dissociative experiences (Sierra and David, 2011). Depersonalization itself typically refers to an unreality of the self. Patients classically describe feelings of remoteness, estrangement from the self, feeling like a robot or automaton, and a flattening of emotional affect (Sierra, 2009; Sierra and David, 2011). The related concept of derealization (DR) which can commonly co-occur with depersonalization, refers more to an unreality of surroundings – where patients typically describe experiencing the world through a fog, a veil, a bubble and being “detached” from their surroundings (Sierra and David, 2011).

The relationship between OBEs and DPD-DR has been questioned. For example, in the OBE the experience is often described as being extremely vivid, convincing, striking, and very real. Individuals often describe a heightened sense of awareness and increased clarity of thought during the experience (see Blackmore, 1982). In contrast, DPD-DR experiences are often described as having a dulled or flattened affect, loss of emotional coloring, and can be somewhat dreamlike (Gabbard et al., 1981, 1982; Twemlow et al., 1982). In addition, DPD-DR experiences typically occur in stressful situations, whereas the OBE can equally occur spontaneously in quite relaxed conditions. These phenomenological and contextual differences have led to the view that OBEs and the ABE reported in DPD-DR are not the same and may reflect quite different neurocognitive underpinnings (Gabbard et al., 1981, 1982; Blackmore, 1982; Twemlow et al., 1982; Gabbard and Twemlow, 1984, 1986; see also Sierra, 2009 for a discussion).

There is some confusion over the terminology used when describing the anomalous experiences reported by DPD-DR patients that may contribute to continued misunderstandings about the prevalence of OBEs in DPD-DR as well as the clinical construct of DPD-DR itself (see Sierra and Berrios, 1997; Medford et al., 2005 for detailed discussions). For example, while some experiences might be described as “disembodied” or “dissociative” OBEs themselves are rarely, if ever, reported by patients with DPD-DR. What patients appear to be describing is that they feel their bodies are unreal and do not belong to them. However, a closer examination of these accounts shows that the perceiving “self” is still typically described as being located inside the physical self – so there is no external “disembodiment” or shift in experiential perspective. The term “disembodiment” and perhaps to a lesser extent “dissociation” can be taken to imply that DPD-DR experiences commonly involve experiences where the perceiving “self” shifts perspective from an egocentric and embodied one, to an exocentric and disembodied one (an OBE). However, for DPD-DR this is rare, so much so that some have noted the complete absence of OBEs in DPD-DR patient populations (Sierra, 2009).

OVERVIEW OF THE CURRENT STUDY

The present study sought to examine cognitive biases in perspective-taking/body-transformation processes that may be implicated in predisposition to hallucinatory experiences that

involve a shift in self-perspective (the OBE). If the striking phenomenological aspects of OBEs are based in some form of involuntary exocentric perspective-taking, then individuals prone to OBEs may also display distinct performance in a deliberate perspective-taking task. An intuitive prediction is that those prone to OBEs would be better at a perspective-taking as they may recruit the same transformational mechanisms underlying the OBE. Although some previous research has shown the opposite pattern, where OBE groups have shown impaired performance (Braithwaite et al., 2011), the actual task employed in these studies has been questioned (Braithwaite and Dent, 2011; Braithwaite et al., 2011).

Therefore, a new perspective-taking task was devised for this study, where a human female avatar could be viewed from either an “Above” viewpoint (above the head of the avatar) or “Below” viewpoint (below the feet of the avatar). Thus, unlike many previous studies, here the avatar was rotated around the horizontal axis and not the more typical vertical axis (or what some describe as around the sagittal plane and not around the transverse plane; Carpenter and Proffitt, 2001; Creem et al., 2001). In addition to any transformation of plane/viewpoint required, half of the stimuli also required a (mental) rotation of the participant’s body in order to fully transform and match their perspective to that of the avatar (see Figure 1).

There were two advantages from these new manipulations. Firstly, these manipulations produced eight separate avatars, four from the above viewpoint and four from the below viewpoint, but two of these four also differed in terms of requiring body-rotation. Previous OBT tasks have typically used only exemplars with two different body positions (e.g., facing/behind). As a consequence the current study is arguably more resistant to the emergence of non-spatial basic contingency-based or rule-based strategies emerging across trials.

Secondly, the use of “Above” viewpoints is more phenomenologically similar to the perspective reported in many visual OBEs (see also Schwabe et al., 2009). As a consequence, the current

perspective-taking transformations are more in line with those implied in accounts of OBEs. Finally, the presence of both a transformation of plane and body-rotation facilitates a separate exploration of these factors in relation to overall body transformations, perspective-taking and spatial processing in relation to OBEs.

In addition to the new behavioral tasks, all participants were measured for their proneness to dissociative anomalous experiences via the administration of the Cambridge Depersonalization Scale (CDS; Sierra and Berrios, 2000), which contains measures of both ABE and anomalous experiences of one’s surroundings (derealization). As noted in the Introduction, there has been some debate as to the relationship between depersonalization and OBEs (Noyes and Kletti, 1976, 1977; Gabbard et al., 1981, 1982; Blackmore, 1982; Gabbard and Twemlow, 1984; see also Sierra, 2009). However, there have been few, if any, experimental investigations of these factors together. Importantly, the ABE measured by the CDS are more related to embodied anomalous experiences, where the self remains within the body and is not transposed into an exocentric perspective. It is not at all clear whether OBE groups also experience elevated levels of these potentially related experiences or whether the OBE tends to occur in isolation to these other experiences. In addition, the CDS also contains a measure of derealization, where individuals report being cut-off and alienated from their surroundings. In light of recent accounts from cognitive neuroscience on the role of a breakdown in multi-sensory integration underlying the OBE, any depletion or disruption in incoming sensory signals from the outside world may act to destabilize internal models of the bodily self. As a consequence, the OBE group may well display elevated signs of derealization, even more so, than the embodied ABE associated with depersonalization *per se*.

MATERIALS AND METHODS

PARTICIPANTS

Sixty-two participants took part in the present study. Of these, 47 (82%) were female and 60 (96%) reported that they were right-handed. None reported any personal medical history of seizure, epilepsy or were diagnosed as having migraine. All participants were undergraduate or postgraduate students (MSc/PhD) from the School of Psychology at the University of Birmingham, UK. Participants ranged in age from 18 to 28 years (average age of 21.5 years). All received course credit for taking part in the study.

QUESTIONNAIRE MEASURES

The Cambridge Depersonalization Scale

The CDS (Sierra and Berrios, 2000) is a 29-item psychometrically established measure of dissociative anomalous experiences associated with the construct of depersonalization (anomalous experiences of the “self”) and derealization (anomalous experiences of one’s surroundings). Two responses to each question are given on 5-point Likert scales, one response for “Frequency” and one for “Duration” and the final score for any item is the summed output of both these responses (giving a potential range of scores between 0 and 290).

It is now recognized that clinically significant depersonalization–derealization (DPD-DR) is best considered as a syndrome rather than a single phenomenon (Sierra, 2009), since it involves

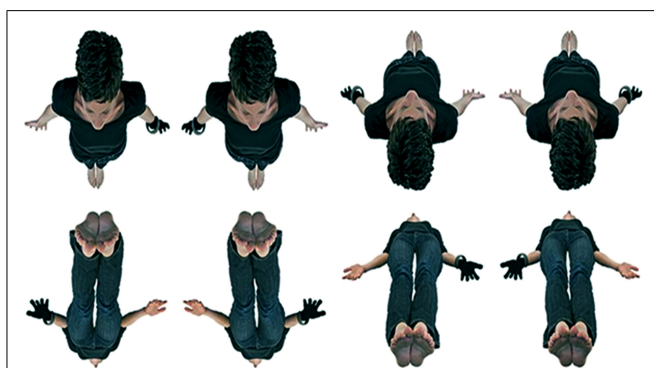


FIGURE 1 | Stimuli from the new Human-own-body-transformation task (HOBT) where avatars from both ‘Above’ (top row) and ‘Below’ (bottom row) perspectives were used. Note, while all avatars require the transformation of plane, the two avatars in the upper row/ left hand side, and lower row right-hand side also require a rotation of body as well. Therefore the stimuli are distinguishable along two a-priori dimensions, one of plane and one of body-rotation.

alterations in the quality of subjective experience across a range of different experiential domains (see, for example, Medford et al., 2005). Although this multi-factorial understanding of DPD-DR has been present in descriptive literature for many decades (see also Ackner, 1954; Sierra and Berrios, 1997) it is only recently that it has been confirmed by empirical phenomenological studies (Sierra et al., 2005; Simeon et al., 2008) which have examined the clustering of CDS items into different factors. In the study by Sierra et al. (2005), CDS items were shown to segregate into four distinct factors, which the authors termed (i) ABE; (ii) emotional numbing (EN, analogous to the term “de-affectualization”), (iii) anomalous sensory recall (ASR), and (iv) alienation from surroundings (AFS or derealization; Sierra et al., 2005). Previous research on patients has shown that a cut-off point of 70 yields a sensitivity of 75% (specificity of 87%) and has high internal consistency (Cronbach's $\alpha = 0.89$) and half-split reliability (0.92; see Sierra and Berrios, 2000; Sierra, 2009). Importantly, it should be noted that there is no explicit question on the CDS for OBEs. The ABE questions typically describe anomalous states that are more associated with embodied perceptions (see General Discussion for further elaborations)¹.

The OBE pre-screen

A pre-screen questionnaire to establish the presence of OBEs and some basic phenomenological information about them was also administered. This questionnaire has been used and detailed in previous studies from our laboratory (Braithwaite et al., 2011, 2013a,b). Participants are initially asked the question: “Have you ever had an experience where you have perceived/experienced the world from a vantage point outside of the physical body?” In addition to this question participants were given further qualifying information that (i) such an experience can feel totally real at the time of the experience with all the phenomenological qualities of veridical perception and (ii) that such experiences can be fleeting and transient or more sustained. If a response of “yes” was provided then additional contextual and situational information about the experience(s) was also ascertained such as how often they had experienced an OBE, whether the experience was visual in nature, whether they saw their physical self during the experience, and the perspective from which they experienced the world or self (above, below, in front, behind, laterally, or other). Associated phenomenology was also documented (e.g., feelings of dizziness, floating sensations, disorientation, dissociation, duality of consciousness, other sensory experiences, etc). This questionnaire also allowed us to ensure the participant themselves had not incorrectly defined their own experiences as OBEs, when in fact they might not be consistent with classical definitions.

PROCEDURE AND STIMULI: PERSPECTIVE-TAKING TASK

All participants took part in a newly devised version of a perspective-taking task, which for clarity and conciseness we

now refer to as the Human OBT (HOBT) task. Unlike previous versions of the OBT task, the present stimuli consisted of both aerial (elevated/above the avatars head) and low (beneath the avatars feet) color photographic views of a human female avatar. In each photograph, the avatar was wearing a distinctive glove/bracelet on one hand. The avatar could be facing in two directions (toward the top or bottom of the screen), from either the elevated or beneath viewpoints, thus generating eight possible exemplar photographs (four from each viewpoint) when combined with the differing hands wearing the glove/bracelet. To successfully solve the task the avatars differed on two main *a priori* dimensions.

For example, all avatars required a transformation of plane where the viewpoint of the participant or the avatar itself could be transformed. In addition, some of the avatars (see **Figure 1**) also required an additional step of mental body-rotation in order to match the perspectives between participant and avatar. The *a priori* prediction was that reaction times (RTs) for those avatars requiring the additional step would be increased. These stimuli were presented centrally, at fixation, against a white background on an 17-inch Samsung CRT monitor coupled to a Pentium PC. The stimuli are shown in **Figure 1**. The experiment was programed in E-prime software v2.1 (Psychology Software Tools).

The stimuli were viewed at an unfixed but general distance of 60 cm and were approximately 110 mm wide \times 75 mm high. Each trial began with the presentation of a black central fixation cross on a white background. The fixation cross was presented for 1000 ms followed by the presentation of the human avatar which remained on the screen until response. There was an inter-stimulus interval of 1000 ms between trials.

All stimuli were presented within one single block of 96 trials (48 per perspective). Participants were instructed to imagine themselves to be in the figure's body position and to adopt the appropriate perspective of the figure. Once done, they had to respond to whether the glove was on the left hand (up-arrow keyboard response) or right hand (down-arrow keyboard response) of the human avatar. The presentation of the different stimuli was randomized within the experimental block of trials. The experiment began with a separate block of 16 practice trials which were not analyzed but used so that participants could learn the correct response-mapping. Participants were instructed to respond as fast and as accurately as they could. The experiment lasted approximately 40 min (including the administration of the questionnaires). The questionnaires were always completed after the perspective-taking task.

RESULTS

For the perspective-taking task, RTs were made fit for analysis in the following way. Firstly, all incorrect responses were identified and removed from the analysis. This revealed an overall response accuracy rate of 94%. Secondly, all outliers (deemed at ± 2.5 standard deviations from the mean) and responses faster than 200 ms were also discarded. Any participant with less than 80% accuracy at the task was also removed from the analysis. This procedure

¹One question on the CDS does ask about feeling “as if” one is outside of the body (question 23) however, this is not regarded as equivalent to the more direct question of actually “perceiving” the world from an external point of view. In addition, further questioning with DPD patients reveals this is rarely, if ever, defined as an OBE by the patient.

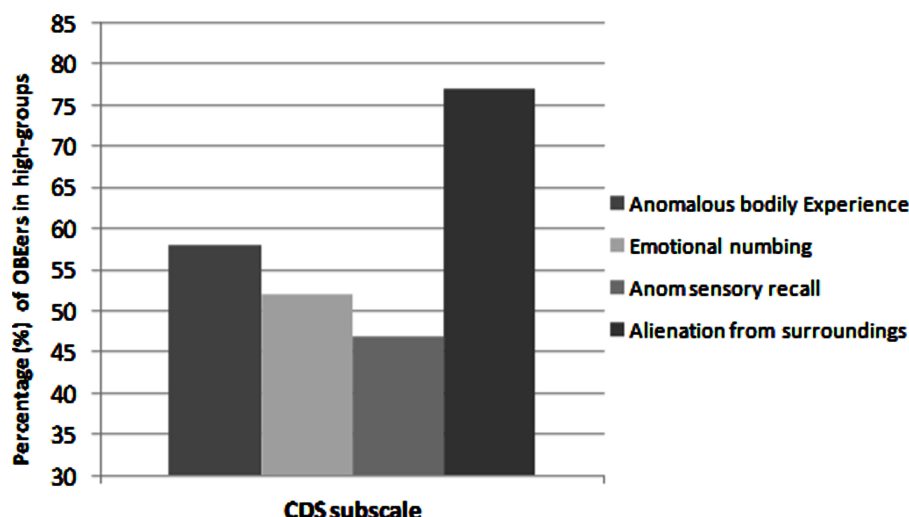


FIGURE 2 | Descriptive statistics (percentages) of those reporting OBEs in the high-groups of each subscale from the Cambridge Depersonalization Scale (CDS).

led to five participants being removed from the analysis². The following analysis was carried out on the remaining sample of 57 participants. An overall measure of performance was calculated where the proportion of correct responses was divided into the RTs providing a measure of efficiency (Townsend and Ashby, 1983; see also Rach et al., 2011). All statistical tests are reported two-tailed and, where necessary, *p*-values have been corrected for multiple comparisons (via the Bonferroni procedure) and corrected degrees of freedom are reported if non-homogenous variability occurred.

Of the remaining participants, 17 (30%) claimed to have experienced an OBE at some point in their life. The remaining 70% made up the non-OBE control group. The OBE pre-screen questionnaire revealed that the entire OBE group reported their experiences had a strong visual component to them, where they experienced themselves or their local environment from an external and exocentric perspective. In addition, all reported an elevated perspective to their experiences, as if they were looking down on the world and/or themselves. Although other multi-sensory information was also noted and contributed to the realism of the experience (e.g., vestibular distortions/floating sensations) in all cases these always co-occurred with visual aspects of the experience.

CAMBRIDGE DEPERSONALIZATION SCALE

Overall summed scores were explored for normality via a Shapiro-Wilk test and were found to be borderline non-normally distributed [$W = 0.96$ ($df = 57$), $p < 0.05$]. As a consequence, these questionnaire data were explored with non-parametric statistics. The overall sample mean score for the CDS was $\bar{X} = 30.5$ (median = 29.3, and range = 0–84). Two participants scored just

above the score of 70 (scores of 71, 84) and one was borderline (score of 66).

A median-split analysis was carried out independently on all four subscales of the CDS and the percentage of those reporting OBEs occurring in the high-groups of these subscales was calculated (see **Figure 2**). This revealed that the high-ABE and AFS groups contained the largest numbers of those reporting OBEs. Interestingly, these descriptive statistics show that 77% of those reporting OBEs placed in the high-AFS group (i.e., increased signs of derealization).

The mean CDS scores for all subscales and for both the OBE group and non-OBE controls are graphically represented in **Figure 3**. These differences were formally compared by a series of Mann-Whitney *U*-tests. Although the largest effects appear to be present for both ABE and AFS measures, after correction for multiple comparisons, only the difference between the groups for the AFS subscale was significant ($U = 176.00$, $Z = -2.88$, $p < 0.005$). The OBE group produced significantly higher scores on measures of AFS ($\bar{X} = 10.8$, $SE = 1.6$) than the control non-OBE group ($\bar{X} = 5.2$, $SE = 1.2$; see **Figure 3**). Although this general pattern also held for measures of ABE (OBE $\bar{X} = 11.8$, $SE = 1.8$; non-OBE control $\bar{X} = 7.5$, $SE = 0.08$), this was not reliable after correction for multiple comparisons ($U = 233.50$, $Z = -1.86$, $p = 0.08$).

Seventy-seven percent of those claiming OBEs in the present sample placed in the high-AFS group (suggesting that the majority of this group displayed elevated signs of derealization experiences). In addition, the OBE group reported significantly higher degrees of AFS relative to the non-OBE control group. The effect for the OBE group to display increased scores on measures of ABE, though showing signs of being present, failed to be reliable. No other factors reliably distinguished the groups.

PERFORMANCE AT THE HOBT TASK

Mean correct efficiency RTs for the HOBT task are plotted in **Figure 4**. Performance at the HOBT task was examined

²All removed participants were from the control group. Exit questioning revealed that some described the experiment as “too hard” so they did not engage fully with the experiment and others that they were confused about the instructions.

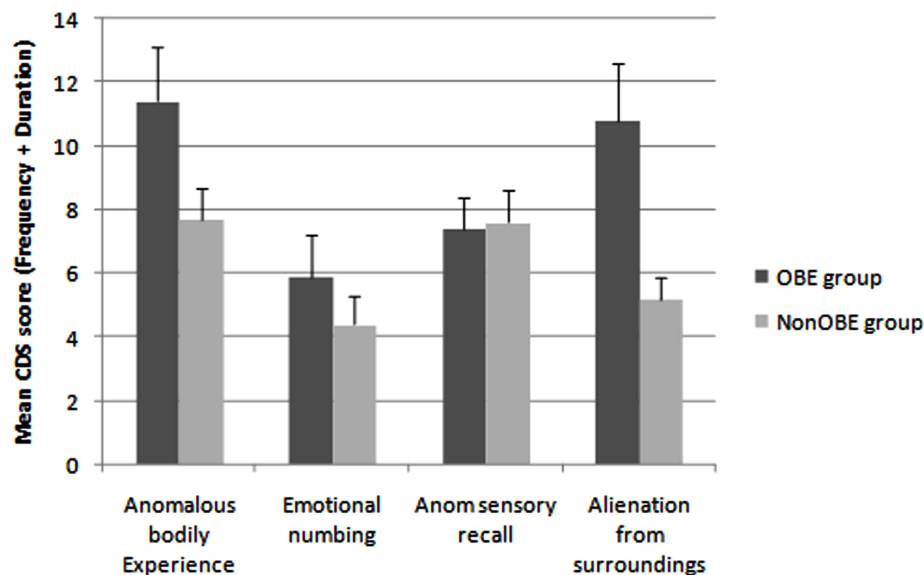


FIGURE 3 | Mean CDS scores for each of the 4 subscales (identified by Sierra and Berrios, 2000) plotted for both the OBE and non-OBE control groups (error bars = 1 SE).

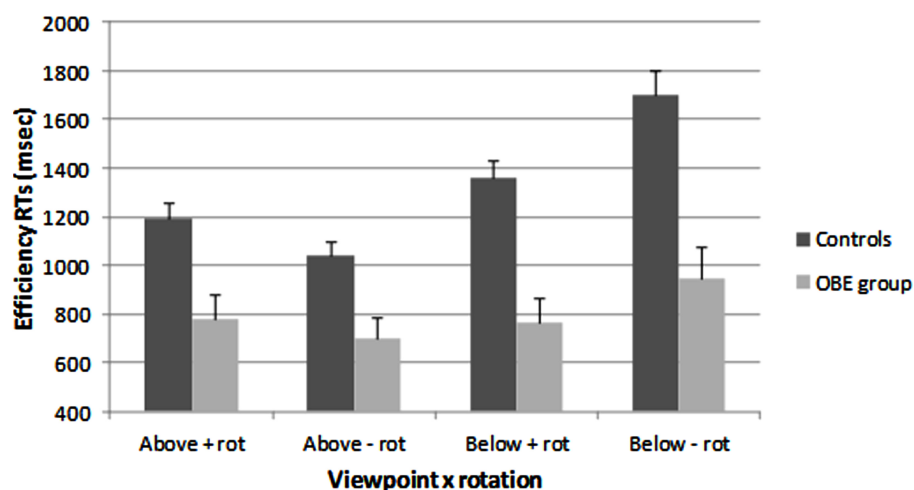


FIGURE 4 | Efficiency RTs for the HOBT task plotted for both Groups (Controls and OBE group) both perspectives (Above and Below) and whether an additional body-rotation was required or not (+ rot = requires body rotation / - rot = does not require body rotation: error bars = 1 SE).

by a 2 (Group: Controls vs OBE group) \times 2 (Viewpoint Above/Below) \times 2 (Body rotation: Rotation vs No rotation) mixed ANOVA applied to the efficiency RTs. The main effect of Group was significant, $F(1, 55) = 24.33$, $p < 0.001$; as was the main effect of Viewpoint, $F(1,55) = 30.70$, $p < 0.001$. On the whole, the OBE group was significantly more efficient ($\bar{X}_{\text{diff}} = 528$ ms) than the non-OBE control group at the HOBT task. In addition, both groups were significantly more efficient overall at Above viewpoints, relative to below viewpoint ($\bar{X}_{\text{diff}} = 264$ ms). In contrast, the main effect of Rotation was not significant, $F(1,55) = 1.67$, $p = 0.202$ ($\bar{X}_{\text{diff}} = 73$ ms). The Viewpoint \times Group and the Viewpoint \times Rotation interactions were significant, $F(1,55) = 10.04$,

$p < 0.005$; and $F(1,55) = 15.64$, $p < 0.001$, respectively. However, the Rotation \times Group interaction was not significant, $F(1,55) = 0.178$, $p = 0.674$. Finally, the three-way interaction between Group \times Viewpoint \times Rotation was not significant, $F(1, 55) = 1.40$, $p = 0.242$.

The significant interactions were explored further via a series of within subjects t -tests carried out separately for each group, for each viewpoint and rotation condition. These data are given in **Table 1**.

To explore the overall cost of viewpoint between the groups, the overall RTs from the “Above” viewpoint were subtracted from RTs from the “Below” viewpoint for both the OBE and control groups.

Table 1 | Breakdown of the separate effects of Viewpoint and Rotation within the two groups.

Condition	\bar{X}_{diff} (ms)	t-Statistic	df	p-Value
Controls: Above rot – No rot	146	2.79	39	< 0.01 *
Controls: Below rot – No rot	–339	2.80	39	< 0.01 *
OBE group: Above rot – No rot	82	1.91	16	= 0.07
OBE group: Below rot – No rot	–180	3.42	16	< 0.005 *

* = results are significant.

This generated two sets of difference scores. These differences were explored via a between-subjects *t*-test which was significant [$t(52.9) = 4.39, p < 0.001$]. On the whole, non-OBE controls were more impaired (by 298 ms) by the cost for below viewpoints than the OBE group.

To summarize, both groups were more efficient at the Above viewpoint compared to the Below viewpoint. In addition, the OBE group was significantly more efficient at all aspects of the HOBT task relative to the non-OBE control group. For Above viewpoints, there was a general trend for a cost to efficiency if an additional body-rotation was required (in addition to any transformation of plane) though this was only reliable for the control group. The pattern of findings for Below viewpoints was reversed with, rather surprisingly, efficiency being improved for those avatars that might require an additional step of body-rotation as well as any transformation of plane. These findings are discussed more fully in Section “General Discussion.”

GENERAL DISCUSSION

The present study examined biases in exocentric perspective-taking/body-transformation processes in relation to predisposition to hallucinatory experiences that involve a shift in self-perspective – the OBE. In addition, signs of embodied anomalous experiences associated with depersonalization/derealization (DPD-DR), with OBE groups, were also explored.

The OBE is, by definition, an anomalous experience revolving around a shift in the perspective of the experiencing self “outside of his/her body.” In line with previous research (Blanke et al., 2005; Braithwaite et al., 2011), a premise of the present study was that if OBEs are based in some form of disruption in the mechanisms underlying stable egocentric processing and/or the efficient use of exocentric perspective-taking processes, then these individuals may display distinct performance at a task which is sensitive to these processes. In addition to this, we also examined the rate and range of other dissociative anomalous experiences to explore their association with the OBE and exocentric perspective-taking.

There was a borderline significant trend for the OBE group to report more additional ABE relative to control groups. This observation for a general trend of elevated egocentric ABEs (associated with depersonalization) for the OBE group is new, though complements other research showing increased somatoform distortions for these groups (Irwin, 2000; Murray and Fox, 2005). Both the ABE subscale in the present study, and the somatoform dissociation scale used by previous studies, include only

items related to either altered bodily sensations, or egocentric dissociative experiences. Clearly, the OBE is a specific form of exocentric ABE and can co-occur with other egocentric dissociative phenomenology. The weaker effects seen here for the ABE subscale are possibly due to the fact that this is a small subscale of items (much smaller than the full measures used in previous studies), containing items more focused on dissociative experiences, rather than specific somatoform distortions (though the two can be related).

In contrast to the pattern seen for all other subscales (ABE, EN, ASR), the OBE group did provide clear and significantly elevated scores on measures of AFS (derealization) compared to the non-OBE control group. Indeed, an exploratory median-split analysis carried out on the whole sample revealed that 77% of the OBE group fell in the high-scoring group for derealization. The relationship between derealization and the OBE is both new and interesting as it might imply that the OBE itself is a response to a temporary lack of connection between the “self,” and the surrounding world.

By this account, the specific neurocognitive biases underlying derealization may increase the disconnection between the bodily self and one’s own surroundings to such an extent that internal representations of the body/self become unstable or degraded in some way. At the very least, incoming sensory information may become ambiguous under conditions of increased derealization. The net consequence of this is that typically stable egocentric representations of the self might become so disrupted that they can no longer support coherent embodied conscious experience. Under some circumstances this might simply result in the dissociative anomalous experiences reported by DPD-DR patients and their non-clinical counterparts (e.g., estrangement from the self, bodies feeling unreal, surroundings feel dreamlike, dull, and deadened). However, in other instances these situations may act as a catalyst for OBEs providing the individual also displays additional cognitive biases in exocentric perspective-taking. This in itself is noteworthy and has implications for the broader debate on whether the OBE is or is not related to DPD-DR (see Sierra et al., 2002; Sierra, 2009; Sierra and David, 2011).

The observation that the OBE group were also significantly more efficient at the objective HOBT perspective-taking task relative to the non-OBE control group is particularly noteworthy. This was the case across all viewpoints and body-rotation permutations of the stimuli. Both groups found the Above viewpoints easier than the Below viewpoints (see also Schwabe et al., 2009 for similar findings with only control groups). This is to be expected and likely reflects both a greater familiarity with seeing bodies from elevated/above viewpoints and also the clear view of the head/shoulder region may act as a useful anchor point (e.g., Kessler and Rutherford, 2010), with which to carry out the transformations necessary to complete the task efficiently and successfully.

An unexpected result was the diverse role of the “Rotation” factor across the different viewpoints. For Above views, there was a general cost to efficiency if both a transformation of the body and plane was required. This cost was significant for control groups, and borderline reliable for the OBE group (see Table 1). This overall finding is in line with our

a priori intuitive prediction that avatars involving two separate transformations (of both plane and body) will be less efficient than those avatars only involving one transformation. The exact opposite pattern occurred for below viewpoints, where RTs were generally increased, but where efficiency was actually benefited by the apparent needs of both a transformation of plane and body-rotation and hampered where apparently only one transformation was required. This result is supported further in that it was observed for both the OBE and control groups.

One possible explanation is that for the “Below – Rotation” condition, and this condition alone, participants may not be carrying out the spatial transformation in a similar manner to the other instances. For example, for both “Above” viewpoints, a clear and familiar view is provided and a salient anchor point (i.e., the head) contribute to the transformations required to efficiently solve the task at hand. Here, either a transformation of plane, of body-rotation, or both are required. It is also, due to the familiar perspective, quite salient which of these processes are best suited to the situation.

However, the “Below – Rotation” condition, presents a view of an avatar which we rarely, if ever experience in daily life: it would require us either looking up at people through a glass floor, or watching superman flying over our heads. In contrast, the “Below + Rotation” condition is identical to a person lying on a bed, thus, a quite familiar view. We therefore suggest that in the “Below – Rotation” condition, instead of a simple transformation of plane, participants may first rotate the whole avatar (like hands rotating on a clock face), in order to place the head toward the top, but in so doing, this now generates the need for an additional body-rotation. Therefore this condition may actually elicit two rotational strategies rather than our assumed one transformation – thus impacting on the efficiency of performance. As suggested, this may be due to the absence of a salient anchor point and unusual view of the human body with which to assign the appropriate initial transformation (e.g., Grabowski, 1999; Kessler and Rutherford, 2010).

EMBODIED AND DISEMBODIED DISSOCIATIVE ANOMALOUS EXPERIENCES

The present study provides preliminary evidence for fractionating the unitary notion of “dissociation” underlying ABE. We suggest that one important factor for consideration when examining the mechanisms underlying dissociative states is whether the dissociation being examined is from an egocentric or “embodied” perspective or whether it is from an exocentric or “disembodied” perspective (or indeed both; e.g., as in cases of heautoscopy; Brugger et al., 1997; Brugger, 2002). As a consequence it might be helpful to conceptually view the legion of dissociative states of the self as being representative of either “*embodied dissociation*” (e.g., dissociative experiences reported in depersonalization, schizophrenic loss of body boundaries, autoscopia, sensed-presence experiences) where the perceiving “self” remains firmly located within the physical body, or “*disembodied dissociation*” (i.e., OBEs) where the perceiving self appears liberated from its egocentric physical moorings. Only the latter implies a bias for additional exocentric perspective-taking

processes underlying the phenomenology of the anomalous experience^{3,4}.

Although speculative, this view is supported by findings from the present study as well as the broader literature. The crucial and major difference between the groups in the current investigation appears to have been the presence of exocentric OBEs, which may have resulted from the co-presence of elevated signs of derealization and biases in exocentric perspective-taking processes. It was clearly the case that the OBE group experienced other forms of egocentric ABEs, but the presence of these additional egocentric ABEs did not appear to be as strongly related to performance at the exocentric perspective-taking task.

Therefore, although the OBE group was a group which reported additional non-exocentric ABEs, performance at the HOBT task appeared to be related more to the co-presence of *disembodied dissociative* experiences that may well have been reliant on an exocentric representation of the self in space (the OBE). The control group, by definition, did not report any instances of *disembodied dissociative* experiences. In addition, their performance at the exocentric perspective-taking task was significantly less efficient than that of the OBE group.

Interestingly, Sierra (2009) notes that while the concept of disembodiment does imply an experience where the “self” is localized outside one’s physical body (analogous to the OBE), in cases of depersonalization, disembodiment is certainly not taken to imply a shift in perspective of the experiencing self at all. Instead, with depersonalization, patients describe “*not really being there*” in an egocentric sense – but do not claim to occupy any external perspective. This supports our argument here that terms like disembodiment and dissociation require a more considered usage when examining cases of OBEs relative to seemingly similar situations like DPD-DR. It would appear that there has been some equivocation over the use of terms like disembodiment over the years which, in no small way, has contributed to confusion over depersonalization and other ABE.

Sierra’s (2009) salient observation shows that the term “*disembodiment*” has often been taken to describe both; (i) what is, in reality, a reduction in saliency of the embodied sense of self – where one is still embodied (egocentrically), but this is greatly weakened/diluted as well as; (ii) being completely disembodied (exocentrically) into another spatial location (the OBE). Because both these factors can occur together and can be dissociated, we recommend abandoning using the term disembodiment for both cases and those representing the former situation.

The revised taxonomy argued for here would help navigate around such confusion, as the concept of disembodiment would only be used for instances where exocentric perspectives are

³This suggestion assumes that the term “disembodied” be interpreted more literally. By this taxonomy, one cannot have an egocentric disembodied experience, but one can have an egocentric dissociative experience.

⁴Importantly, “*embodied dissociation*” and “*disembodied dissociation*” are not being argued to be absolute – more so that these processes likely co-exist as relative biases – where one dictates and defines anomalous conscious experience at any given time. For example, while primarily disembodied, OBEs can consist of a minor co-awareness of the physical self. Nonetheless, disembodied processes dominate the phenomenology and realism of the experience. Exploring the presence, degree, and interplay of such biases, across different OBEs and associated experiences, will be an exciting avenue for future research.

experienced and dominate consciousness. As a consequence of this redefinition, ABEs described by patients with DPD-DR would not be defined as disembodied – though they are clearly dissociative. In other words, one can be dissociated from the self (i.e., estranged from the self) while not necessarily being disembodied from the self.

One argument against this position might be instances where patients may describe no salient experiential perspective, and instances of heautoscopy – where dual egocentric and exocentric perspectives appear to co-exist, are thus not easily accommodated within this re-description. However, our conception is supported by a clear division in empirical performance at a more objective behavioral task, and not just subjective reports in interviews or via questionnaire measures. In addition, the proposed conception does help to; (i) differentiate many dissociative experiences from a variety of neurological, clinical, and psychotic conditions; (ii) adds clarity to the confusion surrounding the nature of ABE in depersonalization; (iii) more clearly highlights the important differences between ABE in depersonalization and the OBE, and; (iv) implicates the possible presence or absence of certain neural networks (exocentric perspective-taking/self-perspective inhibition). Furthermore, identifying experiences that lie outside of these boundaries is still helpful for the development of scientific theory.

In terms of the actual mechanisms mediating the increased efficiency seen for participants predisposed to OBEs, one may think of these simply as an increased ability in exocentric perspective-taking *per se* (i.e., the ability to simply adopt an external point of view). However although intellectually seductive, to some extent these findings may also index a greater ability to suppress the egocentric point of view. There is growing evidence for the existence of both mechanisms of self-perspective inhibition (Vorauer and Ross, 1999; Ruby and Decety, 2004; Samson et al., 2005; van der Meer et al., 2011) and the excitation of exocentric perspectives (Ruby and Decety, 2001; Saxe et al., 2006; Lambrey et al., 2008; see also Zacks et al., 1999, 2000). These may work in concert to achieve exocentric representations underlying striking and convincing multi-sensory hallucinations of the self like the OBE. Both processes may also enjoy diverse neurocognitive underpinnings. One prediction here is that self-perspective inhibition may not, on its own, be sufficient for an OBE to occur. Under these circumstances, individuals may simply report embodied dissociative experiences (e.g., estrangement from the self or “*not being there*”). The *disembodied dissociative* experiences reported by those having OBEs may require additional, alternative and exocentric representations of the self in space.

Interestingly, uniting these themes into a coherent and more comprehensive account of dissociative experiences might also help illuminate theories of both depersonalization and OBEs. For example, as dissociative ABEs reported in depersonalization appear to be entrenched in egocentric/embodied representations, they might reflect an increased and aberrant weighting of internal bodily experiences (perhaps in an attempt to re-establish the egocentric self which is disintegrating). This aberrant weighting or attentional-shift directed toward internal bodily sensations may itself increase the saliency of internal

and interoceptive body-sensations and thus contribute to some of the embodied ABEs reported by DPD-DR patients. This would also be consistent with the observation that clinical cases of DPD-DR have identified the presence of hyperreflexivity – where some patients can become obsessive and display an aberrant focus on bodily sensations (Medford et al., 2005; Sierra, 2009; Sass et al., 2013). Similar observations have been made in studies showing that OBE groups can also display increased signs of somatoform dissociation/distortion, revolving around a heightened and magnified sense of self and self-consciousness (Murray and Fox, 2005).

Such a shift to internal representations might also contribute to altered experiences of one's own surroundings, as attention and processing would be drawn away from processing salient external signals. Conceivably this might contribute, in part, to the nature of the particular phenomenological characteristics of derealization experiences (e.g., observers feel cut-off/detached from the world). If the observer does not have access to additional biases in exocentric representational systems, then they remain embodied, but dissociated and depersonalized. However, in other cases where aberrant activation in exocentric representations also contribute to the experience, which are also temporarily more stable than disrupted egocentric and embodied representations, then an OBE might be more likely to develop.

The present findings may also speak in some way to the ongoing debate over the concepts of depersonalization and derealization – where it has been argued that “pure” cases of derealization are rare in the clinical literature and thus it may not actually reflect a separate construct (see Sierra et al., 2002). Although our present findings are based only on two of the four measures from the CDS, the current findings do imply a stronger effect for derealization (relative to the ABEs associated more with the construct of depersonalization) in relation to OBEs. This provides some tentative support for the view that derealization experiences may well reflect distinct underlying mechanisms, at least for non-clinical hallucinators.

REMAINING ISSUES

Although there are many variants of perspective-taking tasks in the literature, it is not always clear-cut that the processes required to complete them successfully necessarily recruit exocentric perspective-taking. For example, Braithwaite and Dent (2011) were the first to have questioned these assumptions in relation to the evidence recruited for the standard OBT task used by Blanke and colleagues to examine disruptions in body-transformation/perspective-taking processes (e.g., Blanke et al., 2005; Arzy et al., 2006; Mohr et al., 2006; Easton et al., 2009). One limitation with these earlier incarnations of the OBT task is that it typically recruited only two perspectives in the exemplar stimuli and alternative strategies could easily be developed and used within a block of trials (see also Gardner and Potts, 2011; Gronholm et al., 2012; Kessler and Wang, 2012; May and Wendt, 2012; see also Pezzulo et al., 2013). Although often empirical, it is important to remain aware of the different transformational processes (e.g., perspective-taking, object-rotation, if/then strategies) that may be apparent in a given task (Hegarty and Waller, 2004).

These limitations should also be considered in relation to the current task. In the context of the current debate it is important to ask if; (i) the tasks used can be reasonably assumed to measure rotational processes (either exocentric perspective-taking, or mental rotation); and (ii) that these particular mechanisms are functionally implicated in disembodied dissociative experiences (i.e., the OBE). Although always problematic to separate, some of the current findings do suggest that rotational/transformational processes, more than alternative non-transformational ones, are indeed playing a role in the current task.

For example, the main effects of Viewpoint, the Viewpoint \times Rotation, and Viewpoint \times Group interactions would not be expected from some basic form of if/then rule or similar trial-by-trial strategy. These components should be irrelevant to such rule-based strategies. The current HOBt task used two different body positions, from two different viewpoints, and not just a binary viewpoint manipulation (as has been the case with a number of studies; Blanke et al., 2005; Arzy et al., 2006; Mohr et al., 2006; Easton et al., 2009). So the development of alternative strategies, while not impossible, would have to cope with much greater trial-by-trial unpredictability, impairing basic contingency-based and rule-based strategies. In addition, it is not at all clear how or why such contingency-based strategies could explain the effects of Group also seen in the present findings – unless it is argued that such non-spatial strategies relate to the mechanisms underlying the exocentric OBE in some meaningful way.

Furthermore, previous independent investigations that have used the standard OBT task have reported significant costs to RT performance, not benefits, for both OBE samples (Braithwaite et al., 2011) and those showing elevated signs of perceptual aberrations linked to schizotypy (Mohr et al., 2006). This is in contrast to the large and significant improvements to task efficiency found in the present study. Collectively, these findings suggest that the present HOBt task is both methodologically improved and not equivalent to the performance reported for the more traditional version of the task.

Whether the current task predominantly recruits object-rotation or exocentric perspective-taking in the form of mental self-rotation (e.g., Kessler and Thomson, 2010) remains to be explored with future experimentation. In fact, different conditions of the HOBt task may have triggered different strategies of object- vs. self-rotation. We have argued that the condition with the longest RTs, the “Below – Rotation” condition, may have required an initial rotation of the avatar into a more familiar orientation, which is an example of mental object-rotation, while the subsequent steps in this condition as well as the default transformations in the other three conditions may have consisted in mental self-rotation. This is clearly speculative but could be resolved in future studies making use of posture manipulations. Kessler and colleagues (Kessler and Rutherford, 2010; Kessler and Thomson, 2010; Kessler and Wang, 2012) have recently shown that a body posture that anticipates the direction of mental self-rotation (akin to exocentric perspective-taking) facilitates the transformation, while an incongruent posture delays the process. Importantly, body posture does not affect mental object-rotation (Kessler and Thomson, 2010, Experiment 3). This pattern of results could help in shedding light on the processes engaged by OBE participants

during exocentric perspective-taking (i.e., would they show a posture congruency effect or not?). According to their symptomology of perceiving themselves outside their body, we would expect them to engage in self-rotation/exocentric perspective-taking rather than object-rotation whenever possible, making them the highly efficient perspective takers we observed in the current study.

Finally, the current findings also have important bearings on perspective processing in social interactions. Firstly, an intriguing future research question will be if and how OBE participants make use of their efficient perspective-taking skills during social interaction: are they more inclined to adopt another’s perspective in conversation than control participants or are their perspective-taking skills rather confined to visuo-spatial scenarios and completely independent of a social context? We believe that the latter is unlikely in the light of our current findings. In addition and on an anecdotal point, as part of our research programs into anomalous experiences, we have encountered some participants with social phobias/agoraphobias that have reported learning to consciously will and “use” the disembodied viewpoint of the OBE to manage stressful social situations. Here the experience makes the observer feel removed from the direct social context causing stressful reactions.

By moving away from the schematic drawings of the classic OBT task (which have produced mixed results and might not index exocentric perspective-taking) toward more naturalistic photographs of bodies in more varied postures in space, we have enhanced the task’s social dimension especially since these changes have increased the likelihood that motor resonance mechanisms are engaged in order to process difficult body postures (cf. Kessler and Miellet, 2013).

In social interaction, the latter often takes on the form of implicit mimicry, i.e., the so-called “chameleon effect,” which has been shown to enhance pro-social behavior and attitudes (e.g., Chartrand and Bargh, 1999; van Baaren et al., 2009; for review, see Niedenthal et al., 2005). Furthermore, direct effects of posture, posture resonance, and other body-related processes on the speed of egocentric transformations have been recently shown by Kessler and Thomson (2010, especially Experiment 4) and others (e.g., Lenggenhager et al., 2008; Falconer and Mast, 2012; van Elk and Blanke, 2013). Therefore, investigating embodied perspective-taking during realistic social interaction in relation to dissociative traits (e.g., embodied vs. disembodied dissociative experiences) could be a somewhat contra-intuitive, yet highly interesting addition to the field of social cognitive neuroscience.

CONCLUSION

The present study investigated biases in perspective-taking processes that may be implicated in predisposition to hallucinatory experiences that involve a shift in self-perspective (the OBE). The OBE group were much more efficient at a perspective-taking task relative to a control group – supporting the view that the prevalence of the OBE is associated with biases in perspective-taking ability. In addition, the OBE group displayed significantly more signs of derealization experiences – which we speculate may underlie a propensity to experience ambiguous sensory information from the outside world and may contribute to destabilize the

typically coherent sense of self. The current findings also support a fractionating of the unitary notion of dissociation relative to whether embodied or disembodied dissociative experiences are reported. Future studies are planned to investigate the role of both self-perspective inhibition and exocentric perspective-taking underlying these and other related ABEs.

ACKNOWLEDGMENTS

The present study was supported by a Bursary grant awarded from The Bial Foundation (#01/10) to the primary author. We gratefully thank the foundation for its generous support of our research. We thank Emma Brogna for pilot work on the present stimuli and our OBE group for coming forward with their experiences. This project was carried out at the primary authors Selective Attention and Awareness Laboratory at the University of Birmingham, UK.

REFERENCES

- Ackner, B. (1954). Depersonalization: I. Aetiology and phenomenology. *J. Ment. Sci.* 100, 838–853. doi:10.1192/bjp.100.421.838
- Arzy, S., Thut, G., Mohr, C., Michel, C. M., and Blanke, O. (2006). Neural basis of embodiment: distinct contributions of temporoparietal junction and extrastriate body area. *J. Neurosci.* 26, 8074–8081. doi: 10.1523/JNEUROSCI.0745-06.2006
- Blackmore, S. J. (1982). *Beyond the Body: An Investigation into Out-of-the-Body Experiences*. London: Heinemann.
- Blackmore, S. J. (1987). Where am I? Perspectives in imagery and the out-of-body experience. *J. Ment. Imagery* 11, 53–66.
- Blanke, O., and Arzy, S. (2005). The out-of-body experience. Disturbed self-processing at the temporal-parietal junction. *Neuroscientist* 11, 16–24. doi: 10.1177/1073858404270885
- Blanke, O., and Metzinger, T. (2009). Full-body illusions and minimal phenomenal selfhood. *Trends Cogn. Sci.* 13, 7–13. doi: 10.1016/j.tics.2008.10.003
- Blanke, O., and Mohr, C. (2005). Out-of-body experience, heautoscopy, and autoscopic hallucination of neurological origin. Implications for neurocognitive mechanisms of corporeal awareness and self-consciousness. *Brain Res. Rev.* 50, 184–199. doi: 10.1016/j.brainresrev.2005.05.008
- Blanke, O., Mohr, C., Michel, C. M., Pascual-leone, A., Brugger, P., Seeck, M., et al. (2005). Linking out-of-body experience and self-processing to mental own-body imagery at the temporoparietal junction. *J. Neurosci.* 19, 550–557. doi: 10.1523/JNEUROSCI.2612-04.2005
- Braithwaite, J. J., Brogna, E., Bagshaw, A. P., and Wilkins, A. J. (2013a). Evidence for elevated cortical hyperexcitability and its association with out-of-body experiences in the non-clinical population: new findings from a pattern-glare task. *Cortex* 49, 793–805. doi: 10.1016/j.cortex.2011.11.013
- Braithwaite, J. J., Brogna, E., Brincat, O., Stapley, L., Wilkins, A. J., and Takahashi, C. (2013b). Signs of increased cortical hyperexcitability selectively associated with spontaneous anomalous bodily experiences in a nonclinical population. *Cogn. Neuropsychiatry* doi: 10.1080/13546805.2013.768176 [Epub ahead of print].
- Braithwaite, J. J., and Dent, K. (2011). New perspectives on perspective-taking mechanisms and the out-of-body experience. *Cortex* 47, 628–632. doi: 10.1016/j.cortex.2010.11.008
- Braithwaite, J. J., Samson, D., Apperly, I., Brogna, E., and Hulleman, J. (2011). Cognitive correlates of the spontaneous out-of-body experience (OBE) in the psychologically normal population: evidence for an increased role of temporal-lobe instability, body-distortion processing, and impairments in own-body transformations. *Cortex* 47, 839–853. doi: 10.1016/j.cortex.2010.05.002
- Brugger, P. (2002). Reflective mirrors: perspective-taking in autoscopic phenomena. *Cogn. Neuropsychiatry* 7, 179–194. doi: 10.1080/135468002440000076
- Brugger, P., Regard, M., and Landis, T. (1997). Illusory reduplication of one's own body: phenomenology and classification of autoscopic phenomena. *Cogn. Neuropsychiatry* 2, 19–38. doi: 10.1080/135468097396397
- Carpenter, M., and Proffitt, D. R. (2001). Comparing viewer and array mental rotations in different planes. *Mem. Cogn.* 29, 441–448. doi: 10.3758/BF03196395
- Chartrand, T. L., and Bargh, J. A. (1999). The chameleon effect: the perception-behavior link and social interaction. *J. Pers. Soc. Psychol.* 76, 893–910. doi: 10.1037/0022-3514.76.6.893
- Cook, A. M., and Irwin, H. J. (1983). Visuospatial skills and the out-of-body experience. *J. Parapsychol.* 47, 23–35.
- Creem, S. H., Wraga, M., and Proffitt, D. R. (2001). Imagining physically impossible self-rotations: geometry is more important than gravity. *Cognition* 81, 41–64. doi: 10.1016/S0010-0277(01)00118-4
- Eastman, M. (1962). Out-of-the-body experiences. *Proc. Soc. Psych. Res.* 53, 287–309.
- Easton, S., Blanke, O., and Mohr, C. (2009). A putative implication for fronto-parietal connectivity in out-of-body experiences. *Cortex* 45, 216–227. doi: 10.1016/j.cortex.2007.07.012
- Falconer, C. J., and Mast, F. W. (2012). Balancing the mind: vestibular induced facilitation of egocentric mental transformations. *Exp. Psychol.* 59, 332–339. doi: 10.1027/1618-3169/a000161
- Gabbard, G. O., and Twemlow, S. W. (1984). *With the Eyes of the Mind*. New York: Praeger. doi:10.1007/BF01073367
- Gabbard, G. O., and Twemlow, S. W. (1986). An overview of altered mind/body perception. *Bull. Menninger Clin.* 50, 351–366.
- Gabbard, G. O., Twemlow, S. W., and Jones, F. C. (1981). Do “near death experiences” occur only near death? *J. Nerv. Ment. Dis.* 169, 374–377. doi: 10.1097/00005053-198106000-00006
- Gabbard, G. O., Twemlow, S. W., and Jones, F. C. (1982). Differential diagnosis of altered mind/body perception. *Psychiatry* 45, 361–369.
- Gardner, M., and Potts, R. (2011). Domain general mechanisms account for imagined transformations of whole body perspective. *Acta Psychol.* 137, 371–381. doi: 10.1016/j.actpsy.2011.04.006
- Grabowski, J. (1999). A uniform anthropomorphological approach to the human conception of dimensional relations. *Spat. Cogn. Comput.* 1, 349–363. doi: 10.1023/A:1010031428440
- Green, C. E. (1968). *Out-of-the-Body Experiences*. London: Hamish Hamilton. doi:10.1007/BF01073268
- Gronholm, P. C., Flynn, M., Edmonds, C. J., and Gardner, M. R. (2012). Empathic and non-empathic routes to visuospatial perspective-taking. *Conscious. Cogn.* 21, 494–500. doi: 10.1016/j.concog.2011.12.004
- Hegarty, M., and Waller, D. A. (2004). Dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence* 32, 175–191. doi: 10.1016/j.intell.2003.12.001
- Irwin, H. J. (1985). *Flight of Mind: A Psychological Study of the Out-of-Body Experience*. Metuchen, NJ: Scarecrow Press.
- Irwin, H. J. (2000). The disembodied self: an empirical study of dissociation and the out-of-body experience. *J. Parapsychol.* 64, 261–276.
- Kessler, K., and Miellet, S. (2013). Perceiving conspecifics as integrated body-gestalts is an embodied process. *J. Exp. Psychol. Gen.* 142, 774–790. doi: 10.1037/a0029617
- Kessler, K., and Rutherford, H. (2010). The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial preposition. *Front. Psychol.* 1:213. doi: 10.3389/fpsyg.2010.00213
- Kessler, K., and Thomson, L. A. (2010). The embodied nature of spatial perspective taking: embodied transformation versus sensorimotor interference. *Cognition* 114, 72–88. doi: 10.1016/j.cognition.2009.08.015
- Kessler, K., and Wang, H. (2012). Spatial perspective taking is an embodied process, but not for everyone in the same way: differences for gender and social skills score. *Spat. Cogn. Comput.* 12, 133–158. doi: 10.1080/13875868.2011.634533
- Lambrey, S., Amorim, M. A., Samson, S., Noulhiane, M., Hasboun, D., Dupony, S., et al. (2008). Distinct visual perspective-taking strategies involve the left and right medial temporal lobe structures differently. *Brain* 131, 523–534. doi: 10.1093/brain/awn317
- Lenggenhager, B., Lopez, C., and Blanke, O. (2008). Influence of galvanic vestibular stimulation on egocentric and object-based mental transformations. *Exp. Brain Res.* 184, 211–221. doi: 10.1007/s00221-007-1095-9
- May, M., and Wendt, M. (2012). Separating imagined spatial transformations and spatial compatibility effects in the own body transformation task. *Cogn. Process.* 13, 257–260. doi: 10.1007/s10339-012-0455-y
- Medford, N., Sierra, M., Baker, D., and David, A. S. (2005). Understanding and treating depersonalisation disorder. *Adv. Psychiatr. Treat.* 11, 92–100. doi: 10.1192/apt.11.2.92

- Mohr, C., Blanke, O., and Brugger, P. (2006). Perceptual aberrations impair mental own-body transformations. *Behav. Neurosci.* 120, 528–534. doi: 10.1037/0735-7044.120.3.528
- Murray, C. D., and Fox, J. (2005). Dissociational body experiences: differences between respondents with and without prior out-of-body-experiences. *Br. J. Psychol.* 96, 441–456. doi: 10.1348/000712605X49169
- Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth-Gruber, S., and Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Pers. Soc. Psychol. Rev.* 9, 184–211. doi: 10.1207/s15327957pspr0903_1
- Noyes, R., and Kletti, R. (1976). Depersonalization in the face of life threatening danger: a description. *Psychiatry* 39, 19–27.
- Noyes, R., and Kletti, R. (1977). Depersonalization in response to life threatening danger. *Compr. Psychiatry* 18, 375–384. doi: 10.1016/0010-440X(77)90010-4
- Overney, L. S., Arzy, S., and Blanke, O. (2009). Deficient mental own-body imagery in a neurological patient with out-of-body experiences due to cannabis use. *Cortex* 45, 228–235. doi: 10.1016/j.cortex.2008.02.005
- Palmer, J. (1978). The out-of-body experience: a psychological theory. *Parapsychol. Rev.* 9, 19–22.
- Pezzulo, G., Barsalou, L., Cangelosi, A., Fischer, M., McRae, K., and Spivey, M. (2013). Computational grounded cognition: a new alliance between grounded cognition and computational modeling. *Front. Psychol.* 3:612. doi: 10.3389/fpsyg.2012.00612
- Popescu, S. T., and Wexler, M. (2012). Spontaneous body movements in spatial cognition. *Front. Psychol.* 3:136. doi: 10.3389/fpsyg.2012.00136
- Rach, S., Diedrich, A., and Colonius, H. (2011). On quantifying multisensory interaction effects in reaction time and detection rate. *Psychol. Res.* 75, 77–94. doi: 10.1007/s00426-010-0289-0
- Ruby, P., and Decety, J. (2001). Effect of subjective perspective taking during simulation of action: a PET investigation of agency. *Nat. Neurosci.* 4, 546–550. doi:10.1038/87510
- Ruby, P., and Decety, J. (2004). How do you feel versus how do you think she would feel? A neuroimaging study of perspective taking with social emotions. *J. Cogn. Neurosci.* 16, 988–999. doi: 10.1162/0898929041502661
- Samson, D., Apperly, I. A., Kathirgamanathan, U., and Humphreys, G. W. (2005). Seeing it my way: a case of a selective deficit in inhibiting self-perspective. *Brain* 128, 1102–1111. doi: 10.1093/brain/awh464
- Sass, L., Pienkos, E., Nelson, B., and Medford, N. (2013). Anomalous self-experience in depersonalization and schizophrenia: a comparative investigation. *Conscious. Cogn.* 22, 430–441. doi: 10.1016/j.concog.2013.01.009
- Saxe, R., Jamal, N., and Powell, L. (2006). My body or yours? The effect of visual perspective on cortical body representations. *Cereb. Cortex* 16, 178–182. doi: 10.1093/cercor/bhi095
- Schwabe, L., Lenggenger, B., and Blanke, O. (2009). The timing of temporoparietal and frontal activations during mental own body transformations from different visuospatial perspectives. *Hum. Brain Mapp.* 30, 1801–1812. doi: 10.1002/hbm.20764
- Sierra, M. (2009). *Depersonalization: A New Look at a Neglected Syndrome*. Cambridge: Cambridge University Press.
- Sierra, M., Baker, D., Medford, N., and David, A. S. (2005). Unpacking the depersonalization syndrome: an exploratory factor analysis on the Cambridge Depersonalization Scale. *Psychol. Med.* 35, 1523–1532. doi: 10.1017/S0033291705005325
- Sierra, M., and Berrios, G. E. (1997). Depersonalization: a conceptual history. *Hist. Psychiatry* 8, 213–229. doi: 10.1177/0957154X9700803002
- Sierra, M., and Berrios, G. E. (2000). The Cambridge Depersonalization Scale: a new instrument for the measurement of depersonalization. *Psychiatry Res.* 93, 153–164. doi: 10.1016/S0165-1781(00)00100-1
- Sierra, M., and David, A. S. (2011). Depersonalization: a selective impairment of self-awareness. *Conscious. Cogn.* 20, 99–108. doi: 10.1016/j.concog.2010.10.018
- Sierra, M., Lopera, F., Lambert, M. V., Phillips, M. L., and David, A. S. (2002). Separating depersonalisation and derealisation: the relevance of the “lesion method”. *J. Neurol. Neurosurg. Psychiatry* 72, 530–532. doi:10.1136/jnnp.72.4.530
- Simeon, D., Kozin, D. S., Segal, K., Lerch, B., Dujour, R., and Biesbrecht, T. (2008). De-constructing depersonalization: further evidence for symptom clusters. *Psychiatry Res.* 157, 303–306. doi: 10.1016/j.psychres.2007.07.007
- Townsend, J. T., and Ashby, F. G. (1983). *The Stochastic Modeling of Elementary Psychological Processes*. Cambridge: Cambridge University Press.
- Twemlow, S. W., Gabbard, G. O., and Jones, F. C. (1982). The out-of-body experience: a phenomenological typology based on questionnaire responses. *Am. J. Psychiatry* 139, 450–455.
- van Baaren, R., Janssen, L., Chartrand, T. L., and Dijksterhuis, A. (2009). Where is the love? The social aspects of mimicry. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 2381–2389. doi: 10.1098/rstb.2009.0057
- van der Meer, L., Groenewold, N. A., Nolen, W. A., Pijnenborg, M., and Aleman, A. (2011). Inhibit yourself and understand the other: neural basis of distinct processes underlying Theory of Mind. *Neuroimage* 56, 2364–2374. doi: 10.1016/j.neuroimage.2011.03.053
- van Elk, M., and Blanke, O. (2013). Imagined own-body transformations during passive self-motion. *Psychol. Res.* doi: 10.1007/s00426-013-0486-8 [Epub ahead of print].
- Vorauer, J. D., and Ross, M. (1999). Self-awareness and feeling transparent: failing to suppress one's self. *J. Exp. Soc. Psychol.* 35, 415–440. doi: 10.1006/jesp.1999.1388
- Zacks, J., Rypma, B., Gabrieli, J. D. E., Tversky, B., and Glover, G. H. (1999). Imagined transformations of bodies: an fMRI investigation. *Neuropsychologia* 37, 1029–1040. doi: 10.1016/S0028-3932(99)00012-3
- Zacks, J. M., Mires, J., Tversky, B., and Hazeltine, E. (2000). Mental spatial transformations of objects and perspective. *Spat. Cogn. Comput.* 2, 315–332. doi: 10.1023/A:1015584100204

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 26 June 2013; accepted: 10 October 2013; published online: 30 October 2013.

Citation: Braithwaite JJ, James K, Dewe H, Medford N, Takahashi C and Kessler K (2013) Fractionating the unitary notion of dissociation: disembodied but not embodied dissociative experiences are associated with exocentric perspective-taking. *Front. Hum. Neurosci.* 7:719. doi: 10.3389/fnhum.2013.00719

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2013 Braithwaite, James, Dewe, Medford, Takahashi and Kessler. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Strategy modulates spatial perspective-taking: evidence for dissociable disembodied and embodied routes

Mark R. Gardner^{1*}, Mark Brazier¹, Caroline J. Edmonds² and Petra C. Gronholm³

¹ Department of Psychology, University of Westminster, London, UK

² School of Psychology, University of East London, London, UK

³ Health Service and Population Research Department, Institute of Psychiatry, King's College London, London, UK

Edited by:

Klaus Kessler, University of Glasgow, UK

Reviewed by:

Nicole David, University Medical Center Hamburg-Eppendorf, Germany
Andrew David Ridley Surtees, Université catholique de Louvain, Belgium

*Correspondence:

Mark R. Gardner, Department of Psychology, University of Westminster, 309 Regent Street, W1B 2UW, London, UK
e-mail: m.gardner@westminster.ac.uk

Previous research provides evidence for a dissociable embodied route to spatial perspective-taking that is under strategic control. The present experiment investigated further the influence of strategy on spatial perspective-taking by assessing whether participants may also elect to employ a separable “disembodied” route loading on inhibitory control mechanisms. Participants ($N = 92$) undertook both the “own body transformation” (OBT) perspective-taking task, requiring speeded spatial judgments made from the perspective of an observed figure, and a control task measuring ability to inhibit spatially compatible responses in the absence of a figure. Perspective-taking performance was found to be related to performance on the response inhibition control task, in that participants who tended to take longer to adopt a new perspective also tended to show a greater elevation in response times when inhibiting spatially compatible responses. This relationship was restricted to those participants reporting that they adopted the perspective of another by reversing left and right whenever confronted with a front-view figure; it was absent in those participants who reported perspective-taking by mentally transforming their spatial orientation to align with that of the figure. Combined with previously published results, these findings complete a double dissociation between embodied and disembodied routes to spatial perspective-taking, implying that spatial perspective-taking is subject to modulation by strategy, and suggesting that embodied routes to perspective-taking may place minimal demands on domain general executive functions.

Keywords: perspective-taking, own body transformation, strategy, embodiment, social, response inhibition

INTRODUCTION

Spatial perspective-taking underlies successful social interactions (Tversky and Hard, 2009), for instance when giving directions or demonstrating how to perform a task. Furthermore, spatial perspective-taking itself may be an intrinsically social process, when the novel perspective one adopts is that occupied by another person, rather than a position external to that occupied by any other body (Stocker, 2012). Although it has been well established that qualitatively different underlying processes subserve different kinds of perspective-taking (e.g., Michelon and Zacks, 2006; David et al., 2008; Cohen et al., 2009), the manner in which these various perspective-taking mechanisms rely on “embodied” cognitions such as the mental simulation of body movements has yet to be fully specified—despite this being an active line of enquiry (e.g., Kessler and Rutherford, 2010; Kessler and Thomson, 2010). Investigation of embodied perspective-taking may help to elucidate how the spatial and social domains impinge upon perspective-taking. Outstanding issues include identifying the types of perspective-taking that are possible via a “disembodied” route that engages response inhibition rather than motor simulation or social processes, as well

as the role played by endogenous control processes in selecting between multiple perspective-taking routes. Consequently, the aim of the current study was to examine how these two types of executive processes influence perspective-taking, by assessing whether the strategy that participants report using moderates the relationship between perspective-taking ability in the “own body transformation” (OBT) task (e.g., Zacks et al., 1999; Blanke et al., 2005; Mohr et al., 2010) and ability to perform a control task indexing disembodied response inhibition processes.

The prevailing view is that spatial perspective-taking via imagined transformations of one's own egocentric perspective is an embodied process (e.g., Kessler and Rutherford, 2010; Kessler and Thomson, 2010), in the sense that it is performed via mental simulation of the sensorimotor mechanisms involved in actual self rotation (Lenggenhager et al., 2008). The finding that the speed and accuracy of taking another's viewpoint depends upon the degree of angular disparity between one's own and a target's frame of reference provides evidence for an analogue transformational process sharing at least some of the properties of self motion (Zacks and Michelon, 2005). Support for the involvement of deliberate motor simulation is provided by reports

that postural congruence between participants and targets facilitates perspective-taking performance (Kessler and Rutherford, 2010; Kessler and Thomson, 2010). In addition, an individual's motor capability appears to modulate the extent that motor simulation is engaged in perspective-taking. For instance, skill at performing rotational movements has been found to facilitate perspective-taking during a mental body rotation task (Steggemann et al., 2011), and attentional biases associated with participants' own handedness have been found to extend to left-right judgments made from a schematic figure's perspective in the OBT task (Gardner and Potts, 2010). Furthermore, patients with left spatial neglect have even been found to recover information that is unavailable from an egocentric perspective when space is imagined from an opposite perspective (Becchio et al., 2013). Of particular relevance, the degree of amelioration of the neglected side is greatest in an embodied condition when a person is seen to be present in the novel perspective. These findings provide converging evidence for embodied processes contributing to perspective-taking.

Nonetheless, under certain circumstances disembodied processes appear sufficient to account for perspective-taking. For instance, determining which objects can be *seen* from another person's perspective appears to involve line-of-sight computation without the need for transformations of one's own perspective (Michelon and Zacks, 2006). The determination of spatial relationships relative to a third party perspective within the OBT task has also been accounted for in terms of domain general response selection processes and spatial compatibility, either alone (Gardner and Potts, 2011), or in combination with imagined perspective transformations (May and Wendt, 2012). In these cases, a conflict arises between information coded relative to one's own bodily position and information coded for the adopted perspective (May, 2004; Michelin and Zacks, 2006). Thus, the cognitive demands of perspective-taking are at least in part due to the need to inhibit prepotent responses relating to one's own perspective (cf Leslie et al., 2005). In support of this view, the ability to adopt a third party perspective has been shown to be disrupted when performed alongside a secondary task loading on response inhibition processes (Qureshi et al., 2010). Thus, "disembodied" executive functions, including response inhibition, may contribute to perspective-taking alongside, or in place of, a cognitively efficient "embodied" route.

One possibility is that separable embodied and disembodied perspective-taking processes (May and Wendt, 2012), may in fact be distinct routes to perspective-taking controlled by higher level strategy (Gronholm et al., 2012). Although many have proposed that utilization of different strategies could explain variation in perspective-taking performance (Michelon and Zacks, 2006; Thakkar et al., 2009; Mohr et al., 2010; Thakkar and Park, 2010), the role of strategy has rarely been considered explicitly (Amorim, 2003). Initial evidence has been reported indicating the presence of a dissociable strategy associated with embodied perspective-taking (Lenggenhager et al., 2008; Gronholm et al., 2012). For instance, the disruption to mental transformations arising from galvanic vestibular stimulation was found to be restricted to participants reporting that they had employed transformations

of their own perspective, rather than an object based strategy (Lenggenhager et al., 2008). Using the OBT task, Gronholm et al. (2012) found a selective association between trait level empathy and perspective-taking ability that was restricted to participants using an embodied perspective transformation strategy, as opposed to disembodied strategy of reversing left and right whenever confronted with a front-view figure (Gronholm et al., 2012). This finding is consistent with mental simulation playing a common role for embodied spatial perspective-taking as well as social processes such as empathy and Theory of Mind (Ruby and Decety, 2004). However, to date, no equivalent independent evidence appears to be available for disembodied perspective-taking strategies.

The current study was designed to assess further the influence of strategy on perspective-taking in the OBT-task, by examining whether the strategy that participants report using moderates the relationship between perspective-taking and response inhibition abilities. Previous work using an individual differences approach has found that perspective taking is associated with response inhibition ability (Qureshi, 2008). Here, we examine whether this association is strategy-specific. In the present study, participants undertook both the OBT perspective-taking task, requiring speeded spatial judgments made from the perspective of an observed figure, and the "Transpose" task, a disembodied control task measuring ability to inhibit spatially compatible responses. We predicted that if there are dissociable embodied and disembodied routes to spatial perspective-taking that are modulated by high level strategy, then self-reported strategy should moderate the relationship between performance on the OBT and Transpose tasks. Specifically, we predicted a positive relationship between performance for the OBT and Transpose tasks that would be restricted to those who reported that they adopted the perspective of another by transposing left and right whenever confronted with a front-view figure; no association was predicted for those who in an embodied manner mentally transformed their perspective to align with that of the external figure.

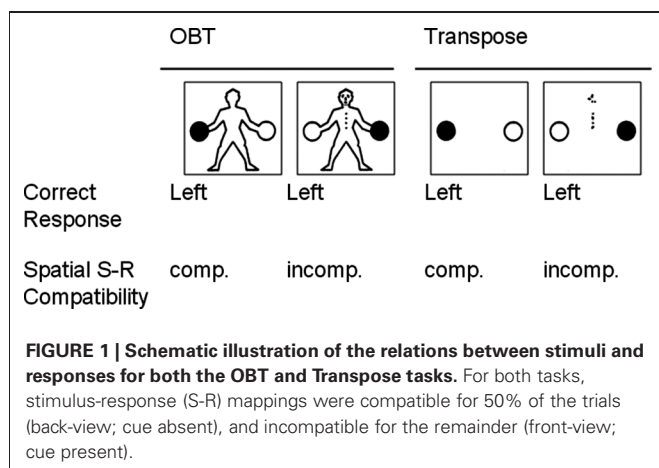
MATERIALS AND METHODS

PARTICIPANTS

Ninety two volunteers (69 female, 23 male), recruited primarily from the university community, took part in the study. Their ages ranged from 19 to 66 years (mean \pm SD = 24.4 \pm 9.4 years). All had normal, or corrected to normal, vision, and provided informed consent in accordance with the local (University of Westminster) ethics approval.

OBT TASK

The OBT task was adapted from that reported previously (Gardner and Potts, 2011, Experiment 1A), as summarized below. Four basic stimuli each depicting a schematic human figure holding a black ball in one hand and a white ball in the other, were presented to participants. The figure could be seen either from front- or back-view, and held the black ball in either the left or right hand (see **Figure 1**, illustrating left hand stimuli). The outline shape of the figure was identical whether it was front- or back-facing. Consequently, the only aspects of the stimulus indi-



cating that the figure was front-facing were the marks indicating the buttons and facial features.

Participants were verbally instructed to imagine taking the perspective of the figure through an embodied mental transformation in order to make a spatial judgment as to which hand the figure was holding the black ball. Standardized instructions to this effect were also delivered via the E-prime program. Each participant was required to rest their index fingers on the response keys (left index finger on the “A” key for a “left” response on a QWERTY keyboard, and right finger on the “L” key for a “right” response). This resulted in S-R mappings that were spatially compatible for 50% of the trials (back-view) and spatially incompatible for the remainder (front-view; see Figure 1).

TRANSPOSE TASK

The Transpose task (Gardner and Potts, 2011) served as a disembodied control task measuring ability to inhibit spatially compatible responses. The stimuli consisted of two balls, one black and one white, in identical locations to those appearing in the OBT task, but in this case presented without a human figure holding them. The black ball could appear on the left or the right, and was presented either alone, in the cue-absent condition, or accompanied by an abstract visual cue, in the cue-present condition. This abstract visual cue consisted of the features that made up the OBT figure’s face and buttons presented in a scrambled configuration. Thus, these stimuli served as non-embodied variants of those employed for the OBT task (see Figure 1).

Participants were instructed to report the location of the black ball from their own viewing perspective by pressing the corresponding key when the abstract visual cue was absent. On trials in which the cue was present, participants were required to transpose left and right when responding (e.g., if the black ball was on the right, the correct response was to press the left response key). Thus, just as in the OBT task, the mapping between stimulus location and response was spatially incompatible for 50% of the trials, and these trials were signaled by equivalent visual information—those marks distinguishing front- and back-view stimuli in the OBT task. The Transpose task should thus place similar demands on response inhibition processes as the

OBT task for those participants adopting a transposing strategy, given that it is operationalized in a comparable manner.

PROCEDURE

All participants performed the OBT task, followed by the Transpose task. The order of these tasks was not counterbalanced in order to prevent expected carry-over to the strategy adopted for the OBT task if participants had experienced the Transpose task first. On each trial, a central black fixation cross was presented for 1400 ms against a white background. This was immediately followed by the stimulus which was displayed for 2100 ms, or until a response had been made. This was followed by visual feedback on whether the response was correct or incorrect, presented for 1500 ms. On any given trial, the stimulus was randomly displaced in the picture plane (range of -50° to $+50^\circ$, in 10° intervals) to introduce further variability in the stimulus set. Each task comprised 132 trials split into two equal blocks, allowing all stimulus combinations to be presented on three occasions in a random order [left vs. right (2) x compatible vs. incompatible (2) x picture plane orientation (11)]. Stimulus presentation and data collection were controlled by a personal computer running E-Prime experiment generator software (Schneider et al., 2002).

Immediately after these tasks were completed participants were asked to report on the strategy they had used during the OBT task, based on which they were categorized into “perspective transformers” or “spatial transposers” in accordance with earlier work (Gronholm et al., 2012). This was intended to discriminate strategies on the basis of embodiment. Those who reported to have always/usually used the “flipping left and right strategy” were classified as (disembodied) spatial transposers, whereas those who always/usually “imagined myself taking the figure’s position” were classified as (embodied) perspective transformers. Participants were classified as perspective transformers also if they reported having used both strategies equally often.

RESULTS

Participants were excluded from the analysis due to an error rate (ER) of above 15% on either the Transpose task ($N = 5$, all female), or the OBT task ($N = 14$, 12 female). The sample that was subjected to analysis thus comprised 73 participants (51 female). In order to measure the relative increase in response times for the incompatible versus compatible condition, a “Composite response time (RT)” for both tasks was computed for each participant according to the formula: Composite RT = (incompatible RT—compatible RT)/compatible RT—see Gronholm et al. (2012). Shapiro-Wilks test indicated these data to be normally distributed: OBT task, $W = 0.973$, $p = .124$; Transpose tasks, $W = 0.989$, $p = .793$.

REPORTED STRATEGY USE, AND PERFORMANCE ON PERSPECTIVE-TAKING AND RESPONSE INHIBITION TASKS

According to self-report, for the OBT task 43 participants (59%) adopted the disembodied transposing strategy and 29 (40%) adopted the putatively embodied perspective transformation strategy. Data on strategy use was unavailable for one further participant. The difference between these proportions was not statistically significant, $p = .125$, binomial test. By adopting

the same classification criterion as Gronholm et al. (2012), the embodied perspective transformation subgroup included 11 participants (38%) that reported having used both strategies. The strategy subgroups were not found to differ in terms of gender distribution, $\chi^2 = 1.69$, $p = .194$, nor age, $t(69) = 0.01$, $p = .992$ (embodied perspective transformation: 79% female, age (mean \pm SD) = 24 ± 9.2 yrs; disembodied transposing strategy: 65% female, age = 24 ± 9.5 yrs).

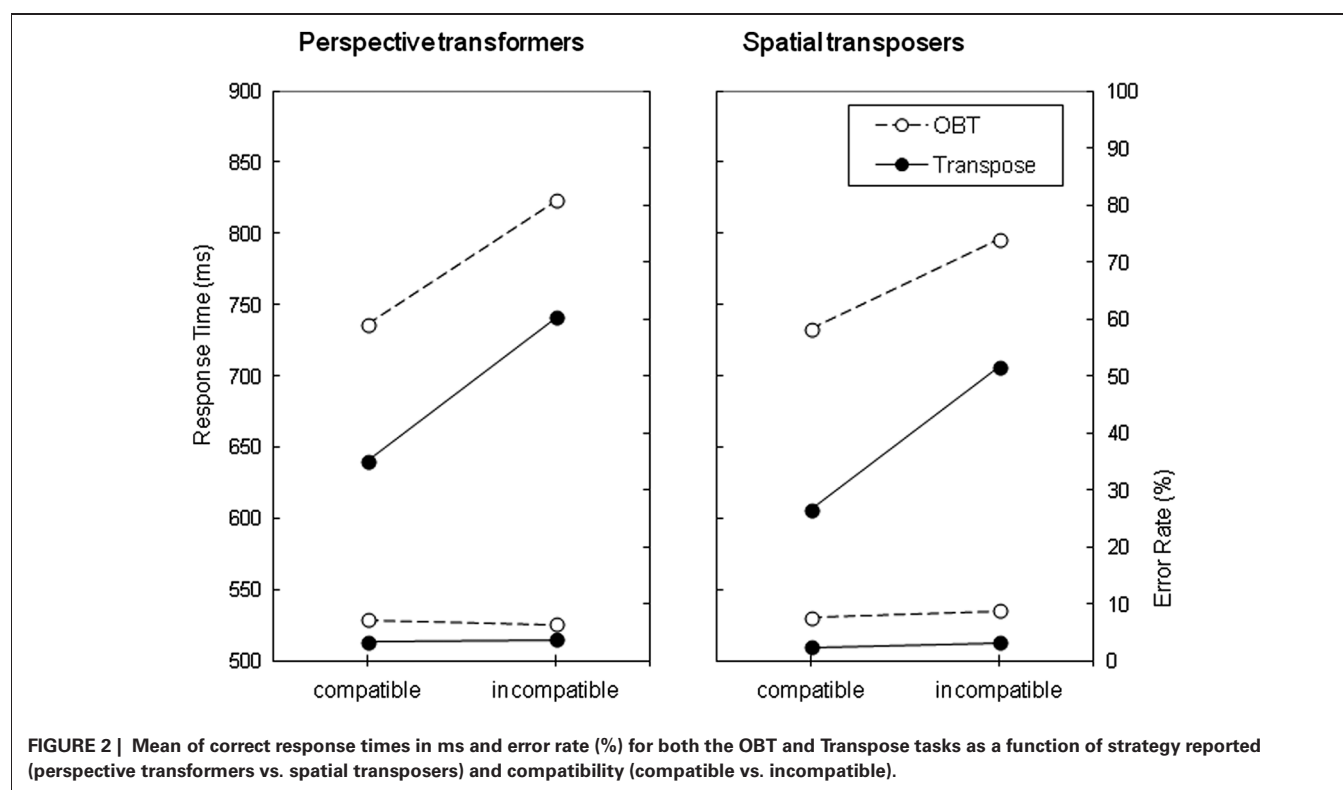
Figure 2 illustrates RT and ER performance on both the OBT and Transpose tasks categorized by the strategy reported, and the S-R compatibility of the stimuli. RTs appeared to be longer for the OBT task than the Transpose task, and longer for the incompatible condition relative to the compatible condition, irrespective of the strategy reported. These impressions were confirmed by a 3-way mixed model Analysis of Variance (ANOVA) where Task (OBT vs. Transpose) and Compatibility (compatible vs. incompatible) were within subject factors and Strategy (perspective transformers vs. spatial transposers) was a between subject factor. This revealed main effects of Compatibility, $F(1,70) = 161$, $p < .001$ and Task, $F(1,70) = 70.1$, $p < .001$, neither of which interacted with Strategy, $F_s < 1$. Furthermore, the main effect of Strategy was not significant, $F < 1$. An interaction between Task and Compatibility was found, $F(1,70) = 5.48$, $p < .022$, consistent with a higher elevation of response times for the incompatible relative to the compatible condition in the Transpose task (mean \pm SD = $17 \pm 12\%$) than in the OBT task ($11 \pm 11\%$). This phenomenon also did not interact with Strategy, $F(1,70) = 1.017$, $p = .317$.

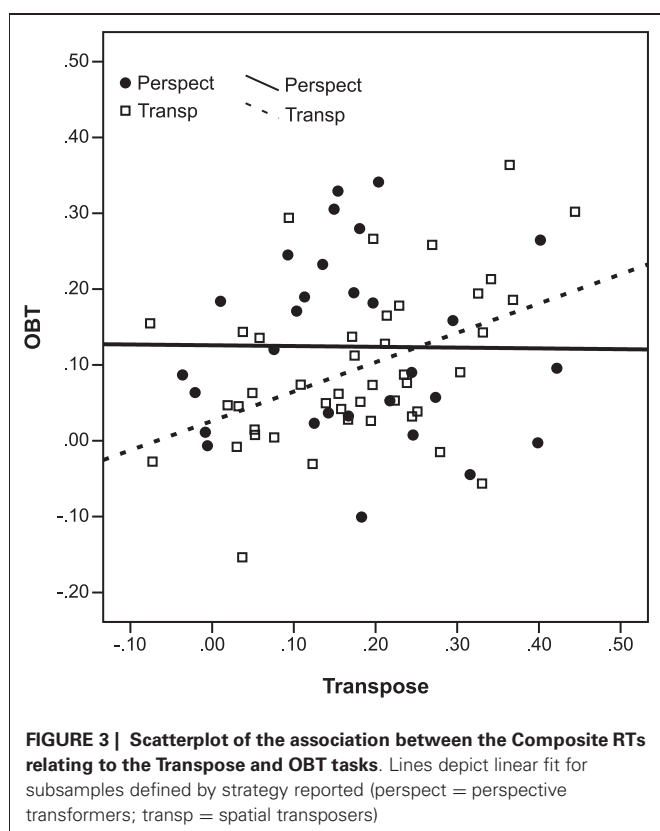
An equivalent 3-way ANOVA was also performed on the ER data depicted in **Figure 2**. This revealed that ER was higher for the

OBT task (mean \pm SD = $7.5 \pm 4.1\%$) than the Transpose task ($3.1 \pm 2.8\%$), $F(1,70) = 68.2$, $p < .001$. However neither the main effect for Compatibility, $F < 1$, nor that for Strategy, $F < 1$, were statistically significant. Strategy was found to moderate the size of the Task effect, $F(1,70) = 4.04$, $p < .048$. The degree to which participants showed greater accuracy for the Transpose compared with OBT task was slightly greater for those reporting having adopted the disembodied spatial transposing strategy (difference in ER, mean \pm SD = $5.4 \pm 4.4\%$, $t(42) = 8.1$, $p < .001$), than for those reporting having adopting a perspective transformation strategy, ($3.3 \pm 4.4\%$, $t(28) = 4.0$, $p < .001$). No other interactions were statistically significant.

STRATEGY AND THE RELATIONSHIPS BETWEEN PERSPECTIVE-TAKING AND RESPONSE-INHIBITION ABILITIES

We examined whether self-reported strategy moderated the relationships between perspective-taking and response inhibition abilities by assessing correlations both within subgroups employing each type of strategy and collapsed across these subgroups, see **Figure 3**. When strategy was disregarded ($N = 73$), a positive relationship was found between performance on the OBT and Transpose tasks as measured by Composite RT, $r = .245$, $p = .036$. When the correlations were repeated within subgroups, the relationship between perspective-taking and response inhibition as measured by the OBT and Transpose tasks was found to be moderated by strategy. For the subgroup that reported having employed the disembodied spatial transposing strategy ($N = 43$), a highly significant positive correlation was found, $r = .449$, $p = .003$. Whereas, for the subgroup that





reported having employed an embodied perspective transformation strategy ($N = 29$), there was no correlation between these tasks, $r = -0.011$, $p = .956$. Nor were the tasks correlated when the 11 participants that reported having used both strategies were removed from the perspective transformation subgroup, $r = -0.065$, $p = .799$, $N = 18$. The difference in correlation coefficients for perspective-transforming and transposing subgroups was statistically significant, $Z = 1.96$, $p = .05$ (Snedecor and Cochran, 1967).

DISCUSSION

The present study sought to clarify the cognitive processes involved in spatial perspective-taking by assessing whether strategy moderates the relationship between performance in tasks designed to measure perspective-taking (OBT) and response inhibition (Transpose). For the Transpose task, RTs were elevated for the incompatible relative to compatible condition, consistent with the costs of inhibiting a prepotent spatially compatible response in response to a cue (Gardner and Potts, 2011). For the OBT task, RTs were elevated for the front- relative to back-view condition, consistent with, depending upon putative route, either the costs of an embodied imagined transformation of perspective (e.g., Zacks et al., 1999; Blanke et al., 2005; Mohr et al., 2010), or the costs of inhibiting a spatially compatible response in response to the appearance of the front-view of the figure (Gardner and Potts, 2011, see also May and Wendt, 2012). The two strategy subgroups were not found to differ on overall speed of responding, or size of compatibility effect, in either the OBT

or Transpose tasks. However, as predicted, participants' self-report of which of these two strategies they had employed for the OBT task was found to moderate the relationship between the degree of elevation in response times resulting from incompatibility in the OBT and Transpose tasks. Specifically, response inhibition ability, indexed by the Transpose task, was found to be related to perspective-taking ability—but selectively for those reporting that they had adopted the disembodied spatial transposing strategy. This relationship was absent in those reporting having adopted an embodied perspective transformation strategy.

The selective association found between performance on the Transpose and OBT tasks implies that response inhibition ability predicts perspective-taking ability only among those that choose to take on another's perspective using a "spatial transposing" strategy—that is, by reconfiguring spatial relationships as they appear from one's own perspective. This association complements earlier work (Qureshi, 2008), by showing that the association between response inhibition and perspective-taking also generalizes to the perspective-taking performance measured by the OBT task, more specifically the relative ability to adopt a perspective differing from one's own by 180° compared to 0° . The selective association is particularly important in providing evidence that this disembodied spatial transposing strategy is dissociable from an embodied perspective transformation strategy. Previously, the existence of this route was only implied by the absence of an association between trait empathy and perspective-taking ability otherwise present for those reporting having performed perspective transformations (Gronholm et al., 2012). Furthermore, the transposing subgroup also showed greater improvement in accuracy between the OBT and Transpose tasks. This could be explained in terms of their strategy for perspective-taking rendering the OBT task computationally equivalent to the Transpose task which leads to greater carry over from practice in comparison to the perspective transformation subgroup. These dissociations provide support for an independent disembodied route, consistent with findings for sensorimotor interference within the spatial updating literature for imaginal perspective changes in remembered environments (May, 2004).

Combined with earlier results (Gronholm et al., 2012), the present findings complete a double dissociation, implying that the spatial transposing and perspective transformation strategies reflect two separable routes to spatial perspective-taking. Where previously we described these strategies as empathic and non-empathic (Gronholm et al., 2012), we now suggest that this dissociation might be better characterized as between "embodied" perspective transformations and "disembodied" routes (Kessler and Rutherford, 2010; Stocker, 2012; Becchio et al., 2013; Tomasino and Rumiati, 2013). The embodied route, probably mediated by mental simulation of self motion (Lenggenhager et al., 2008; Stegmann et al., 2011), appears linked to social determinants such as trait empathy (Gronholm et al., 2012). The disembodied route, which the present results suggest involves the deliberate reconfiguration of spatial relationships as they appear from one's own point of view, may be completely insensitive to social context or whether the new perspective is a position occupied by a person. This dis-

sociation builds upon evidence suggesting dissociable processes for level 1 and level 2 perspective-taking (Michelon and Zacks, 2006; Kessler and Thomson, 2010), by implying that further fractionation is possible purely within the level 2 perspective-taking involved in the OBT task, confirming the hitherto untested hypotheses of other authors (e.g., Thakkar et al., 2009; Mohr et al., 2010; Thakkar and Park, 2010).

These results also inform debate on the suitability of the OBT and related tasks to measure spatial perspective-taking (Gardner and Potts, 2011; May and Wendt, 2012, under review). Given the way that it is operationalized with only four types of stimuli, the OBT task may be particularly susceptible to low-level alternative strategies. However, there are at least three reasons not simply to dismiss the OBT task as a test of spatial perspective-taking on the basis that it may be solved by the reconfiguration of spatial relationships as they appear from one's own position. First, this issue does not appear to be unique to the OBT task. A similar mechanism could contribute to performance in other tasks employing laterality judgments (e.g., Michelin and Zacks, 2006; Kessler and Thomson, 2010, see May and Wendt, under review), although it is less likely to extend to tasks requiring participants to imagine the appearance of an array from a novel perspective (e.g., Langdon and Coltheart, 2001). Second, evidence that imitation also imposes a demand to inhibit incompatible S-R mappings (e.g., Ishikura and Inomata, 1995; Heyes and Ray, 2004; Jackson et al., 2006; Chiavarino et al., 2007), suggests that spatial transposing may be pervasive in face-to-face social interactions. Third, the present results imply that although the low-level reconfiguration of spatial relationships may contribute to performance in the OBT task, this may be restricted to a subset of participants adopting a particular "spatial transposing" strategy. Thus, this finding implies that identifying the interpersonal determinants of strategy selection may be a worthwhile avenue for research in spatial perspective-taking, and social interaction more generally (see Mohr et al., 2013).

The finding that the two dissociable perspective-taking processes may be reliably categorized by self-reported strategy also implies that the route to perspective-taking may be endogenously triggered. This evidence contrasts with other research showing exogenous triggering of embodied perspective-taking, either by revealing enhanced perspective-taking for body present compared with body absent conditions (Becchio et al., 2013), or by showing that congruence between the postures of the participant and an avatar facilitated performance (Kessler and Thomson, 2010). However, our finding that the presentation of a figure is not sufficient to elicit embodied perspective-taking corresponds to the finding that galvanic vestibular stimulation selectively disrupts mental task performance for participants adopting an egocentric rather than object-based transformation strategy (Lenggenhager et al., 2008). It also complements work demonstrating that the presence of a body was neither a necessary condition for response latency being related to the extent of imagined self rotation (Michelon and Zacks, 2006), nor for congruence effects between the participant's body position and direction of imagined rotation (Kessler and Thomson, 2010). Although research to date implies that embodied perspective-taking can be

both endogenously and exogenously driven, the significance of the present results is that they demonstrate endogenous driven embodied perspective-taking for a task that, by inviting participants to step into the shoes of the schematic figure, might have been assumed likely to have triggered an embodied route exogenously. This corresponds with the view that the strategic modulation of embodied and disembodied routes is pervasive in various domains of cognition, including object based mental rotation, and language learning (Tomasino and Rumiati, 2013).

The absence of a correlation for the perspective transformation subgroup between perspective-taking and response inhibition, and the statistically significant difference in correlation coefficient compared to the spatial transformation subgroup, implies this executive function is not involved in the "embodied" route to perspective-taking to the same extent as for the "disembodied" route adopted by spatial transposers. This is consistent with earlier findings implying the fractionation of perspective-taking processes into cognitively efficient and cognitively demanding components (Michelon and Zacks, 2006; Samson et al., 2010; Qureshi et al., 2010), and with the different developmental trajectories of perspective-taking and executive function (Dumontheil et al., 2010). Previously, the cognitively demanding perspective-taking process was taken to be either the calculation of spatial relationships relative to an alternative viewpoint (Michelon and Zacks, 2006—"level 2" knowledge; Flavell et al., 1981) or the selection of either the self or other perspective (Qureshi et al., 2010). Whereas, the cognitively efficient process in both studies was the calculation about what is visible from another viewpoint (level 1 knowledge). By contrast, in the present study, both routes involve the calculation of spatial relationships, but only the disembodied route appears to load onto response inhibition. This raises the possibility that a dedicated, domain specific, route also exists for level 2 perspective-taking, provided that one uses an embodied strategy (cf Amorim et al., 2006). We speculate that this route may place relatively light demands on domain general resources, although further research is required to assess this possibility.

Limitations of the correlational method adopted in the present study should be acknowledged. On one hand, our choice of the Transpose task as a measure of response inhibition could have elevated correlations in both subgroups due to shared variance attributable to procedural similarity between the OBT and Transpose tasks. On the other hand, methodological limitations could contribute to the absence of a statistically significant correlation for the embodied perspective transformation subgroup; the perspective transformation subgroup was smaller, and potentially more heterogeneous than the spatial transformation subgroup. Although the magnitude of the correlation coefficients reported here between perspective-taking and response inhibition should therefore be interpreted with caution, critically these correlations were found to differ significantly between subgroups. Differences in sample size would have reduced the power of this test, and shared variance attributable to task similarity would be expected to affect both subgroups equally. Nonetheless, our findings should ideally be corroborated using an experimental approach such as the dual task methodology used by Qureshi et al. (2010), despite contrasting selective associations being an

established methodology for revealing evidence for dissociable processes (e.g., Asendorpf et al., 2002).

Finally, it should be noted that participants' performance for the OBT and Transpose tasks did not show the close equivalence found in other research using the same tasks (Gardner and Potts, 2011). In the present study, overall performance was found to be better for the Transpose than the OBT task, both in terms of shorter RTs and fewer errors, and the size of the compatibility effect was greater in the Transpose task than in the OBT task. At first glance, these findings might be taken to imply that the Transpose task is not a good control for the OBT task, or, alternatively, that the disembodied spatial transposing route is more efficient than the embodied perspective transformation route. However, in the current experiment, all participants completed the Transpose task after the OBT task in order not to influence the strategy employed for the OBT task—a likely possibility had order been counterbalanced. Therefore, the between task differences in performance could be accounted for by practice in the first task (OBT) leading to better performance in a similar second task (Transpose), particularly for the compatible trials, and particularly for those adopting a disembodied spatial transposing strategy. Such differences were not found when different participants completed the two tasks (Gardner and Potts, 2011).

In conclusion, our main finding was a selective association whereby response inhibition was related to perspective-taking ability only among participants adopting a “spatial transpos-

ing” strategy—that is, by reconfiguring spatial relationships as they appear from one's own perspective. Combined with earlier results (Gronholm et al., 2012), this evidence completes a double dissociation between two independent routes to perspective-taking in the OBT task. We propose that these routes either recruit “embodied” egocentric mental transformation processes, or involve the “disembodied” reconfiguration of spatial relationships. The contributions made by these findings are that they elucidate the processes involved in perspective-taking, imply that perspective-taking route is under higher order control, and lend support to the hypothesis that embodied routes to perspective-taking place minimal demands on domain general executive functions.

ACKNOWLEDGMENTS

We thank both reviewers for helpful comments on a previous version of this manuscript. We also gratefully acknowledge the contribution made by all our research participants, including students taking 1PSY509 at the University of Westminster during the Autumn of 2012.

The Author, Petra Gronholm, receives funding support from the National Institute for Health Research (NIHR) Mental Health Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the author and not necessarily those of the NHS, the NIHR or the Department of Health.

REFERENCES

- Amorim, M. A. (2003). “What is my avatar seeing?": the coordination of “out-of-body” and “embodied” perspectives for scene recognition across views. *Vis. Cogn.* 10, 157–199. doi: 10.1080/713756678
- Amorim, M.-A., Isableu, B., and Jaraya, M. (2006). Embodied spatial transformations: “Body analogy” for the mental rotation of objects. *J. Exp. Psychol. Gen.* 135, 327–347. doi: 10.1037/0096-3445.135.3.327
- Asendorpf, J. B., Banse, R., and Mücke, D. (2002). Double dissociation between implicit and explicit personality self-concept: the case of shy behavior. *J. Pers. Soc. Psychol.* 83, 380–393. doi: 10.1037/0022-3514.83.2.380
- Becchio, C., Del Giudice, M., Dal Monte, O., Latini-Corazzini, L., and Pia, L. (2013). In your place: neuropsychological evidence for altercentric remapping in embodied perspective taking. *Soc. Cogn. Affect. Neurosci.* 8, 165–170. doi: 10.1093/scan/nsr083
- Blanke, O., Mohr, C., Michel, C. M., Pascual-Leone, A., Brugger, P., Seeck, M., et al. (2005). Linking out-of-body experience and self processing to mental own-body imagery at the temporoparietal junction. *J. Neurosci.* 25, 550–557. doi: 10.1523/JNEUROSCI.2612-04.2005
- Chiavarino, C., Apperly, I., and Humphreys, G. (2007). Exploring the functional and anatomical bases of mirror-image and anatomical imitation: the role of the frontal lobes. *Neuropsychologia* 45, 784–795. doi: 10.1016/j.neuropsychologia.2006.08.007
- Cohen, M. X., David, N., Vogeley, K., and Elger, C. E. (2009). Gamma band activity in the human superior temporal sulcus during mentalizing from nonverbal social cues. *Psychophysiology* 46, 43–51. doi: 10.1111/j.1469-8986.2008.00724.x
- David, N., Aumann, C., Santos, N. S., Bewernick, B. H., Eickhoff, S. B., Newen, A., et al. (2008). Differential involvement of the posterior temporal cortex in mentalizing but not perspective taking. *Soc. Cogn. Affect. Neurosci.* 3, 279–289. doi: 10.1093/scan/nsn023
- Dumontheil, I., Apperly, I. A., and Blakemore, S. J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Dev. Sci.* 13, 331–338. doi: 10.1111/j.1467-7687.2009.00888.x
- Flavell, J. H., Everett, B. A., Croft, K., and Flavell, E. R. (1981). Young children's knowledge about visual perception: further evidence for the level 1–level 2 distinction. *Dev. Psychol.* 17, 99–103. doi: 10.1037/0012-1649.17.1.99
- Gardner, M. R., and Potts, R. (2010). Hand dominance influences the processing of observed bodies. *Brain Cogn.* 73, 35–40. doi: 10.1016/j.bandc.2010.02.002
- Gardner, M. R., and Potts, R. (2011). Domain general mechanisms account for imagined transformations of whole body perspective. *Acta Psychol.* 137, 371–381. doi: 10.1016/j.actpsy.2011.04.006
- Gronholm, P., Flynn, M., Edmonds, C., and Gardner, M. (2012). Empathic and non-empathic routes to visuospatial perspective-taking. *Conscious. Cogn.* 21, 494–500. doi: 10.1016/j.concog.2011.12.004
- Heyes, C., and Ray, E. (2004). Spatial S–R compatibility effects in an intentional imitation task. *Psychon. Bull. Rev.* 11, 703–708. doi: 10.3758/BF03196623
- Ishikura, T., and Inomata, K. (1995). Effects of angle of model demonstration on learning of motor skill. *Percept. Mot. Skills* 80, 651–658. doi: 10.2466/pms.1995.80.2.651
- Jackson, P., Meltzoff, A., and Decety, J. (2006). Neural circuits involved in imitation and perspective-taking. *Neuroimage* 31, 429–439. doi: 10.1016/j.neuroimage.2005.11.026
- Kaiser, S., Walther, S., Nennig, E., Kronmüller, K., Mundt, C., et al. (2008). Gender specific strategy use and neural correlates in a spatial perspective-taking task. *Neuropsychologia* 46, 2524–2531. doi: 10.1016/j.neuropsychologia.2008.04.013
- Kessler, K., and Rutherford, H. (2010). The two forms of visuospatial perspective taking are differently embodied and subserve different spatial prepositions. *Front. Psychol.* 1:213. doi: 10.3389/fpsy.2010.00213
- Kessler, K., and Thomson, L. A. (2010). The embodied nature of spatial perspective taking: embodied transformation versus sensorimotor interference. *Cognition* 114, 72–88. doi: 10.1016/j.cognition.2009.08.015
- Langdon, R., and Coltheart, M. (2001). Visual perspective-taking and schizotypy: evidence for a simulation-based account of mentalizing in normal adults. *Cognition* 82, 1–26. doi: 10.1016/S0010-0277(01)00139-1
- Lenggenhager, B., Lopez, C., and Blanke, O. (2008). Influence of galvanic vestibular stimulation on egocentric and object-based mental transformations. *Exp. Brain Res.* 184, 211–221. doi: 10.1007/s00221-007-1095-9

- Leslie, A. M., German, T. P., and Polizzi, P. (2005). Belief-desire reasoning as a process of selection. *Cogn. Psychol.* 50, 45–85. doi: 10.1016/j.cogpsych.2004.06.002
- May, M. (2004). Imaginal perspective switches in remembered environments: transformation versus interference accounts. *Cogn. Psychol.* 48, 163–206. doi: 10.1016/S0010-0285(03)00127-0
- May, M., and Wendt, M. (2012). Separating mental transformations and spatial compatibility effects in the own body transformation task. *Cogn. Process.* 13, 257–260. doi: 10.1007/s10339-012-0455-y
- Michelon, P., and Zacks, J. M. (2006). Two kinds of visual perspective taking. *Percept. Psychophys.* 68, 327–337. doi: 10.3758/bf03193680
- Mohr, C., Rowe, A. C., and Blanke, O. (2010). The influence of sex and empathy on putting oneself in the shoes of others. *Br. J. Psychol.* 101, 277–291. doi: 10.1348/000712609X457450
- Mohr, C., Rowe, A. C., Kurokawa, I., Dendy, L., and Theodoridou, A. (2013). Bodily perspective taking goes social: the role of personal, interpersonal and intercultural factors. *J. Appl. Soc. Psychol.* 43, 1369–1381. doi: 10.1111/jasp.12093
- Qureshi, A. W. (2008). *The Cognitive Bases of Mental State Reasoning in Adults*. Unpublished doctoral dissertation, Birmingham, UK: University of Birmingham. <http://etheses.bham.ac.uk/301/>
- Qureshi, A. W., Apperly, I. A., and Samson, D. (2010). Executive function is necessary for perspective selection, not level-1 visual perspective calculation: evidence from a dual-task study of adults. *Cognition* 117, 230–236. doi: 10.1016/j.cognition.2010.08.003
- Ruby, P., and Decety, J. (2004). How would you feel versus how do you think she would feel? A neuroimaging study of perspective-taking with social emotions. *J. Cogn. Neurosci.* 16, 988–999. doi: 10.1162/0898929041502661
- Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., and Bodley Scott, S. E. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people see. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 1255–1266. doi: 10.1037/a0018729
- Schneider, W., Eschman, A., and Zuccolotto, A. (2002). *E-Prime: User's guide*. Psychology Software Incorporated.
- Snedecor, G. W., and Cochran, W. G. (1967). *Statistical Methods*. 6th Edn. Ames: Iowa State University Press.
- Steggemann, Y., Engbert, K., and Weigelt, M. (2011). Selective effects of motor expertise in mental body rotation tasks: comparing object-based and perspective transformations. *Brain Cogn.* 76, 97–105. doi: 10.1016/j.bandc.2011.02.013
- Stocker, K. (2012). Toward an embodiment-disembodiment taxonomy. *Cogn. Process.* 13, 347–350. doi: 10.1007/s10339-012-0495-3
- Thakkar, K. N., Brugger, P., and Park, S. (2009). Exploring empathic space: correlates of perspective transformation ability and biases in spatial attention. *PLoS One* 4:e5864. doi: 10.1371/journal.pone.0005864
- Thakkar, K. N., and Park, S. (2010). Empathy, schizotypy, and visuospatial transformations. *Cogn. Neuropsychiatry* 15, 477–500. doi: 10.1080/13546801003711350
- Tomasino, B., and Rumiati, R. I. (2013). At the mercy of strategies: the role of motor representations in language understanding. *Front. Psychol.* 4:27. doi: 10.3389/fpsyg.2013.00027
- Tversky, B., and Hard, B. M. (2009). Embodied and disembodied cognition: spatial perspective-taking. *Cognition* 110, 124–129. doi: 10.1016/j.cognition.2008.10.008
- Zacks, J. M., and Michelon, P. (2005). Transformations of visuospatial images. *Behav. Cogn. Neurosci. Rev.* 4, 96–118. doi: 10.1177/1534582305281085
- Zacks, J., Rypma, B., Gabrieli, J. D. E., Tversky, B., and Glover, G. H. (1999). Imagined transformations of bodies: an fMRI investigation. *Neuropsychologia* 37, 1029–1040. doi: 10.1016/S0028-3932(99)00012-3

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 May 2013; accepted: 22 July 2013; published online: 13 August 2013.
 Citation: Gardner MR, Brazier M, Edmonds CJ and Gronholm P (2013) Strategy modulates spatial perspective-taking: evidence for dissociable disembodied and embodied routes. *Front. Hum. Neurosci.* 7:457. doi: 10.3389/fnhum.2013.00457
 Copyright © 2013 Gardner, Brazier, Edmonds and Gronholm. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Relating spatial perspective taking to the perception of other's affordances: providing a foundation for predicting the future behavior of others

Sarah H. Creem-Regehr*, Kyle T. Gagnon, Michael N. Geuss and Jeanine K. Stefanucci

Department of Psychology, University of Utah, Salt Lake City, UT, USA

Edited by:

Klaus Kessler, University of Glasgow, UK

Reviewed by:

Michael A. Riley, University of Cincinnati, USA

Donatella Spinelli, Università di Roma "Foro Italico," Italy

*Correspondence:

Sarah H. Creem-Regehr,
Department of Psychology,
University of Utah, 380 S. 1530 E.,
Room 502, Salt Lake City, UT 84112,
USA
e-mail: sarah.creem@psych.
utah.edu

Understanding what another agent can see relates functionally to the understanding of what they can do. We propose that spatial perspective taking and perceiving other's affordances, while two separate spatial processes, together share the common social function of predicting the behavior of others. Perceiving the action capabilities of others allows for a common understanding of how agents may act together. The ability to take another's perspective focuses an understanding of action goals so that more precise understanding of intentions may result. This review presents an analysis of these complementary abilities, both in terms of the frames of reference and the proposed sensorimotor mechanisms involved. Together, we argue for the importance of reconsidering the role of basic spatial processes to explain more complex behaviors.

Keywords: affordances, perspective taking, perception and action, spatial cognition, motor simulation

How can different people look at the same object or event and perceive (pretty much) the same thing?... What is even more intriguing is the possibility that I can perceive the meaning afforded by the existing layout of surfaces in the environment for another person as well as for me. What underlies the commonality of perception across diverse individuals? (Mark, 2007, p. 108)

Humans are inherently social beings as evident by the fact that we live in families, work in groups, share meals with one another, relax with friends, and are often entertained by watching the lives of other humans. This is not a new idea, but rather the motivation for establishing the field of social psychology. Furthermore, the "ecological dominance—social competition model" proposed by Alexander (1990), suggests that one of the most influential evolutionary pressures that shaped human intelligence was "...a within-species co-evolutionary arms race in which success depended on effectiveness in social competition" (pp. 4–7). Whether one is trying to gauge an enemy's weakness, or striving to cooperate with a friend, the ability to predict the future behavior of other humans allows actors to adjust their current behavior, providing them with a powerful social advantage (for a review, see Flinn et al., 2005).

Predicting the future behavior of others involves both an understanding of what another person is capable of *doing* and an understanding of their current *goals*. Studies that have explored how a viewer makes judgments of another's action capabilities—other's *affordances*—have revealed that viewers can adequately judge what another is capable of performing when provided information about this others' ability to act (e.g., body dimensions or kinematic information). The ability to take the spatial perspective of another person may provide information about the goals of this other person by revealing their line of sight.

While judging other's affordances and spatial perspective taking are often studied under the disciplines of perception and spatial cognition, we propose that these two abilities may also work together to build a foundation for *social* cognition. Our goal is to review the literature from both domains to determine how *spatial perspective taking* and the *perception of other's affordances* work together to predict the behavior of others. In addition, we will review neurological evidence that may provide a biological mechanism common to both processes.

We begin with a review of the behavioral evidence demonstrating that observers have an understanding of what others can do through explicit judgments of affordances for another agent. Second, we review the evidence that spatial perspective taking can reveal the intentions of another agent. Then, we consider how spatial perspective taking and judging affordances for others may be integrated to provide an observer with the information necessary to predict the behavior of others. Next, we consider two distinct but not necessarily exclusive accounts of the underlying mechanisms of social perception and action—motor resonance/simulation (Sebanz et al., 2003; Bosbach et al., 2005; Gallese and Sinigaglia, 2011) and ecological approach/information-based (Marsh et al., 2006; Ramenzoni et al., 2008b). We discuss evidence for the possibility of shared mechanisms with spatial perspective taking and similarities and differences between the way frames of reference are used. Self-judgments are made with respect to the viewer's reference frame (*egocentric*). An important theme is whether judgments about another agent use a transformation of the viewer's reference frame onto the other's egocentric reference frame to update spatial relations (termed *egocentric transformation*), or the use of an *allocentric* frame—the use of relative spatial relations between two points outside of one's egocentric frame. We conclude with a

discussion of how both abilities, judging other's affordances and taking the perspective of another, while likely different processes, rely on a social context and support the broader goal of social coordination.

PERCEIVED SELF-AFFORDANCES

Knowing what another person is capable of doing is often considered in the context of the theory of affordances (Gibson, 1979). Gibson's (1979) ecological theory of perception stated that the perception of the environment is directly related to the actions that one is capable of performing in the environment. The term affordance is used to describe the fit between environmental (perceived through the senses) and person features (e.g., size of the body or kinematic capabilities; Michaels and Carello, 1981; Turvey, 1992; Stoffregen, 2003; Plumert et al., 2004). For example, a tree branch lying on the ground can afford sitting, stepping on, or stepping over. A tree branch placed sufficiently higher does not afford sitting or stepping over, but may instead afford walking under. In sum, affordances are opportunities for action present in the environment that are defined by the observer's action capabilities (Turvey, 1992; Stoffregen, 2003).

People are able to judge whether an environment affords a particular action without executing the actual action (termed affordance judgment) and scale environmental features to their abilities (Mark, 1987; Warren and Whang, 1987). For example, Warren and Whang (1987) found that people required apertures to be 1.16 times their shoulder width when judging whether an aperture afforded non-rotated passage. They also found that this affordance was scaled to the eye height of the participant suggesting that the visual information was related to body dimensions and abilities. Other body dimensions are taken into account for other types of actions. For instance, the maximum climbable surface has been found to be about 0.88 times the length of the actor's leg (Warren, 1984; Mark and Vogeley, 1988). The critical boundary has been identified for a number of different actions including grasping (Newell et al., 1989) sitting (Mark, 1987), and reaching (Carello et al., 1989).

Affordances can also be learned or recalibrated to fit new capabilities or novel environments (Wagman and Taylor, 2005; Ishak et al., 2008). Ishak et al. (2008) demonstrated that participants were able to recalibrate decisions about whether their hand could fit through an aperture when their hand was made larger. Wagman and Taylor (2005) manipulated the width of participants by having them hold a t-shaped object at their waist. They showed that participants almost instantly recalibrated judgments of passage through an aperture when their body size was widened by holding the pole. They attributed the immediacy of recalibration to the ability of participants to determine the length of the pole by wielding it prior to judgments. Higuchi et al. (2004) investigated the ability of novice wheelchair users to judge their ability to pass through an aperture when in the wheelchair. They found that novice users often judged apertures to be passable when they would not actually fit through in the wheelchair (aperture to wheelchair width ratio of .92). While participants' judgments improved after 8 days of practice with the wheelchair, they did not reach levels observed in

baseline performance (without the wheelchair). Under a different paradigm, Mark (1987) and Mark et al. (1990) investigated how an actor comes to know the specific relationship between an environmental extent and their action capability. Mark (1987) altered standing eye height by requiring participants to wear 10 cm blocks underfoot. They then judged their ability to sit on surfaces of different heights. Without practice sitting, participants' judgments of what they could sit on returned to the critical boundary when not wearing blocks over the course of 30 trials. Mark et al. (1990) then systematically manipulated information available to the participant when wearing the blocks. They found that participants were able to recalibrate their judgments of suitability to their new height when they were able to locomote, move their heads or eyes, or lean to the side. Restricting visual information by providing only monocular viewing through a peephole or restricting movement by requiring participants to rest their heads against a wall significantly reduced participants' ability to recalibrate information and judge suitability with blocks underfoot.

This body of work is important because it shows that people are fairly accurate in judging what they are capable of performing in an environment. This work also demonstrates that people are able to quickly calibrate their affordance judgments to changes in their ability to act. The rate of recalibration is often determined by the degree to which observers experience or gain information about the change to their capabilities. Others have theorized that flexibility in affordance judgments and the performance of actions is necessary to deal with changes in the demands of the situation, changes to the criteria for success (the goal), and changes with the availability of visual information (Fajen et al., 2009). Importantly, this work demonstrates that all of the information necessary to judge and carry out an action is available to the person in the ambient stimulus arrays in which the person is immersed.

PERCEIVING OTHERS' AFFORDANCES

As introduced above, affordances for the self are typically grounded in an egocentric frame of reference and scaled in terms of one's body dimensions with respect to the current viewpoint. However, when judgments of other people's affordances are made, it is possible that observers switch to an allocentric frame of reference. We define allocentric judgments as those that are relative judgments made between two points, outside of the self. As such, the environment is scaled to the other's body rather than to one's own (Stoffregen et al., 1999). Rochat (1995) examined reaching affordances of children and adults, asking whether young children distinguish reachability for themselves and others. The findings revealed that both children and adults scaled their judgments of reaching to their own physical characteristics in the self-judgments and to the other's physical characteristics for judgments of the other. In addition, all subjects showed the ability to take into account the other's change in reaching height when viewing the other on "tip-toes." These findings suggest an early ability to switch from an egocentric to an allocentric frame of reference in this task. More recent studies with adults have focused on judging others' affordances when the action involves either a single other person

or the potential actions of dyads (the observer and another person).

AFFORDANCES FOR ONE OTHER

For single-person affordances, multiple studies have shown that observers accurately scale environmental features to the action capabilities of the actor being observed (Stoffregen et al., 1999; Ramenzoni et al., 2008b,c, 2010). Stoffregen et al. (1999) examined observers' abilities to perceive the maximum height at which another actor could sit. In this extensive study, the observer judged their own and another actor's affordances for sitting, while varying the height of the other actor as well as the viewer's experience with observing kinematic displays of the other actor perform non-sitting actions. They found that affordances of others were scaled with respect to the actor's leg length. In addition, Ramenzoni et al. (2008b) tested judgments of maximum reaching height of the self and another with the goal of testing whether eye-height information would be used. Observers judged how high they or a different sized actor could reach an object while the observer stood on the floor, or one of two different sized steps. The other actor always remained standing on the floor. Judgments, when scaled to the observer's reaching height for the self and to the actor's reaching height for the other, were near 1.0, indicating that estimates were very accurate, both for self and other. These results support the notion that affordances are scaled to the intrinsic units of the observer (in self-judgments) or actor (in other judgments). Mark (2007) summarizes a series of studies following up on these findings, replicating the effect for sitting, climbing, and stepping affordances. These studies argue for the claim that an allocentric frame of reference is adopted when judging affordances for others and that observers can do this in the context of judging their own affordances as well-switching easily from an egocentric to an allocentric framework.

Some actions, like jumping-and-reaching, require the observer to have information about the actor's kinematic abilities and not just information about the size of the actor (Weast et al., 2011). Stoffregen et al. (1999) found that when observers were provided with the appropriate information about the underlying dynamic actor properties, they could accurately judge the other's ability. In addition, Ramenzoni et al. (2010) asked whether a learning paradigm would influence maximum jump to reach estimates for another actor over multiple repeated trials in a similar manner as self-judgments. They found an increase in accuracy across trials for self-estimates, but not for actor estimates. The lack of changes over time in the other's judgments suggests that judgments of others are not dependent on judgments of self. However, their second study tested the influence of watching the actor perform a task related in dynamics (lifting) or unrelated to the dynamics of jumping and reaching (torso-twist) on judgments of reach-by-jump for both the self and the other. They found that watching an actor perform a related task improved the accuracy of the estimates of the actor's capabilities but watching the unrelated task did not help. The second experiment showed that experience with another's kinematic abilities facilitates related affordance judgments, suggesting the importance of calibrating the observer to specific action-relevant information about the actor's capabilities.

Weast et al. (2011) investigated how expertise influenced the perception of affordances for others. They found that basketball players were better than novices at judging the jump-and-reach height of another actor but that basketball players were no better than novices at judging a non-sports-relevant action (sitting height). In their second study, they demonstrated that with exposure to kinematic information, basketball players', but not novices, judgment of maximum jump-and-reach improved. This finding suggested that basketball players had enhanced sensitivity to kinematic information. These findings emphasize the claim that the relationship between the other's physical body parameters (e.g., size and capabilities) and the environment, as well as degree to which someone has experience with a specific relationship, is critical in informing decisions about others' ability to act in the environment.

Another series of studies examined the ability and accuracy of adults to judge reachability of children (Cordovil and Barreiros, 2010, 2011) generally supporting the claim that observers scale affordances to the other's body, but also showing less consistent overestimation in judgments of children's reaching compared to adult self-judgments. As in Ramenzoni et al. (2010); Cordovil et al. (2013) asked whether accuracy in judging another's affordance may be a function of experience or practice. Cordovil et al. (2013) tested adults' judgments of the maximum standing reachability, reach and jump reachability, and step-length of a 5 year old boy, before and after observing the boy perform the action. They found that viewing the boy's actual affordance improved the more complex affordances (jump-to-reach and step-length) but had little effect on the basic reaching while standing judgment. The observation/practice manipulation suggests that when given more information about the relationship between the other actor and the environment, observers can calibrate the information to adjust their response.

A somewhat different take-home message comes from Ramenzoni et al. (2008a) in a study of perceived maximum reach by jumping. The observer's capability to jump was manipulated by wearing ankle weights. Judgments were made both for the self and for another actor who did not wear ankle weights. Interestingly, estimations of jumping-reach height were lowered not only for the self, but also for the other actor, specifically after the observer walked while wearing the weights. The effect of ankle weights to reduce the critical boundary of reach by jumping is consistent with the body of work showing that effort or behavioral potential influences spatial judgments (Proffitt, 2006). However, what is unique about these findings is that the manipulation affected judgments of what someone else could do. These results support a social context underlying perceived affordances and suggest that judging others action capabilities may rely somewhat on how the observer herself can act. Thus, the task becomes one at least partially based in the observer's egocentric frame of reference. Notably, in this study, observers may not have had sufficient information about the actor's jumping ability to rely solely on the relationship between the other and the environment to make their judgment. The influence of the ankle weights on judgments of the actor's capabilities may be erased if sufficient information about the actor's kinematic capabilities is provided.

AFFORDANCES FOR DYADS

Another way in which researchers have assessed the ability to judge others' affordances has been to examine dyads or joint actions. This work looks at how observers are able to make decisions about actions when these actions are to be performed in correspondence with another person. This is especially interesting because, unlike the single-person judgments, observer and actor actions necessarily have a direct influence on one another. Further, different action capabilities may result as two observers coordinate their actions (Isenhower et al., 2010). Chang et al. (2009) took this approach in an environment-person-person system, testing whether adults would accurately estimate their ability to pass through an aperture while walking through with a child. The adult and child were attached with a Velcro strip at the child's elbow and the adult's wrist. The results showed that adults were able to accurately perceive affordances for passage with the child. Consistent with the self and single-actor studies, the results revealed that judgments were scaled additively to the intrinsic units of the adult shoulder width + child shoulder width.

Similarly, Davis et al. (2010) assessed how two adults performed the joint action of walking through an aperture. First, they established that a "shared" model, rather than an additive model, better predicted the critical boundary for the dyad's actual passage. This showed that the critical aperture width was less than the sum of the critical aperture widths for each actor separately suggesting that coordinated actions are scaled to the combined action capabilities of the two actors. Further, they examined the influence of action-observation experience on perceived affordances for passage of the self and the other actor. Participants either viewed the other actor walk, walked alongside the actor, or viewed the actor standing only. As in previous work, the ratio of critical width to actual shoulder width (scaled to the participant for self-judgments and the actor for other judgments) were nearly identical, suggesting the ability to use the other's intrinsic scale to make estimates. However, the dyad estimates were significantly underestimated with respect to the actual joint affordance. Furthermore, unlike some of the previous work, there was no effect of the increased action-observation conditions. The reduced accuracy in response is similar to the person-plus-tool studies mentioned above (Higuchi et al., 2004; Wagman and Taylor, 2005), and is likely a result of insufficient information or lack of experience walking as a dyad.

The body of literature on perceiving affordances for one other and for dyads suggests that observers are capable of judging what another person can perform. These judgments are likely completed by using an allocentric frame of reference, and they reveal what actions another person is or is not capable of performing in the current environment. In addition, an observer's judgments about another person are scaled to the action capabilities of the other person or the other person + self system. When making judgments about actions that require more than a relative size comparison, observer's judgments about another's affordances improve when they see the actor perform similar dynamic movements. There is also evidence that when an observer is not provided with kinematic information about the

actor that the observer may use their own ability as a baseline to judge what another could perform. Notably, much of the existing literature involves judgments of others in tasks such as walking through apertures that does not involve critical time constraints. It may be that in more interactive dyadic tasks, such as lifting a box together, different information relevant to action coordination is used (see later section on synergistic accounts). In all, the evidence points to the use of an allocentric frame of reference generally used for perceiving other's affordances, with the influence of an egocentric frame of reference when there is insufficient information available about the other's capabilities.

OTHER'S AFFORDANCES: SUMMARY AND CONCLUSION

There is clear evidence for the human ability to judge what others can do, as well as to use what others can do to influence their own action judgments. Together, this work reinforces the idea that others' affordances are used as an important component in the broader problem of predicting the future behavior of others. However, if humans only had at their disposal the ability to judge action capabilities for another, they would have to consider all of the affordances that a given environment offers to this other person. This would be a rather cumbersome way to predict the behavior of others, unless there was a meaningful way to focus on only a few affordances. The theory of affordances (reviewed above) may provide some insight to this problem. When perceiving affordances for oneself, observers orient their senses to the properties of the environment that are necessary for perceiving a particular affordance. For example, if someone intends to grasp an object sitting atop a tall shelf, they will likely look in the direction of the object. If they can reach the object they will then do so, otherwise they will likely look around for a heightened surface that affords standing/climbing and will use this surface to reach the object. Therefore, assuming that other people also orient their senses to pick up information relevant to a potential action, an observer can simply identify where this other person is looking and consider the actions that this spatial location may afford for the other person. Much of the research that has examined our ability to detect where another person is looking, what they can/cannot see, and their spatial relationship to other objects in the environment is called spatial perspective-taking and will be reviewed next.

SPATIAL PERSPECTIVE TAKING

Research on spatial perspective taking has a long history across both developmental and cognitive psychology ranging from Piaget's classic three mountain task (Piaget and Inhelder, 1967) to a comparison of physical and imagined body rotations (Rieser, 1989). The role that spatial perspective taking plays in spatial memory and navigation has also been examined (Loomis et al., 1999; Shelton and McNamara, 2004). Perspective-taking research is also interested in how observers determine what another person can or cannot see, and is often called joint (shared) visual attention (Frischen et al., 2007). In general, spatial perspective taking encompasses a class of phenomena that involve accessing spatial information relative to a viewpoint different from one's own egocentric viewpoint. Importantly, we will examine whether

these abilities may allow an observer to suppose the intentions of another person.

Spatial perspective taking can be differentiated into Level-1 perspective taking (PT-1) and Level-2 perspective taking (PT-2) based on developmental stages and proposed underlying processes (Salatas and Flavell, 1976; Kessler and Rutherford, 2010). PT-1 is often defined as a visibility task in which an observer determines what another person can or cannot see. One of the first studies examining this type of task with adults was aimed at establishing shared common ground in a virtual environment. Kelly et al. (2004) asked observers in the real world or in a virtual environment to judge whether another agent could see a given target in the environment. The scene was purposefully chosen (or created in VR) so that there was an occluding building, and the viewer was given instructions to judge which parts of the scene were visible from the other's viewpoint and which were occluded by the building. They indicated this on a photograph of the scene (in the real world) or by pointing to the location in the virtual world. Viewers were generally good at this task across both environments, but overestimated what the agent could see as the distance between the viewer and the agent increased from 5 to 10 to 15 m. This work suggests that PT-1 may utilize an allocentric frame of reference in which observers visually match various distances and angles to infer the line of sight of another.

In contrast, PT-2 typically requires an observer to identify where in space a target object is located relative to a viewpoint that is different from the observer's current viewpoint. For example, in early work on imagined and real transformations, Rieser (1989) asked participants to learn the location of an array of objects while standing in the middle of the array. While blindfolded, they were asked to point in the direction of a named target from a new imagined viewpoint. Then they were asked to imagine facing in a new direction (rotation task) or to imagine moving to a new target location while continuing to face in the same cardinal direction (translation task). This and other work (e.g., Presson and Montello, 1994; May, 2007) showed a robust angular disparity effect in the imagined rotation task, such that reaction time increased with the increasing disparity between one's actual facing and imagined facing direction. This was significantly different from the virtually flat response time function found in real rotations, suggesting a cost to perform the mental transformation to judge what the spatial layout looked like outside of one's physical viewing perspective. From this work, Rieser (1989) and Presson and Montello (1994) suggested that the angular disparity effect found in PT-2 tasks is due to the increased processing involved in updating self-to-object relationships.

May (2004, 2007) suggested that the angular disparity effect may be due to a conflict of sensorimotor codes. Specifically, a conflict in sensorimotor codes occurs between codes that help identify the location of a target object from the to-be-imagined viewpoint, and the codes that help the observer actually make a pointing response. This was initial evidence that PT-2 involves a shift from one egocentric frame of reference to another egocentric frame of reference. Kessler and Thomson (2010) provided additional support for the use of egocentric

reference frames during PT-2 by showing that the observer may actually imagine rotating her body axes to align with the to-be-imagined perspective. They asked participants to indicate whether an object was located to the left or the right of an avatar situated at 0, 40, 80, 120, or 160° around a circular table with respect to the participant's viewpoint. Importantly, the authors situated the participants at the computer such that their bodies were either facing straight ahead toward the monitor, or at a 40° angle from the monitor. They found an overall effect of body posture that increased monotonically with angular disparity. In other words, observers switch from their current egocentric viewpoint to the egocentric viewpoint of another person in space in order to mentally transform their body axes through the space. May and Wendt (2012) have more recently pointed out that some egocentric mental transformation tasks also face stimulus-response compatibility effects, where spatial conflict may contribute to the apparent mental transformation effects.

Overall, the difference between visibility tasks (PT-1) and determining spatial relationships from a new perspective (PT-2) may be the object relations that are used. Inter-object relations may be used to determine whether something is visible from another's perspective. However, when updating to a new left/right position relative to that perspective, rotation of the viewer's frame of reference is needed. In support of this claim, several have found that left/right decisions involve increasing response time with increasing angular disparity, whereas visibility/front back decision show relatively flat response time functions as a function of angular disparity but increasing response time as a function of distance between the agent and the target (Michelon and Zacks, 2006; Kessler and Rutherford, 2010). In summary, PT-1 appears to rely on an allocentric frame of reference, determining the location of an object with respect to another's viewpoint whereas often PT-2 relies on the *transformation* of the egocentric reference frame onto the other's viewpoint, in order to update object spatial relations with respect to the new viewpoint.

OTHERS AND SPATIAL PERSPECTIVE TAKING

Both PT-1 and PT-2 can contribute to a viewer's ability to predict the behavior of others. Several examples come from the study of spatial language in which different frames of reference may be used to produce spatial descriptions to a partner depending on the social context. Generally, these studies show that attributional cues about the partner influence how people interpret and produce spatial descriptions. When speakers perceive that partners have less knowledge or relevance to the task—due to a number of factors such as lower spatial abilities, less familiarity, less agency, or less information about the viewpoint—then speakers are more likely to take a partner-centered frame of reference (Schober, 2009; Duran et al., 2011; Galati et al., 2013). In other words, when the observer realizes there is less of a shared perspective, they will adjust their language to meet the needs of the partner. When the partner's goals, realism/presence, or shared mutual understanding increase, then speakers are more likely to use their own egocentric perspective.

Further, in a simple, but elegant manipulation of the visual presence and goals of an agent, Tversky and Hard (2009) showed that the presence of another person in a scene changed the way people described the left/right relationship between two objects. Observers viewed a photograph of two objects on a table, with or without a person seated across the table either looking at or reaching for one of the objects. The frequency of reporting the relationship of the two objects from the other's perspective increased with the presence of the person, and increased further when the question referred to action. These results suggest that even outside of an explicit communication task, viewers will spontaneously take the perspective of another person. Spontaneous perspective taking was also seen in Samson et al. (2010), who required a viewer to judge (in a picture) how many discs on a wall could be seen from their own perspective or from an avatar's perspective (a PT-1 visibility task). The number of discs that the avatar could see was either consistent or inconsistent with the number of discs from the viewer's egocentric perspective. Viewers were slower to make their egocentric judgments when there was a conflict with the avatar's perspective, even though the avatar perspective had no direct relevance to their task.

Consistent with these results, implicit perspective taking has also been shown with an action-based mimicry task. For example, participants viewed a virtual tight-rope walking avatar while they were simultaneously asked to imagine also being on a tight-rope (Thiriaux et al., 2009). The participants were told to lean the way the avatar was leaning, not specifying whether to lean as if the avatar was a mirror reflection, or to lean as if they were in the shoes of the avatar. The study found that the participants adopted the viewpoint of the avatar instead of mirroring the avatar nearly 70 percent of the time.

Many of these studies tend to naturally confound body orientation or depicted action with eye gaze. Mazzarella et al. (2012) decoupled action and eye gaze in stimuli depicting another agent to assess when perspective taking would occur. In contrast to Tversky and Hard (2009), they first used an explicit perspective taking task in which participants were instructed to report target location from either an egocentric perspective or the agent's perspective. Participants viewed scenes with an agent positioned across the table with an object. The scenes varied as to whether the agent looked at or grasped the object. Given the explicit task of taking an egocentric or allocentric frame of reference, it is not surprising that viewers made few allocentric errors in the egocentric condition. However, the results also showed that in the explicit allocentric condition, viewers were better in their allocentric judgments when the actor was depicted as grasping the object, with no significant influence of eye gaze. A third experiment distinguished between the effects of grasping and gaze on perspective taking and attentional orienting. When the task was to detect an object after being presented with the agent-in-action/agent-gaze images, participants were faster with the gaze image than the action image. These results suggest that gaze and body/action information may provide different information about others' intentions. Arm/body cues may be more useful in communicating current

goals and eye gaze may indicate what the actor will do in the future.

SPATIAL PERSPECTIVE TAKING WITH OTHERS: SUMMARY AND OPEN QUESTIONS

Overall, the work reviewed on spatial perspective taking with others describes two types of tasks, Level-1 and Level-2, which are both elicited in the context of another agent. First, this work suggests that observers may identify the intentions of another by considering where they are looking (PT-1). Second, this work suggests that the body of the other may indicate current goals of the actor while the eye gaze of the actor may denote future goals. Both could be used to understand the intentions of others. Finally, the work reviewed suggests that PT-1 uses an allocentric frame of reference while PT-2 involves shifting from one egocentric reference frame to another's egocentric reference frame.

Much of the spatial perspective taking research has been designed to understand spatial memory, language, navigation, and overall spatial cognition. However, very little of this work has considered the broader social function of spatial perspective taking—predicting other's behavior in the service of coordinating actions. If spatial perspective taking operates in conjunction with perceiving affordances for others, it may have evolved to help us infer an intention or goal for another person. When used alongside the ability to judge this other person's action capabilities, both may allow humans to make fairly accurate predictions about what another person is likely to do next. In turn, observers are able to adjust their own actions to coincide, cooperate, or compete with another person's current and future behaviors.

SPATIAL PERSPECTIVE TAKING AND PERCEIVED ACTION CAPABILITIES MUTUALLY INFORM BEHAVIOR PREDICTION **HOW LEVEL-1 PERSPECTIVE TAKING AND JUDGING AFFORDANCES FOR OTHERS MAY WORK TOGETHER**

Gibson (1979) argued that all of the information necessary to judge affordances is available to any point of observation (see also Stoffregen et al., 1999; Mark, 2007). Likewise, information specifying one's line of sight is also available in the optic array. Both PT-1 and perceiving affordances for others utilize an allocentric frame of reference because both processes can be carried out using object-to-object relationships, likely with a visual matching strategy. Although it is unknown how humans (or other species) determine where another is looking, it is plausible that visual information regarding the direction of one's gaze is combined with perceptual information identifying the distance and depth of objects in the environment (see Kelly et al., 2004 for a similar view). Together, it may be possible for an observer to see another person and simultaneously know (1) where they are looking and (2) what actions they are capable of performing given the properties of the environment. This would suggest that the line of sight operates to orient the observer's attention to the properties of the environment that must be considered alongside the bodily capabilities of the other person. Such a process is consistent with Kugler and Turvey's (1987) definition of an intention being an attribution that an observer projects on to another person to simplify what behaviors might be expected from this person. They use an example in physics, in which temperature and pressure are

concepts used to understand collective properties of molecules. The temperature of a substance is attributed to the molecules by the observer in an attempt to describe higher level processes when describing the individual movement of each molecule is cumbersome. Much the same, attributing intentions to an actor, is a method by which an observer attempts to reduce the many possible actions available to an actor to a subset few and in so doing describes the demands of the environment that are placed on the actor. Future research should consider testing the possibility that Level-1 perspective taking occurs when attempting to predict the behavior of others.

HOW LEVEL-2 PERSPECTIVE TAKING AND JUDGING AFFORDANCES FOR OTHERS MAY WORK TOGETHER

Level-2 perspective taking is distinguishable from Level-1 based on the extent to which observer-centered spatial transformations are needed (as discussed above). PT-2 reveals to an observer the spatial relationship between a person and objects in the environment. For example, you can sit across the table from a friend, and while your friend's cup may be on your right-hand side, you are able to identify that the cup is on your friend's left-hand side. There are many different models that attempt to account for this ability to discriminate one's own perspective from another. Overwhelmingly the evidence suggests that the observer must imagine a rotation of their body axes or frame of reference, possibly involving the motor, proprioceptive, or vestibular system to accomplish this task (Grabherr et al., 2007; Kessler and Thomson, 2010). PT-2 requires that the observer transform their own egocentric frame of reference to the egocentric frame of reference of another person. This is different from how reference frames are utilized when perceiving affordances for others, as judging another's affordances likely involves a shift from the observer's egocentric frame of reference to an allocentric (other-to-object) frame of reference.

Regardless of the use of different frames of reference, the intentions of another actor may still be inferred through PT-2 when an asymmetry exists between the other's left and right side. For example, if another person is holding a rod in their right hand, their ability to reach to objects differs for their right and left sides (Linkenauger et al., 2009). Thus, one could infer that the actor is more likely to reach with her right hand, an understanding that may be critical for a task involving joint action. However, when a distinction between what is on the left or right of an actor is not needed, PT-2 processes are not likely relied on for judging affordances of others. Instead, the observer can visually match the length of the actor's arm (or arm plus rod) to the distance between the actor and some object, thereby inferring what the actor can do by using an allocentric reference frame from the observer's viewpoint. However, PT-2 perspective taking could be integral for successful communication in which two or more people need to create a common conception of the space (Duran et al., 2011). In addition, PT-2 perspective taking appears to be closely related to path integration during navigation, and developing a geocentric view (bird's eye view) of the space (Loomis et al., 1999). In conclusion, it may be the case that PT-2 perspective taking is not used when determining the intentions of other people unless future coordination is required.

SELF AND OTHER AFFORDANCES MUTUALLY INFORM BEHAVIOR PREDICTION

There are instances in which information about the observer may be used to understand the capabilities of another, and conversely, instances where the capabilities of another influence actions or judgments about the self. For example, in joint action, previous research suggests that observers consider not only their own action capabilities, but also the action capabilities of another person (Sebanz et al., 2006). Even when joint action is not an explicit goal, recent evidence suggests that judging affordances for oneself can be influenced by the action capabilities of another person (Gagnon et al., in preparation). In our own recent work (Gagnon et al., in preparation) we examined both the influence of one's own body size on affordance judgments for another, and the influence of another's size on self-judgments. In a paradigm using judgments of passage through apertures, we found that the judgments for another are scaled to the other's body size, but that there is an additional mutual influence of the self on other judgments and the other on self-judgments.

In addition, Costantini et al. (2011) tested the influence of the affordances of another agent on the spatial alignment effect paradigm (Bub and Masson, 2010)—an effect showing that action-relevant but task-irrelevant objects will facilitate actions when the object is congruent with the action. Previous work showed that in a desktop virtual environment, the presentation of a mug facilitated a grasp response, but only when it reachable by the actor as depicted in the virtual scene (Costantini et al., 2010). Costantini et al. (2011) extended this paradigm and found that the viewer's motor facilitation also occurred when the object was outside of the viewer's reachable space but *within an agent's* reachable space. They suggest that the space in which the actor can perform an action might be "mapped on" to the observer's bodily spatial representation, influencing the observer's own potential to act. This could inform an observer about how another agent perceives a space and capability for action, as well as providing information for joint action (Costantini et al., 2011).

Related spontaneous use of another's potential for action has been demonstrated in a distance judgment task that varied the extent to which another agent could reach a target (Bloesch et al., 2012). Bloesch et al. proposed that if using a tool makes a distance appear closer (see also Witt et al., 2005), then it may be that watching another agent use a tool also influences perceived distance. These predictions held true; observers who watched another actor reach successfully to a target with a reach-extending tool judged the distance to be closer than those who watched an unsuccessful arm-based reach.

As social beings, the mere presence of another person may prompt humans to share (implicitly or explicitly) spatial and proprioceptive information with each other. Oullier et al. (2008) found that when two people see each other performing the same action, they spontaneously synchronize their actions, suggesting a means of information exchange that could coordinate actions. Whether these examples are an instance of a transformation of one's egocentric frame of reference is unknown. Regardless, this work suggests that spatial and proprioceptive information is not necessarily confined to the physical

boundaries of a person, but can be shared amongst two or more people.

POSSIBLE OVERLAPPING MECHANISMS SUPPORTING OTHER'S AFFORDANCES AND SPATIAL PERSPECTIVE TAKING

Given the relationship between judging others' affordances and spatial perspective taking is somewhat unclear from the behavioral work, it may be useful to consider whether the process of judging other's affordances and spatial perspective taking share overlapping processes relying on motor simulation. First, we will review the proposed mechanisms involved in perspective taking, and then relate this to the potential mechanisms involved in perceiving affordances for others.

MECHANISMS FOR SPATIAL PERSPECTIVE TAKING

One explanation for the angular disparity effects present in spatial updating after imagined rotations is sensorimotor interference. Despite evidence for the need for mental transformation of the egocentric reference frame (Rieser, 1989; Presson and Montello, 1994; Easton and Sholl, 1995; Wraga et al., 2000), costs in perspective taking have been attributed to a response-based conflict between one's real and imagined perspective. This is especially apparent in pointing tasks where the correct response is incompatible with the viewer's current physical proprioceptive information for facing orientation (Wraga, 2003; Avraamides et al., 2007) and has been shown to be reduced by disorienting participants before the response (May, 1996). Taken together, this work suggests that sensorimotor processes may underlie spatial perspective taking given the disparity in imagined and real locations influences task performance.

Recent work suggests the influence of the vestibular system in imagined perspective taking as well (Mast et al., 2007). For example, van Elk and Blanke (2013) asked participants to perform imagined viewer rotation while being passively rotated clockwise or counterclockwise. By passively rotating the participants the authors were able to separate some of the proprioceptive cues used in active rotation from the vestibular signals. When the participants were being passively rotated in the same direction that they imagined rotating their viewpoint, reaction times were faster than when the passive rotation was incongruent to the imagined rotation direction. Grabherr et al. (2011) compared patients with unilateral and bilateral vestibular loss on egocentric and object mental transformation tasks. They found that those with bilateral loss showed significantly poorer performance in the egocentric transformation task than unilateral loss patients. In healthy participants, galvanic vestibular stimulation (GVS, direct electrical stimulation of vestibular end organs) has been shown to lead to poorer performance on imagined viewer rotation (Grabherr et al., 2007; Lenggenhager et al., 2007; Dilda et al., 2012).

There are other accounts that may better explain certain types of perspective taking tasks, such as the visibility tasks (PT-1) described above. For example, there is evidence that for visibility tasks, judgments about what another can do may be solved based on visual-spatial processing that do not require a shift to an imagined viewpoint (Kelly et al., 2004; Michelon and Zacks, 2006; Kessler and Rutherford, 2010). Predicting whether an object is visible from another agent's viewpoint is likely performed

without a transformation of one's egocentric frame of reference. Rather, the answer can be computed based on an object-to-object based strategy, where a mental line is constructed from the agent to the target. While a viewer-based transformation could be used to solve the task, the lack of an angular disparity effect suggests that the line-of-sight computation is used. There is little evidence in support of any body-based simulation underlying this type of judgment. An open question for the current paper is how mechanisms for spatial perspective taking may or may not be related to affordances and how they may work together to coordinate action.

Several of the mechanisms proposed for spatial perspective taking involve sensorimotor processing. Likewise, one dominant account for the understanding of other's actions—particularly the observation of other's overt actions—is also framed in the motor system. If perceiving other's affordances and spatial perspective taking rely on similar mechanisms, then this suggests that they may be functionally related with respect to social coordination. While on one hand motor simulation may underlie both processes, we must concede that it is possible that it does not account for either process. As described above, there is relatively strong support for the use of perceptual information available to the other, not the self, in judging other's affordances. Further, there is evidence that non-motor, visual-spatial processing may be used for at least some Level-1 (Kessler and Rutherford, 2010) and Level-2 (Amorim et al., 2006; Creem-Regehr et al., 2007) perspective taking tasks. We consider the evidence for both motor simulation and non-simulation/visual-information based accounts of perceiving other's affordances below.

MECHANISMS FOR PERCEIVING AFFORDANCES FOR OTHERS

Gibson's (1979) concept of affordances and much of the work following this theoretical viewpoint was concerned with characterizing perception at the level of the observer-environment system. As with any psychological process, one may ask how the process is supported by our biology. While the theory of affordances did not attempt to address questions about the underlying neurocognitive mechanisms involved, there is a related notion of object-based affordances, alluded to in the work of Costantini et al. (2011) above, which elicits motor system activation and could help to explain the mechanisms underlying the prediction and use of other's affordances. Numerous studies with objects have shown that affordances may be automatically activated and lead to subsequent effects on the motor system. For example, a classic behavioral study by Tucker and Ellis (1998) showed a response compatibility effect. When presented with images of objects with handles, responses to an irrelevant stimulus feature were facilitated when the handle orientation was congruent with the hand used to make the response. Neuroimaging has supported this claim, showing that activation of related premotor and parietal cortex results from simply viewing objects such as tools that have affordances (Chao and Martin, 2000; Creem-Regehr and Lee, 2005). It is important to note, however, that goal context has been shown to be important in modulating activity across both cognitive and neural approaches. Buxbaum and Kalenine (2010) provide compelling examples of how motor resonance may only occur in the context of goal-directed, functional representations of objects, rather than simply the structure of the object itself (see

also, Creem-Regehr et al., 2007). Can the neurocognitive notion of object affordances (mostly focused on grasping) be extended to environmental affordances such as those for passing through and sitting? We discuss this possibility in terms of motor resonance theory below.

MOTOR RESONANCE

The “mirror neuron” system is a specific brain mechanism proposed to underlie motor simulation during action observation. Mirror neurons were identified initially in the ventral premotor and parietal cortices of the macaque monkey. They activate both when the monkey performs an action as well as when the monkey observes another human or monkey perform the same action (Gallese et al., 1996; Rizzolatti and Craighero, 2004). A body of work has proposed an analogous system in humans, including the premotor cortex, inferior parietal cortex, and superior temporal sulcus, with specificity to the level of somatotopic representation of specific body parts (Buccino et al., 2001) and tuning to the actual motor capabilities and experiences of the actor (Calvo-Merino et al., 2005, 2006). Subsequent research has defined some mirror neurons as goal-related rather than effector-specific (Fogassi et al., 2005; Rochat et al., 2010). For example, Fogassi et al. (2005) found mirror neurons in the monkey inferior parietal lobule that responded to observation of the same grasping action differentially *as a function of the goal of the action*. Neurons were selective for the goals of grasping-to-eat vs. grasping-to-place. Similarly, in humans, Iacoboni et al. (2005) varied whether an observed grasping movement was performed in the context of goals of drinking or cleaning up. Premotor cortex activity was modulated by the context and intention of the grasp depicted. The importance of understanding a hierarchy of goals has been emphasized by several researchers (Grafton and Hamilton, 2007; Thill et al., 2013). Also, when performing a joint action, there is neural activity associated with coordinated ($\phi 2$) and independent ($\phi 1$) behavior. Topographically, this activity maps well to the mirror neuron system, and $\phi 1$ (independent behavior) may indicate inhibition of the mirror neuron system (Tognoli et al., 2007). Many have proposed that we understand the actions of others by means of a motor or embodied simulation system, although these claims have also stirred much debate. How then, might this mirror system support the judgments of what others can do and see?

The term motor resonance refers to the matching of one's own action to another's (Uithol et al., 2011). As Uithol et al. (2011) described, the term “resonance” comes from the physical phenomenon that two systems oscillate and at the same frequency and phase as one another. However, in the neurocognitive context of mirror neuron systems, resonance is used more broadly to describe a mechanism of emulation, in which viewing an action performed by another leads to activation of neurons in the viewer that represent that action. Viewers understand actions by matching or simulating the action. Furthermore, the analysis by Uithol et al. (2011) differentiates between intrapersonal resonance and interpersonal resonance—a distinction that may be important for the extension to judging other's affordances. Intrapersonal resonance occurs within an individual: a perceptual representation of observed action is activated and at the same time coupled

with a motor representation (Rizzolatti et al., 2001). This notion is supported by the common coding theory (Hommel et al., 2001) in which perception and action share common underlying representations. In interpersonal resonance, there is a functional equivalence between the motor representation of the observer and the actor, emphasizing shared goals or action plans across the two actors (Wilson and Knoblich, 2005).

Although there is an extensive literature on the mirror system mechanisms involved in observation of actions (e.g., Fadiga et al., 1995; Decety et al., 1997; Johnson-Frey et al., 2003; Iacoboni et al., 2005), the problem posed by this review is somewhat different. In most cases of explicit or implicit use of other's affordances and of spontaneous use of another's viewpoint in perspective taking, there is no overt movement of the other agent. It is possible that observers use intrapersonal motor resonance to not only emulate actions, but also to infer and predict future actions (Wilson and Knoblich, 2005; Sebanz et al., 2006). Specifically, experience and capabilities or current bodily state could be used to predict the actions of others. Bosbach et al. (2005) showed the importance of one's proprioceptive body information on action understanding by demonstrating that individuals with impaired sense of touch and proprioception failed to understand another's expectation of weight when observing the action (see also Reed and Farah, 1995; Daems and Verfaillie, 1999 for posture-based effects). Knoblich and colleague's proposal that the observer serves as an initial model for understanding and predicting action could explain some of the results discussed so far. For example, the influence of wearing ankle weights on judging other's jumping ability would relate one's own action capability to judgments for another's capabilities (Ramenzoni et al., 2008a). Likewise, the capability of another agent to reach or not reach a mug could influence one's own likelihood of reaching the mug, leading to more or less priming of the motor system (Costantini et al., 2011). This claim is supported by more recent work (Cardellicchio et al., 2013) which used transcranial magnetic stimulation (TMS) to record the motor-evoked potentials (MEPs) of observers. In a virtual environment display, a mug was presented either within or outside of the observer's reaching space and within or outside of an agent's reaching space. Highest MEPs were measured when the mug was within either the observer's reaching space or the agent's reaching space, compared to when the mug was outside of the observer's reaching space or close to a non-body cylinder (which took the place of the avatar/agent). Finally, in joint actions, there could be neural representations for action based on each actor's capabilities that mutually activate in order to support complementary actions.

INFORMATION-BASED ACCOUNTS

An alternative account of self-other interactions comes from the ecological viewpoint, emphasizing the direct information about the environment available to the viewer. As mentioned in the introduction, this account is not necessarily exclusive of the motor resonance account, but it emphasizes different aspects of the processes of social perception-action. As described earlier, Ramenzoni et al. (2008b) found that viewers used eye-height scaled information to judge accurately what others could reach, suggesting that judging other's affordances relies on viewer-scaled

optical information. Indeed, Ramenzoni et al. (2010) argued that the motor resonance account proposes a “strong dependency on the observer’s own action capabilities” (p. 1117) that is not necessarily supported by the empirical findings. Accounts based in motor simulation place an emphasis on the attributes of the perceiver in judging other’s affordances, rather than the situated perceptual information available to the other agent. In many cases, studies of judging other’s affordances have shown the importance of the perceptual information available to the agent, in contrast to a reliance on the perceiver’s capabilities.

A possible mechanism for this direct use of environmental information may be explained by the *synergistic* approach (Riley et al., 2011). In this approach, observers are thought to be able to coordinate actions with others through a process of reducing each other’s degrees of freedom in movement (dimensional compression) and reacting to the movements of one another (reciprocal compensation) to create a single coordinated system (Riley et al., 2011). The synergistic approach extends the work of Nikolai Bernstein in motor coordination. Bernstein identified that one major problem for any movement system, such as the human body, is in regulating all the possible degrees of freedom inherent to it (e.g., joints, muscle extension/flexion, etc.). Bernstein (1967) proposed that these degrees of freedom may couple together to create a synergy. By allowing for synergies, the overall degrees of freedom are reduced allowing the movement system to work as a single unified system. In applying the synergistic approach at the interpersonal level, Riley et al. (2011), consider how two individuals couple their actions to produce a synergy that ultimately constrains the degrees of freedom in the movement of each individual. Because viewers have access to concurrent visual information from multiple viewpoints and can judge affordances for another with respect to the other’s bodily information in the context of the environment, they can also interpersonally coordinate actions. Overall, the synergistic approach describes a process that allows observers to couple their movements with those of others, which gives rise to dynamic changes that are not independent in the two systems (see Kugler and Turvey, 1987; Turvey and Carello, 1996).

The synergistic approach may better explain phenomena such as understanding the interpersonal exchange in conversations (Condon and Ogston, 1971) and similar affect in interactions between mothers and their children (Cohn and Tronick, 1988) than the motor resonance approach. More related to the current paper, Ramenzoni (2008) asked participants to coordinate holding a stick inside a hollowed circle. When circle size was varied, the task became more or less difficult and as a result, participants’ hand and torso movements were more or less coordinated (see also Riley et al., 2011). The main difference between this approach and that of Sebanz et al. (2006) is the claim that actors’ movements in a coordinated action are not independent of one another, rather they coordinate to form a new entity with which to judge affordances. As such, the motor resonance approach may predict dimensional compression, but it cannot account for reciprocal compensation due to the assumption that the mirror neurons systems of two individuals are independent of one another (Riley et al., 2011). In addition, this approach does not focus on fixed neurological structures causing the activity of other

structures; rather it focuses on the functionality that arises when many neurological structures interact or couple together, reflecting Bernstein’s (1967) original approach to understanding motor coordination.

CONCLUSIONS

Perceiving other’s affordances and spatial perspective taking are two abilities that have traditionally been studied in the domains of perception and spatial cognition, respectively. While typically considered separate abilities, they share a common conceptual foundation of relating self and other perspectives in some way. An observer must determine how another agent can act or see the world. While these are skills that are important fundamentally for an understanding of our spatial environment, we argue that when considered together, they provide a basis for a broader social function of human behavior prediction critical to our social coordination with others. In this paper we aimed to provide a review of the work carried out on other’s affordances and perspective taking to show how they are related in the service of understanding both the actions and intentions of others.

Judging other’s affordances is a means to determine capability for future action. The literature reviewed shows that in circumstances of a single other agent, or in dyads, observers are relatively good at perceiving affordances for others when provided with enough information to scale judgments to the other’s body. However, we have proposed that these laboratory-based affordance judgments are typically more specified in terms of an action-goal than what occurs in the real world where the other’s goal may not be as specified. To solve this problem and identify another’s intentions, the ability of spatial perspective taking may come into play, allowing an observer to further define the intention and goal of the other actor. Support for these two components as complementary processes comes from an analysis of the similarities between the two, on both computational and neural mechanism levels.

An analysis of frames of reference recruited shows us that there are at least three possible frameworks used. The viewer may use their own egocentric frame (as used in judgments of self-affordances), which may also include a reliance on their own possibilities for action when judging for others; alternatively, a viewer’s egocentric frame of reference may be transformed onto the other’s frame of reference, aligning the self and other reference frames, typically used in PT-2 tasks; finally, the viewer may simply use an allocentric frame of reference, computing the relationship between the other and the target object/environment. Current work suggests more overlap in the allocentric computation used in perceiving other’s affordances and PT-1; however, more work is needed to determine whether egocentric spatial transformations may be involved in some affordance judgments. Future studies addressing this question could assess the possible transformation of the egocentric frame by measuring angular disparity effects during explicit or implicit affordance judgments with respect to other agents.

An analysis of motor resonance theory suggests that the sensorimotor mechanisms supporting some forms of perspective taking and perceiving other’s affordances may overlap. This is particularly apparent in circumstances in which there is no

available visual information to make judgments of affordances or perspective—e.g., insufficient information about kinematics or the need for updating of spatial relations in a viewer-centered framework. In these cases, viewers may use motor simulation to judge the capabilities or perspective of others. Furthermore, the spontaneous and mutual influence of another agent and the self, seen in both affordance judgments and perspective taking, also is consistent with shared spatial and proprioceptive information among two people, as well as shared motor processing. In all, we suggest that judging affordances and spatial perspective rely on a combination of direct visual information and motor resonance.

Finally, we have considered how the broader goal of social cognition could be served by two spatial processes, but it is also important to consider the possibility of the inverse. Does social context itself moderate the abilities of perceiving other's affordances and perspective? The underlying rationale is that in order to perform a spatial switch of perspectives, one must understand that other agents have different perspectives. Thus, having a "theory of mind" could be a prerequisite to spatial perspective taking. The influence of social skills on spatial perspective taking has been shown in a number of ways. First, individuals with autism spectrum disorder (ASD) have been studied as a population that is defined with social impairment. Hamilton et al. (2009) showed a subtle distinction between performance on two mental rotation tasks in ASD children, finding impairment on a perspective rotation condition in which the decision required was with respect to what another person could see, but not on an object-rotation condition. Shelton et al. (2012) investigated the influence of social skills on perspective taking by testing a healthy non-clinical population, but using a questionnaire to assess traits

of ASD. In a version of Piaget's three mountain task, they asked observers to choose a picture of a display as it would appear from another perspective. The location of the other's perspective was indicated either by a triangle, camera, or a doll. They found that perspective taking performance was modulated by social skills, but only for the doll, such that better social skills were associated with better perspective taking. Similarly, Kessler and Wang (2012) found that differences in perspective taking emerged as a function of both sex and social skills.

While not directly the same task as the mostly static affordance or spatial judgments focused on in this paper, there is also a recent literature on the influence of social context of others on executed actions. For example, reach-to-grasp kinematics are different when passing an object to a partner compared to placing it in a new location (Becchio et al., 2008) and implicit social requests for an object have been shown to override an initial motor plan (Sartori et al., 2009). Together, this work emphasizes the importance of social context on action planning and the flexibility in online adjustments in action that occur with potential social interactions.

Clearly, there is a need to consider what may seem to be disparate areas of research to understand complex human behaviors, such as social coordination and joint action. This review provides one example for which research on two distinct spatial processes—judgments of others' affordances and spatial perspective taking—may be examined to elucidate potential mechanisms for more complex behaviors.

ACKNOWLEDGMENTS

This work was partially supported by National Science Foundation Grants 0914488 and 1116636.

REFERENCES

- Alexander, R. D. (1990). *How did Humans Evolve? Reflections on the Uniquely Unique Species*. Ann Arbor, MI: Museum of Zoology (Special Publication No. 1); The University of Michigan.
- Amorim, M. A., Isableu, B., and Jarraya, M. (2006). Embodied spatial transformations: "body analogy" for the mental rotation of objects. *J. Exp. Psychol. Gen.* 135, 327–347. doi: 10.1037/0096-3445.135.3.327
- Avraamides, M. N., Ioannidou, L. M., and Kyranidou, M. N. (2007). Locating targets from imagined perspectives: comparing labelling with pointing responses. *Q. J. Exp. Psychol.* 60, 1660–1679. doi: 10.1080/17470210601121833
- Becchio, C., Sartori, L., Bulgheroni, M., and Castiello, U. (2008). The case of Dr. Jekyll and Mr. Hyde: a kinematic study on social interaction. *Conscious. Cogn.* 17, 557–564. doi: 10.1016/j.concog.2007.03.003
- Bernstein, N. A. (1967). *Coordination and Regulation of Movements*. New York, NY: Pergamon Press.
- Bloesch, E. K., Davoli, C. C., Roth, N., Brockmole, J. R., and Abrams, R. A. (2012). Watch this! Observed tool use affects perceived distance. *Psychon. Bull. Rev.* 19, 177–183. doi: 10.3758/s13423-011-0200-z
- Bosbach, S., Cole, J., Prinz, W., and Knoblich, G. (2005). Understanding another's expectation from action: the role of peripheral sensation. *Nat. Neurosci.* 8, 1295–1297. doi: 10.1038/nn1535
- Bub, D. N., and Masson, M. E. (2010). Grasping beer mugs: on the dynamics of alignment effects induced by handled objects. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 341. doi: 10.1037/a0017606
- Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., et al. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *Eur. J. Neurosci.* 13, 400–404. doi: 10.1111/j.1460-9568.2001.01385.x
- Buxbaum, L. J., and Kalenine, S. (2010). Action knowledge, visuomotor activation, and embodiment in the two action systems. *Ann. N.Y. Acad. Sci.* 1191, 210–218. doi: 10.1111/j.1749-6632.2010.05447.x
- Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E., and Haggard, P. (2005). Action observation and acquired motor skills: an fMRI study with expert dancers. *Cereb. Cortex* 15, 1243–1249. doi: 10.1093/cercor/bhi007
- Calvo-Merino, B., Grèzes, J., Glaser, D. E., Passingham, R. E., and Haggard, P. (2006). Seeing or doing? Influence of visual and motor familiarity in action observation. *Curr. Biol.* 16, 1905–1910. doi: 10.1016/j.cub.2006.07.065
- Cardellicchio, P., Sinigaglia, C., and Costantini, M. (2013). Grasping affordances with the other's hand: a TMS study. *Soc. Cogn. Affect. Neurosci.* 8, 455–459. doi: 10.1093/scan/nss017
- Carello, C., Groszofsky, A., Reichel, F. D., Solomon, H. Y., and Turvey, M. T. (1989). Visually perceiving what is reachable. *Ecol. Psychol.* 1, 27–54. doi: 10.1207/s15326969eco0101_3
- Chang, C.-H., Wade, M. G., and Stoffregen, T. A. (2009). Perceiving affordances for aperture passage in an environment-person-person system. *J. Mot. Behav.* 41, 495–500. doi: 10.3200/35-08-095
- Chao, L. L., and Martin, A. (2000). Representation of manipulable man-made objects in the dorsal stream. *Neuroimage* 12, 478–484. doi: 10.1006/nimg.2000.0635
- Cohn, J. F., and Tronick, E. Z. (1988). Mother-infant face-to-face interaction: influence is bidirectional and unrelated to periodic cycles in either partner's behavior. *Dev. Psychol.* 24, 386–392. doi: 10.1037/0012-1649.24.3.386
- Condon, W. S., and Ogston, W. D. (1971). "Speech and body motion synchrony of the speaker-hearer," in *The Perception of Language*, eds D. Horton and J. Jenkins (Columbus, OH: Charles E. Merrill), 150–184.
- Cordovil, R., Andrade, C., and Barreiros, J. (2013). Perceiving children's affordances: recalibrating estimation following single-trial observation of three different tasks.

- Hum. Mov. Sci.* 32, 270–278. doi: 10.1016/j.humov.2013.01.001
- Cordovil, R., and Barreiros, J. (2010). Adults' perception of children's height and reaching capability. *Acta Psychol.* 135, 24–29. doi: 10.1016/j.actpsy.2010.04.008
- Cordovil, R., and Barreiros, J. (2011). Egocentric or allocentric frameworks for the evaluation of other people's reachability. *Hum. Mov. Sci.* 30, 976–983. doi: 10.1016/j.humov.2010.08.011
- Costantini, M., Ambrosini, E., Tieri, G., Sinigaglia, C., and Committeri, G. (2010). Where does an object trigger an action? An investigation about affordances in space. *Exp. Brain Res.* 207, 95–103. doi: 10.1007/s00221-010-2435-8
- Costantini, M., Committeri, G., and Sinigaglia, C. (2011). Ready both to your and to my hands: mapping the action space of others. *PLoS ONE* 6:e17923. doi: 10.1371/journal.pone.0017923
- Creem-Regehr, S. H., Dilda, V., Vicchirilli, A. E., Federer, F., and Lee, J. N. (2007). The influence of complex action knowledge on representations of novel graspable objects: evidence from functional magnetic resonance imaging. *J. Int. Neuropsychol. Soc.* 13, 1009–1020. doi: 10.1017/S1355617707071093
- Creem-Regehr, S. H., and Lee, J. N. (2005). Neural representations of graspable objects: are tools special? *Cogn. Brain Res.* 22, 457–469. doi: 10.1016/j.cogbrainres.2004.10.006
- Creem-Regehr, S. H., Neil, J. A., and Yeh, H. J. (2007). Neural correlates of two imagined egocentric transformations. *Neuroimage* 35, 916–927. doi: 10.1016/j.neuroimage.2006.11.057
- Daems, A., and Verfaillie, K. (1999). Viewpoint-dependent priming effects in the perception of human actions and body postures. *Vis. Cogn.* 6, 665–693. doi: 10.1080/135062899394894
- Davis, T. J., Riley, M. A., Shockley, K., and Cummins-Sebree, S. (2010). Perceiving affordances for joint actions. *Perception* 39, 1624. doi: 10.1068/p6712
- Decety, J., Grezes, J., Costes, N., Perani, D., Jeannerod, M., Procyk, E., et al. (1997). Brain activity during observation of actions. Influence of action content and subject's strategy. *Brain* 120, 1763–1777. doi: 10.1093/brain/120.10.1763
- Dilda, V., MacDougall, H. G., Curthoys, I. S., and Moore, S. T. (2012). Effects of galvanic vestibular stimulation on cognitive function. *Exp. Brain Res.* 216, 275–285. doi: 10.1007/s00221-011-2929-z
- Duran, N. D., Dale, R., and Kreuz, R. J. (2011). Listeners invest in an assumed other's perspective despite cognitive cost. *Cognition* 121, 22–40. doi: 10.1016/j.cognition.2011.06.009
- Easton, R. D., and Sholl, M. J. (1995). Object-array structure, frames of reference and retrieval of spatial knowledge. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 438–500. doi: 10.1037/0278-7393.21.2.483
- Fadiga, L., Fogassi, L., Pavesi, G., and Rizzolatti, G. (1995). Motor facilitation during action observation: a magnetic stimulation study. *J. Neurophysiol.* 73, 2608–2611.
- Fajen, B. R., Riley, M. A., and Turvey, M. T. (2009). Information, affordances, and the control of action in sport. *Int. J. Sport Psychol.* 40, 79.
- Flinn, M. V., Geary, D. C., and Ward, C. V. (2005). Ecological dominance, social competition, and coalitionary arms races: why humans evolved extraordinary intelligence. *Evol. Hum. Behav.* 26, 10–46. doi: 10.1016/j.evolhumbehav.2004.08.005
- Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., and Rizzolatti, G. (2005). Parietal lobe: from action organization to intention understanding. *Science* 308, 662–667. doi: 10.1126/science.1106138
- Frischen, A., Bayliss, A. P., and Tipper, S. P. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychol. Bull.* 133, 694. doi: 10.1037/0033-2909.133.4.694
- Galati, A., Michael, C., Mello, C., Greenauer, N. M., and Avraamides, M. N. (2013). The conversational partner's perspective affects spatial memory and descriptions. *J. Mem. Lang.* 68, 140–159. doi: 10.1016/j.jml.2012.10.001
- Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain* 119, 593–609. doi: 10.1093/brain/119.2.593
- Gallese, V., and Sinigaglia, C. (2011). What is so special about embodied simulation? *Trends Cogn. Sci.* 15, 512–519. doi: 10.1016/j.tics.2011.09.003
- Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston, MA: Houghton Mifflin.
- Grabherr, L., Cuffel, C., Guyot, J.-P., and Mast, F. W. (2011). Mental transformation abilities in patients with unilateral and bilateral vestibular loss. *Exp. Brain Res.* 209, 205–214. doi: 10.1007/s00221-011-2535-0
- Grabherr, L., Karmali, F., Bach, S., Indermaur, K., Metzler, S., and Mast, F. W. (2007). Mental own-body and body-part transformations in microgravity. *J. Vestib. Res.* 17, 279–287.
- Grafton, S. T., and Hamilton, A. F. D. C. (2007). Evidence for a distributed hierarchy of action representation in the brain. *Hum. Mov. Sci.* 26, 590. doi: 10.1016/j.humov.2007.05.009
- Hamilton, A. F. D., Brindley, R., and Frith, U. (2009). Visual perspective taking impairment in children with autistic spectrum disorder. *Cognition* 113, 37–44. doi: 10.1016/j.cognition.2009.07.007
- Higuchi, T., Takada, H., Matsuura, Y., and Imanaka, K. (2004). Visual estimation of spatial requirements for locomotion in novice wheelchair users. *J. Exp. Psychol. Appl.* 10, 55–68. doi: 10.1037/1076-898X.10.1.55
- Hommel, B., Müsseler, J., Aschersleben, G., and Prinz, W. (2001). The theory of event coding (TEC): a framework for perception and action planning. *Behav. Brain Sci.* 24, 849–878. doi: 10.1017/S0140525X01000103
- Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., and Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biol.* 3:e79. doi: 10.1371/journal.pbio.0030079
- Isenhour, R. W., Richardson, M. J., Carello, C., Baron, R. M., and Marsh, K. L. (2010). Affording cooperation: embodied constraints, dynamics, and action-scaled invariance in joint lifting. *Psychon. Bull. Rev.* 17, 342–347. doi: 10.3758/PBR.17.3.342
- Ishak, S., Adolph, K. E., and Lin, G. C. (2008). Perceiving affordances for fitting through apertures. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 1501–1514. doi: 10.1037/a0011393
- Johnson-Frey, S. H., Maloof, F. R., Newman-Norlund, R., Farrer, C., Inati, S., and Grafton, S. T. (2003). Actions or hand-object interactions? Human inferior frontal cortex and action observation. *Neuron* 39, 1053–1058. doi: 10.1016/S0896-6273(03)00524-5
- Kelly, J. W., Beall, A. C., and Loomis, J. M. (2004). Perception of shared visual space: establishing common ground in real and virtual environments. *Presence Teleoper. Virtual Environ.* 13, 442–450. doi: 10.1162/1054746041944786
- Kessler, K., and Rutherford, H. (2010). The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions. *Front. Psychol.* 1:213. doi: 10.3389/fpsyg.2010.00213
- Kessler, K., and Thomson, L. A. (2010). The embodied nature of spatial perspective taking: embodied transformation versus sensorimotor interference. *Cognition* 114, 72–88. doi: 10.1016/j.cognition.2009.08.015
- Kessler, K., and Wang, H. F. (2012). Spatial perspective taking is an embodied process, but not for everyone in the same way: differences predicted by sex and social skills score. *Spat. Cogn. Comput.* 12, 133–158. doi: 10.1080/13875868.2011.634533
- Kugler, P. N., and Turvey, M. T. (1987). *Information, Natural Laws, and Self-Assembly of Rhythmic Movement*. Hillsdale, MI: L. Erlbaum Associates.
- Lenggenhager, B., Lopez, C., and Blanke, O. (2007). Influence of galvanic vestibular stimulation on egocentric and object-based mental transformations. *Exp. Brain Res.* 184, 211–221. doi: 10.1007/s00221-007-1095-9
- Linkenauger, S. A., Witt, J. K., Stefanucci, J. K., Bakdash, J. Z., and Proffitt, D. R. (2009). The effects of handedness and reachability on perceived distance. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 1649–1660. doi: 10.1037/a0016875
- Loomis, J. M., Klatzky, R. L., Golledge, R. G., and Philbeck, J. W. (1999). “Human navigation by path integration,” in *Wayfinding Behavior: Cognitive Mapping and other Spatial Processes*, ed R. G. Golledge (Baltimore, MD: Johns Hopkins University Press), 125–151.
- Mark, L. S. (1987). Eyeheight-scaled information about affordances: a study of sitting and stair climbing. *J. Exp. Psychol. Hum. Percept. Perform.* 13, 361–370. doi: 10.1037/0096-1523.13.3.361
- Mark, L. S. (2007). Perceiving the actions of other people. *Ecol. Psychol.* 19, 107–136. doi: 10.1080/10407410701331967
- Mark, L. S., Balliet, J. A., Craver, K. D., Douglas, S. D., and Fox, T. (1990). What an actor must do in order to perceive the affordance for sitting. *Ecol. Psychol.* 2, 325–366. doi: 10.1207/s1532696eco0204_2
- Mark, L. S., and Voge, D. (1988). A biodynamic basis for perceiving action categories: a study of sitting and stair climbing. *J. Mot.*

- Behav.* 19, 367–384. doi: 10.1080/00222895.1987.10735418
- Marsh, K. L., Richardson, A. E., Baron, R. M., and Schmidt, D. (2006). Contrasting approaches to perceiving and acting with others. *Ecol. Psychol.* 18, 1–38. doi: 10.1207/s15326969eco1801_1
- Mast, F., Bamert, L., and Newby, N. (2007). “Mind over matter? Imagined body movements and their neural correlates,” in *Spatial Processing in Navigation, Imagery, and Perception*, eds F. Mast and L. Jancke (New York, NY: Springer), 353–368. doi: 10.1007/978-0-387-71978-8_18
- May, M. (1996). Cognitive and embodied modes of spatial imagery. *Psychol. Beitr.* 38, 418–434.
- May, M. (2004). Imaginal perspective switches in remembered environments: transformation versus interference accounts. *Cogn. Psychol.* 48, 163–206. doi: 10.1016/S0010-0285(03)00127-0
- May, M. (2007). Imaginal repositioning in everyday environments: effects of testing method and setting. *Psychol. Res.* 71, 277–287. doi: 10.1007/s00426-006-0083-1
- May, M., and Wendt, M. (2012). Separating mental transformations and spatial compatibility effects in the own body transformation task. *Cogn. Process.* 13, S257–S260. doi: 10.1007/s10339-012-0455-y
- Mazzarella, E., Hamilton, A., Trojano, L., Mastromauro, B., and Conson, M. (2012). Observation of another's action but not eye gaze triggers allocentric visual perspective. *Q. J. Exp. Psychol.* 65, 2447–2460. doi: 10.1080/17470218.2012.697905
- Michaels, C. F., and Carello, C. (1981). *Direct Perception*. Englewood Cliffs, NJ: Prentice-Hall.
- Michelon, P., and Zacks, J. M. (2006). Two kinds of visual perspective taking. *Percept. Psychophys.* 68, 327–337. doi: 10.3758/BF03193680
- Newell, K., Scully, D., Tenenbaum, F., and Hardiman, S. (1989). Body scale and the development of prehension. *Dev. Psychobiol.* 22, 1–13. doi: 10.1002/dev.420220102
- Oullier, O., de Guzman, C. G., Jantzen, K. J., Lagarde, J., and Kelso, J. A. S. (2008). Social coordination dynamics: measuring human bonding. *Soc. Neurosci.* 3, 178–192. doi: 10.1080/17470910701563392
- Piaget, J., and Inhelder, B. (1967). *The Child's Conception of Space*. (Original work published 1948). New York, NY: Norton.
- Plumert, J. M., Kearney, J. K., and Cremer, J. F. (2004). Children's perception of gap affordances: bicycling across traffic-filled intersections in an immersive virtual environment. *Child Dev.* 75, 1243–1253. doi: 10.1111/j.1467-8624.2004.00736.x
- Presson, C. C., and Montello, D. R. (1994). Updating after rotational and translational body movements: coordinate structure of perspective space. *Perception* 23, 1447–1455. doi: 10.1068/p231447
- Proffitt, D. R. (2006). Embodied perception and the economy of action. *Perspect. Psychol. Sci.* 1, 110–122. doi: 10.1111/j.1745-6916.2006.00008.x
- Ramenzoni, V. C. (2008). *Effects of Joint Task Performance on Interpersonal Postural Coordination*. Cincinnati, OH: Unpublished doctoral dissertation; University of Cincinnati.
- Ramenzoni, V. C., Davis, T. J., Riley, M. A., and Shockley, K. (2010). Perceiving action boundaries: learning effects in perceiving maximum jumping-reach affordances. *Atten. Percept. Psychophys.* 72, 1110–1119. doi: 10.3758/APP.72.4.1110
- Ramenzoni, V. C., Riley, M. A., Shockley, K., and Davis, T. J. (2008a). Carrying the height of the world on your ankles: encumbering observers reduces estimates of how high an actor can jump. *Q. J. Exp. Psychol.* 61, 1487–1495. doi: 10.1080/17470210802100073
- Ramenzoni, V. C., Riley, M. A., Shockley, K., and Davis, T. J. (2008b). An information-based approach to action understanding. *Cognition* 106, 1059–1070. doi: 10.1016/j.cognition.2007.04.008
- Ramenzoni, V. C., Riley, M. A., Shockley, K., and Armstrong, R. (2008c). Tuning in to another person's action capabilities: perceiving maximal jumping-reach height from walking kinematics. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 919–928. doi: 10.1037/0096-1523.34.4.919
- Reed, C., and Farah, M. J. (1995). The psychological reality of the body schema: a test with normal participants. *J. Exp. Psychol. Hum. Percept. Perform.* 23, 334–343. doi: 10.1037/0096-1523.21.2.334
- Rieser, J. J. (1989). Access to knowledge of spatial structure at novel points of observation. *J. Exp. Psychol. Learn. Mem. Cogn.* 15, 1157–1165. doi: 10.1037/0278-7393.15.6.1157
- Riley, M. A., Richardson, M. J., Shockley, K., and Ramenzoni, V. C. (2011). Interpersonal synergies. *Front. Psychol.* 2:38. doi: 10.3389/fpsyg.2011.00038
- Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192. doi: 10.1146/annurev.neuro.27.070203.144230
- Rizzolatti, G., Fogassi, L., and Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat. Rev. Neurosci.* 2, 661–670. doi: 10.1038/35090060
- Rochat, M. J., Caruana, F., Jezzini, A., Intskirveli, I., Grammont, F., Gallese, V., et al. (2010). Responses of mirror neurons in area F5 to hand and tool grasping observation. *Exp. Brain Res.* 204(4), 605–616. doi: 10.1007/s00221-010-2329-9
- Rochat, P. (1995). Perceived reachability for self and for others by 3- to 5-year old children and adults. *J. Exp. Child Psychol.* 59, 317–333. doi: 10.1006/jecp.1995.1014
- Salatas, H., and Flavell, J. H. (1976). Perspective taking: the development of two components of knowledge. *Child Dev.* 47, 103–109. doi: 10.2307/1128288
- Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., and Scott, S. E. B. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people see. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 1255–1266. doi: 10.1037/a0018729
- Sartori, L., Becchio, C., Bara, B. G., and Castiello, U. (2009). Does the intention to communicate affect action kinematics? *Conscious. Cogn.* 18, 766–772. doi: 10.1016/j.concog.2009.06.004
- Schober, M. F. (2009). Spatial dialogue between partners with mismatched abilities. *Spat. Lang. Dialogue* 1, 23–39. doi: 10.1093/acprof:oso/978019554201.003.0003
- Sebanz, N., Bekkering, H., and Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends Cogn. Sci.* 10, 70–76. doi: 10.1016/j.tics.2005.12.009
- Sebanz, N., Knoblich, G., and Prinz, W. (2003). Representing others' actions: Just like one's own? *Cognition* 88, B11–B21. doi: 10.1016/S0010-0277(03)00043-X
- Shelton, A. L., Clements-Stephens, A. M., Lam, W. Y., Pak, D. M., and Murray, A. J. (2012). Should social savvy equal good spatial skills? The interaction of social skills with spatial perspective taking. *J. Exp. Psychol. Gen.* 141, 199–205. doi: 10.1037/a0024617
- Shelton, A. L., and McNamara, T. P. (2004). Spatial memory and perspective taking. *Mem. Cogn.* 32, 416–426. doi: 10.3758/BF03195835
- Stoffregen, T. A. (2003). Affordances as properties of the animal-environment system. *Ecol. Psychol.* 15, 115–134. doi: 10.1207/S15326969ECO1502_2
- Stoffregen, T. A., Gorday, K. M., Sheng, Y.-Y., and Flynn, S. B. (1999). Perceiving affordances for another person's actions. *J. Exp. Psychol. Hum. Percept. Perform.* 25, 120–136. doi: 10.1037/0096-1523.25.1.120
- Thill, S., Caligiore, D., Borghi, A. M., Ziemke, T., and Baldassarre, G. (2013). Theories and computational models of affordance and mirror systems: an integrative review. *Neurosci. Biobehav. Rev.* 37, 491–521. doi: 10.1016/j.neubiorev.2013.01.012
- Thirioux, B., Jorland, G., Bret, G., Tramus, M. H., and Berthoz, A. (2009). Walking on a line: a motor paradigm using rotation and reflection symmetry to study mental body transformations. *Brain Cogn.* 70, 191–200. doi: 10.1016/j.bandc.2009.02.006
- Tognoli, E., Lagarde, J., De Guzman, C. G., and Kelso, J. A. S. (2007). The phi complex as a neuro-marker of human social coordination. *Proc. Natl. Acad. Sci. U.S.A.* 104, 8190–8195. doi: 10.1073/pnas.0611453104
- Tucker, M., and Ellis, R. (1998). On the relations between seen objects and components of potential actions. *J. Exp. Psychol. Hum. Percept. Perform.* 24, 830–846. doi: 10.1037/0096-1523.24.3.830
- Turvey, M. T. (1992). Affordances and prospective control: an outline of the ontology. *Ecol. Psychol.* 4, 173–187. doi: 10.1207/s15326969eco0403_3
- Turvey, M. T., and Carello, C. (1996). “Dynamics of Bernstein's levels of synergies,” in *Dexterity and Its Development*, eds M. L. Latash, and M. T. Turvey (Mahwah, NJ: Erlbaum), 339–377.
- Tversky, B., and Hard, B. M. (2009). Embodied and disembodied cognition: spatial perspective taking. *Cognition* 110, 124–129. doi: 10.1016/j.cognition.2008.10.008
- Uithol, S., van Rooij, I., Bekkering, H., and Haselager, P. (2011). Understanding motor resonance. *Soc. Neurosci.* 6, 388–397. doi: 10.1080/17470919.2011.559129
- van Elk, M., and Blanke, O. (2013). Imagined own-body transformations during passive self-motion. *Psychol. Res.* doi: 10.1007/s00426-013-0486-8. [Epub ahead of print].
- Wagman, J. B., and Taylor, K. R. (2005). Perceiving affordances for

- aperture crossing for the person-plus-object system. *Ecol. Psychol.* 17, 105–130. doi: 10.1207/s15326969ecol702_3
- Warren, W. H. (1984). Perceiving affordances: visual guidance of stair climbing. *J. Exp. Psychol. Hum. Percept. Perform.* 10, 683–703. doi: 10.1037/0096-1523.10.5.683
- Warren, W. H., and Whang, S. (1987). Visual guidance of walking through apertures: body-scaled information for affordances. *J. Exp. Psychol. Hum. Percept. Perform.* 13, 371–383. doi: 10.1037/0096-1523.13.3.371
- Weast, J. A., Riley, M. A., and Shockley, K. (2011). The influence of athletic experience and kinematic information on skill-relevant affordance perception. *Q. J. Exp. Psychol.* 64, 689–706. doi: 10.1080/17470218.2010.523474
- Wilson, M., and Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychol. Bull.* 131, 460–473. doi: 10.1037/0033-2909.131.3.460
- Witt, J. K., Proffitt, D. R., and Epstein, W. (2005). Tool use affects perceived distance but only when you intend to use it. *J. Exp. Psychol. Hum. Percept. Perform.* 31, 880–888. doi: 10.1037/0096-1523.31.5.880
- Wraga, M. (2003). Thinking outside the body: an advantage for spatial updating during imagined versus physical self-rotation. *J. Exp. Psychol. Learn. Mem. Cogn.* 29, 993–1005. doi: 10.1037/0278-7393.29.5.993
- Wraga, M., Creem, S. H., and Proffitt, D. R. (2000). Updating displays after imagined object and viewer rotations. *J. Exp. Psychol. Learn. Mem. Cogn.* 26, 151–168. doi: 10.1037/0278-7393.26.1.151
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Received: 01 July 2013; accepted: 03 September 2013; published online: 24 September 2013.
- Citation: Creem-Regehr SH, Gagnon KT, Geuss MN and Stefanucci JK (2013) Relating spatial perspective taking to the perception of other's affordances: providing a foundation for predicting the future behavior of others. *Front. Hum. Neurosci.* 7:596. doi: 10.3389/fnhum.2013.00596
- This article was submitted to the journal *Frontiers in Human Neuroscience*. Copyright © 2013 Creem-Regehr, Gagnon, Geuss and Stefanucci. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Does that look heavy to you? Perceived weight judgment in lifting actions in younger and older adults

Corrina Maguinness^{1,2}, Annalisa Setti^{3,4*}, Eugenie Roudaia² and Rose Anne Kenny^{2,3}

¹ School of Psychology, Trinity College Dublin, Dublin, Ireland

² Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland

³ The Irish Longitudinal Study on Ageing, Trinity College Dublin, Dublin, Ireland

⁴ School of Applied Psychology, University College Cork, Cork, Ireland

Edited by:

Antonia Hamilton, University of Nottingham, UK

Reviewed by:

Simone Schütz-Bosbach, Max Planck Institute for Human Cognitive and Brain Sciences, Germany
Richard Ramsey, Bangor University, UK

*Correspondence:

Annalisa Setti, The Irish Longitudinal Study on Ageing, School of Psychology, Lincoln Gate, Trinity College Dublin, Dublin 2, Ireland
e-mail: asetti@tcd.ie

When interpreting other people's movements or actions, observers may not only rely on the visual cues available in the observed movement, but they may also be able to "put themselves in the other person's shoes" by engaging brain systems involved in both "mentalizing" and motor simulation. The ageing process brings changes in both perceptual and motor abilities, yet little is known about how these changes may affect the ability to accurately interpret other people's actions. Here we investigated the effect of ageing on the ability to discriminate the weight of objects based on the movements of actors lifting these objects. Stimuli consisted of videos of an actor lifting a small box weighing 0.05–0.9 kg or a large box weighing 3–18 kg. In a four-alternative forced-choice task, younger and older participants reported the perceived weight of the box in each video. Overall, older participants were less sensitive than younger participants in discriminating the perceived weight of lifted boxes, an effect that was especially pronounced in the small box condition. Weight discrimination performance was better for the large box compared to the small box in both groups, due to greater saliency of the visual cues in this condition. These results suggest that older adults may require more salient visual cues to interpret the actions of others accurately. We discuss the potential contribution of age-related changes in visual and motor function on the observed effects and suggest that older adults' decline in the sensitivity to subtle visual cues may lead to greater reliance on visual analysis of the observed scene and its semantic context.

Keywords: action perception, motion perception, visuomotor, sensorimotor, embodied cognition, motor simulation, weight judgment, aging

INTRODUCTION

Imagine being in a coffee shop and looking at a cup placed on a counter. The cup is completely opaque and you do not know whether it is full or empty. Now imagine your friend reaching for and lifting the cup to move it to another table. By observing the strength of their grip and the speed of their movement, you can immediately deduce that the cup is full, even though you still cannot see what is inside it. What's more, you can also deduce whether they knew that the cup was full or incorrectly expected it to be empty. As such, observing the actions of others involves a form of experience sharing (Brown and Brüne, 2012; Limanowski and Blankenburg, 2013), from which we can derive meaningful information about the agent's intentions and expectations as well as the characteristics of the object acted upon. This information can in turn inform our own interactions with the environment.

Our ability to understand the actions of others (action understanding or action interpretation) is likely mediated by multiple levels of analysis (Grafton and Hamilton, 2007; Thioux et al., 2008), including deducing *how* an action is performed (e.g., with the hand or with the full body), *what* the action is (e.g., lifting a cup) and *why* it is occurring (e.g., to refill the cup because it is empty) (Thioux et al., 2008). The ageing process is accompanied by perceptual and physical changes that may impact the

ability to interpret others' actions at these multiple levels of analysis. However, to date, the relationship between ageing, action perception, and judgment of object properties remains relatively unexplored. In younger adults, it has been suggested that the spatiotemporal information derived from action observation engages internal motor simulation of the observed action (Gallese et al., 2004; Knoblich and Sebanz, 2006) and that action understanding and action execution have a shared coding system (Gallese et al., 2004; Knoblich and Sebanz, 2006), as they have been shown to involve overlapping brain regions (Gallese et al., 1996; Rizzolatti et al., 1996). These shared systems may afford our understanding of actions toward objects (Buccino et al., 2004; Hamilton et al., 2006; Ramsey and Hamilton, 2012), as well as intransitive actions such as walking or dancing (Buccino et al., 2001; Calvo-Merino et al., 2005). Although such mechanisms may inform our understanding of *how* and *what* actions are performed, it has been suggested that when people infer the unobservable aspects of the action, such as *why* the action is being performed, they engage an extended network beyond the sensorimotor system to support such "mentalizing" or "theory of mind" processing (Spunt et al., 2011). Other studies have suggested a role for the motor system in conjunction with other brain networks typically involved in theory of mind processing for action interpretation (De Lange et al.,

2008; Ramsey and Hamilton, 2012; see also Keyser and Gazzola, 2007).

Motor engagement in action observation is largely modulated by the motor repertoire of the observer (Calvo-Merino et al., 2005). Evidence from healthy and patient populations suggests that spatial awareness of our own and others' body positions (Marzoli et al., 2011, 2013) and sensations arising from our body contribute to interpreting the actions of others (Hamilton et al., 2004; Bosbach et al., 2005; Ní Choisdealbha et al., 2011). For example, when Hamilton et al. (2004) asked participants to judge the weight of a box lifted by an agent while concurrently lifting a box themselves, they noted that the weight of the physically lifted box directly affected perceptual weight judgments. Participants judged the box being lifted by the agent to be heavier when they were physically lifting a light box, and vice versa. In a follow-up study, Hamilton et al. (2006) showed that the magnitude of the bias induced by the motor system on perceptual weight judgments was associated with activation of a specific cluster of visual and motor regions in the brain, leading the authors to suggest that the perceptual and motor systems are not distinct, but interact and influence each other at various levels.

The ageing process is accompanied by declines in motor abilities across a range of tasks. For example, older adults demonstrate differential velocity profiles, decreased fluidity, and increased variability in simple action execution (Cooke et al., 1989; Seidler et al., 2002, for review, see Seidler et al., 2011). The ability to imitate and replicate more complex movement sequences is also negatively affected by ageing (Maryott and Sekuler, 2009; Caçola et al., 2013). Older adults also show declines in the ability to judge the position of their body in space and appear to rely on additional sensory information, largely vision, to compensate for their decline in proprioception (Seidler-Dobrin and Stelmach, 1998; Romero et al., 2003; Barrett et al., 2013). Moreover, Diersch et al. (2012) demonstrated that when online visual information is interrupted, older adults show deficits in predicting the correct time course of action sequences. This indicates that the ability to mentally represent and predict action sequences declines with ageing (see also Saimpont et al., 2009; Gabbard et al., 2011; Diersch et al., 2012). Thus, declines in motor ability with ageing, together with changes in internal forward models of action representation (Diersch et al., 2012), may lead older adults to become more reliant on visual analysis of observed action sequences for action interpretation and inference on object properties. Interestingly, Poliakoff et al. (2010) observed that patients with Parkinson's disease can still perform perceptual weight judgments, however, they may rely more on visual analysis due to declines in the motor system (Poliakoff et al., 2010; Poliakoff, 2013). Thus, while embodied simulation may in part underlie action perception, when we cannot put ourselves "in other people's shoes" through simulation, or when this is not useful to action perception, visual analysis may support action understanding (Brady et al., 2011). Yet, little is known as to how motor changes in non-pathological ageing may affect the interpretation of other people's actions and whether a similar visual strategy may be engaged with advancing age.

Although action execution and action interpretation appear to interact, it is also important to note that they may not bear a direct

correspondence. For example, Hamilton et al. (2007) demonstrated that the most reliable physical cues as to the weight of a lifted item do not correspond to the perceptual cues that individuals use when making a weight judgment. Auvray et al. (2011) observed similar discrepancies and suggest that individuals do not engage an "exact copy" of action execution when making perceptual judgments, but rather exploit the most diagnostic visual cues, such as acceleration. Indeed, motion cues such as velocity and acceleration can be used to determine the weight of lifted objects even when visual information is only provided by moving point light displays (Shim and Carlton, 1997). Moreover, the embodied nature of forward models has been questioned, as it has been suggested that motor activation may relate less to "mirroring" or directly matching the actions of others, but rather to anticipating future compatible actions (Csibra, 2007). It has also been suggested that action understanding may be achieved through visual analysis alone without the need for direct embodied simulation (for review, see Giese and Poggio, 2003). This is largely related to our direct visual experience of naturally occurring sequences. The changes that we encounter in action sequences in a natural environment are gradual and are governed by natural laws. Through our constant exposure to naturally occurring sequences, our perceptual system can learn to predict the continuation and outcomes of observed actions (Giese and Poggio, 2003; Perrett et al., 2009). Indeed the spatial and temporal constraints observed in naturally occurring sequences can have a direct effect on our ability to encode (Wallis, 1998; Wallis and Bülthoff, 2001) and, in turn, anticipate the sequence outcome (Perrett et al., 2009). Such visual analysis abilities may be compromised in older adults.

Ageing is associated with deterioration in visual motion perception. For example, older adults are less accurate than younger adults at processing information in biological motion displays (Billino et al., 2008; Pilz et al., 2010; Insch et al., 2012; Legault et al., 2012), suggesting that their ability to process motion cues relevant to action may be impaired. However, age-related declines in motion perception are not limited to biological motion, as other forms of motion perception are also vulnerable to the ageing process (Billino et al., 2008). Older adults are less sensitive to detecting and discriminating the direction of motion in random-dot patterns, a class of stimuli commonly used to address the mechanisms underpinning motion perception (Snowden and Kavanagh, 2006; Bennett et al., 2007; Roudaia et al., 2010; Hutchinson et al., 2012). Older adults are also less sensitive to changes in the speed of moving stimuli (Scialfa et al., 1991; Snowden and Kavanagh, 2006). Thus, age-related declines in visual motion perception may limit older adults' ability to perform visual analysis of observed actions and therefore potentially negatively affect action perception in older adults.

In sum, healthy ageing is accompanied by declines in the ability to perform fine motor movements and declines in visual motion perception, both of which may compromise older adults' ability to interpret other people's actions accurately, either through a reduced ability to extract relevant cues from visual observation and/or through reduced internal simulation of observed actions. In the present study, we examined whether ageing may impact on action understanding by examining the ability of younger and older adults to derive information about the weight of an object,

based on the movements of an actor lifting the object. This task is likely to engage aspects of action understanding pertaining to *how* the action is performed (e.g., lifting the box with the hand or with full body motion; the grip and speed of the movements), and *what* the action is (e.g., lifting a small or a large box). It is a naturalistic task with which both younger and older adults have direct experience in everyday life and it is known to provide a reliable measure of sensitivity to interpret the actions of others (Hamilton et al., 2007). Furthermore, the task has been shown to engage both the perceptual and motor systems of the observer (Hamilton et al., 2004, 2007; Poliakoff et al., 2010). Stimuli consisted of a series of videos showing lifting actions of a small box with light weights and a large box with heavy weights. Small box lifts displayed upper limb motion that engaged the forearm and hand and large box lifts displayed the full body motion of the actor lifting the box from the floor. An additional set of videos contained motions that showed the lifting actions of an actor who was told incorrect information about the weight they were about to lift. This deceptive information altered the actors' movement profile, resulting in exaggerated motion that may provide greater visual cues to support weight judgment. The manipulations of box weight category and the actors' movement profile allows for exploration of the relative contribution of visual cues and motor engagement in perceptual weight judgment performance in ageing. For example, although the weights lifted in the large box condition can challenge the ageing motor system via simulation, the perceptual cues pertaining to the weight lifted may be more salient in this condition than in the small box condition (Bosbach et al., 2005). We also collected self-report measures of motor ability (Potter et al., 2009) in the older adult group to assess how perceived motor ability may be related to their capacity to interpret lifting actions.

MATERIALS AND METHODS

PARTICIPANTS

Seventeen younger adults (all female) aged 21–28 years (mean age = 24.6 years; $SD = 1.9$ years) and 19 community-dwelling older adults, recruited through an active choral society (18 female) took part in this study. Participation was voluntary and individuals did not receive monetary compensation for their time. Data from two older participants were excluded from the analysis reported below: data from one male participant was removed to maintain consistency with the all-female sample in the younger group and data from one female participant were removed because the participant did not understand the task. The remaining 17 older adults were aged 68–84 years (mean age = 74 years; $SD = 4.4$ years). All younger and older participants reported to be right hand dominant and all reported normal or corrected to normal vision. All participants wore their usual corrective lenses, if needed, at the time of testing. All participants were not suffering from psychiatric or neurological illness by self-report and all provided written informed consent. Our younger and older samples were not strictly matched for years of education, however, older adults had secondary level education or higher and younger adults were college students. The experiments reported here were approved by the St. James Hospital Ethics Committee and conformed to the Declaration of Helsinki.

STIMULI AND APPARATUS

Video stimuli

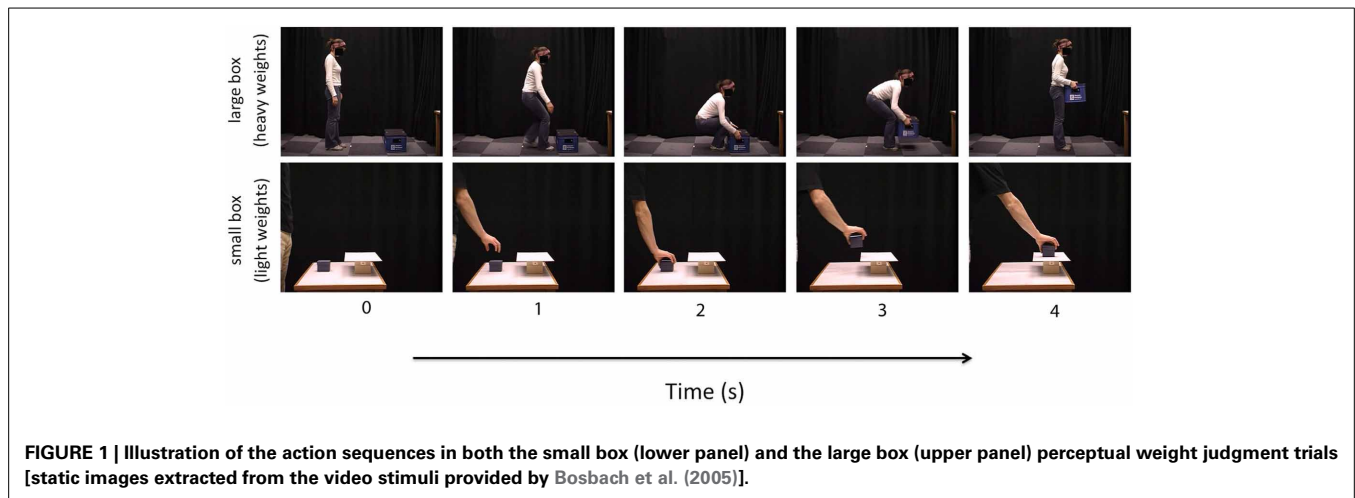
Stimuli were made available by the authors of Bosbach et al. (2005). Stimuli consisted of 8 videos of a male actor lifting a small box and 8 videos of a female actor lifting a large box. The small box videos displayed the right arm and hand of the actor lifting the small box from a table and putting it on a small shelf. The large box videos displayed the full body of the actor lifting a large box from the floor. In all videos, the external features of the box remained constant, but the weight of the box varied (see **Figure 1**). The small box weighed 50, 300, 600, or 900 g. and the large box weighed 3, 6, 12, or 18 kg. For both the small and large boxes, four non-deceptive videos showed the actor lifting the box after being told correct information about the weight of the box and four deceptive videos showed the actor lifting the box after being told incorrect information about the weight of the box (e.g., lighter than the true weight of the box). All videos showed the actor and the box from the side-view. Each video was approximately 4 s in length and was displayed at a rate of 25 frames per second. Participants viewed the videos at a distance of 60 cm and the images in the videos subtended a visual angle of approximately 14° horizontally and 11° vertically. The experiment was driven by Presentation® software and was presented on a Sony Vaio PC laptop with a 14 inch LCD screen.

Perceived Motor-Efficacy Scale for Older Adults

All older adult participants completed a subset of 19 items taken from the Perceived Motor-Efficacy Scale for Older Adults (Potter et al., 2009). This questionnaire measures the self-reported ability to engage in a number of everyday manual activities and has been shown to relate to actual physical ability. The selected items assess the perceived capability to execute tasks that engage precise manual hand movements and activities that engage full body movements, i.e., activities most relevant to the current experiment. The Appendix contains a list of all administered items. Each item was followed by a 0–10 rating points scale (0 = strongly disagree; 10 = strongly agree).

PROCEDURE

For the computer-based experiment, participants were seated at a distance of approximately 60 cm from the screen. They were instructed that they would view a number of videos of a person lifting either a small or a large box and that following each video presentation they would be asked to estimate the weight of the box the actor lifted by choosing one of four weight options shown onscreen (50, 300, 600, 900 g. for small boxes and 3, 6, 12, 18 kg. for large boxes). Participants were told that one option was always correct. Participants were offered the choice to view weight options in ounces and pounds and a number of the older adult sample opted for this option. On each trial, the video was presented for 4 s, which was then followed immediately by the response screen. Older participants responded verbally and the experimenter entered their responses by pressing the corresponding button on the keyboard. Younger adults responded by pressing the appropriate button themselves. In all cases, the button press immediately initiated the beginning of the next video.



The experiment was presented in four blocks: two blocks contained only non-deceptive videos and two blocks contained both non-deceptive and deceptive videos. The blocks containing only non-deceptive videos were always shown first, however, the order of the small and large box blocks was counterbalanced across participants. In the non-deceptive blocks, each of the four weights was repeated 3 times in random order. In the deceptive blocks, each weight was repeated once in the deceptive and once in the non-deceptive form. Each block was preceded by two practice trials to familiarize the participants with the task. Excluding practice trials, the computer task comprised of 40 trials in total, 24 trials in the non-deceptive blocks and 16 trials in the deceptive block and was approximately 10 min in duration. Following the computer based task, older adult participants completed the questionnaire comprised of the 19 selected items from the Perceived Motor Efficacy Scale for Older Adults (Potter et al., 2009). The experimenter read aloud each item and asked the participant how strongly they agreed with the statement on a scale ranging from 0 (strongly disagree) to 10 (strongly agree). The participant's response was recorded by the experimenter on the sheet. The questionnaire took approximately 5-10 min to administer. Younger adults did not complete the questionnaire, as it is specifically designed to assess perceived motor ability in older adults and as such is not informative for a younger population. All younger adults were active and were not suffering from any mobility impairments.

ANALYSIS

Data for non-deceptive videos were analyzed using the mean weight estimates, as well as signal-detection measures of sensitivity (d') and response bias (c) (Macmillian and Creelman, 2005). Mean weight estimates for each non-deceptive video were calculated by averaging the weights reported in three trials in the non-deceptive block and one trial in the deceptive block. Linear regression was used to obtain the slope and intercept of the best-fit line for each individual's estimated weights as a function of the physical weight of the box, for the small and large boxes separately. In this analysis, accurate perception of the weights would yield a slope of 1 and an intercept of 0, while a slope of

0 would indicate no relationship between perceived and actual weight.

d' scores for discriminating between each pair of adjacent weights were calculated for each participant according to the standard procedure for one-dimensional classification experiments (Macmillian and Creelman, 2005). Cumulative d' scores were then obtained by summing the d' scores for discriminating weights (W) W1 from W2, W2 from W3, and W3 from W4, yielding an overall measure of sensitivity for the small and the large box conditions. The loglinear adjustment method was used to adjust for extreme values of hits and false alarms (Stanislaw and Todorov, 1999). Similar methods were used to obtain the cumulative response bias (c) scores for each participant in the small and large box conditions.

Due to the limited number of deceptive trials, it was impossible to calculate d' and c measures for this condition, therefore, data were analyzed by obtaining the slope and intercept of the best-fit line to the weight estimates for the small and large box condition separately.

Whereas the mean weight estimates, and the fitted regression lines, are contaminated with participants' response bias, the d' measure represents an unbiased estimate of the participant's sensitivity for discriminating the weights (Macmillian and Creelman, 2005). The measure of response bias (c) was used to determine whether participants showed a preference to use either the higher or the lower end of the weight scale.

Slope and intercept values of the linear regression fits, and d' scores were analyzed using separate 2×2 mixed-design analyses of variance (ANOVA) with Age (older and younger) as the between-subjects factor and Box Type (small or large) as the within-subjects factor. c scores across Age and Box Type were tested against zero using one sample t -tests.

RESULTS

NON-DECEPTIVE TRIALS

Figure 2 shows the group average mean weight estimates of younger and older participants for non-deceptive videos in the small and large box conditions, as well as individual subjects' regression line fits. The 2 (Age) \times 2 (Box Type) ANOVA on

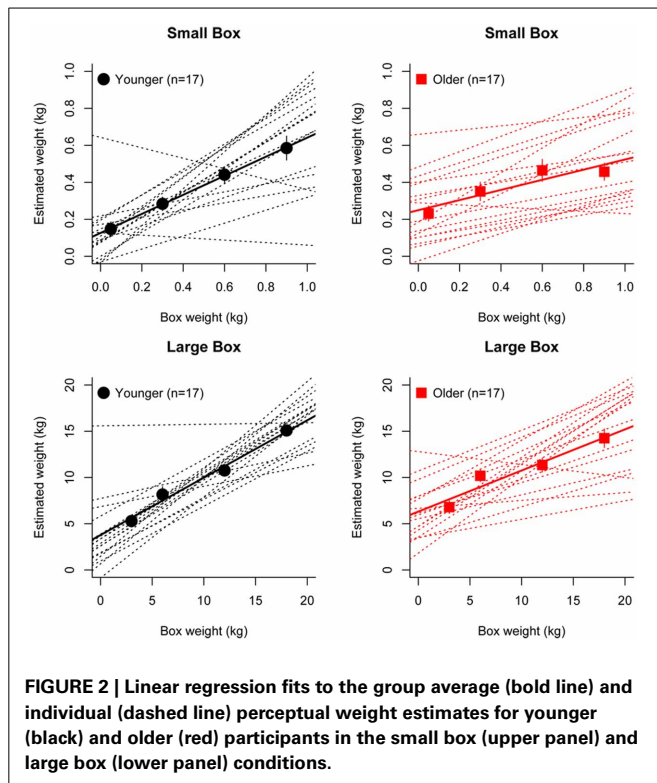


FIGURE 2 | Linear regression fits to the group average (bold line) and individual (dashed line) perceptual weight estimates for younger (black) and older (red) participants in the small box (upper panel) and large box (lower panel) conditions.

slope values revealed a significant main effect of Age [$F_{(1, 32)} = 8.56, p = 0.006$], as slopes were shallower in the older group (mean = 0.36) compared to the younger group (mean = 0.57). There was also a significant main effect of Box Type [$F_{(1, 32)} = 6.43, p = 0.02$], with shallower slopes in the small box (mean = 0.39) compared to the large box (mean = 0.53) conditions (see Figure 3). There was no significant Age \times Box Type interaction [$F_{(1, 32)} < 1$]. The 2 (Age) \times 2 (Box Type) ANOVA on intercept values revealed significant main effects of Age [$F_{(1, 32)} = 5.32, p = 0.03$], with higher intercepts in the older group compared to the younger group. The main effect of Box Type was also significant [$F_{(1, 32)} = 71.33, p < 0.001$], as intercepts in the large box were higher than in the small box. The Age \times Box Type interaction was also significant [$F_{(1, 32)} = 4.4, p = 0.04$]. Tests of simple main effects revealed a significant effect of Age for the small box [$F_{(1, 32)} = 4.47, p = 0.04$; mean younger = 0.13, mean older = 0.25] and the large box [$F_{(1, 32)} = 4.86, p = 0.03$; mean younger = 3.78, mean older = 6.31] conditions (see Figure 3). Thus, older participants showed overall shallower slopes and higher intercepts for both small and large box conditions.

Sensitivity d' analysis

Figure 4 (left) shows the mean sensitivity (d') scores for younger and older participants in the small and large box conditions. Higher d' scores represent better discrimination ability. As can be seen in the figure, older participants showed overall poorer sensitivity for discriminating weights than younger participants, especially in the small box condition. A 2 (Age) \times 2 (Box Type) ANOVA on d' scores revealed a significant main effect

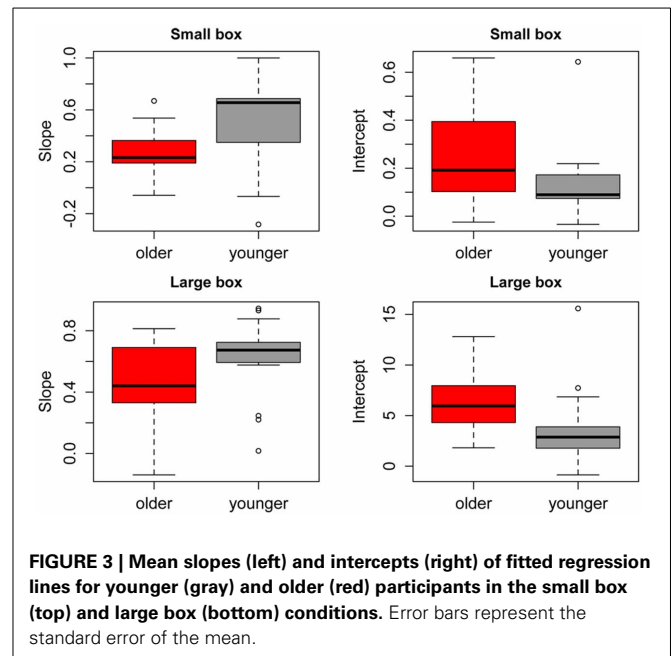


FIGURE 3 | Mean slopes (left) and intercepts (right) of fitted regression lines for younger (gray) and older (red) participants in the small box (top) and large box (bottom) conditions. Error bars represent the standard error of the mean.

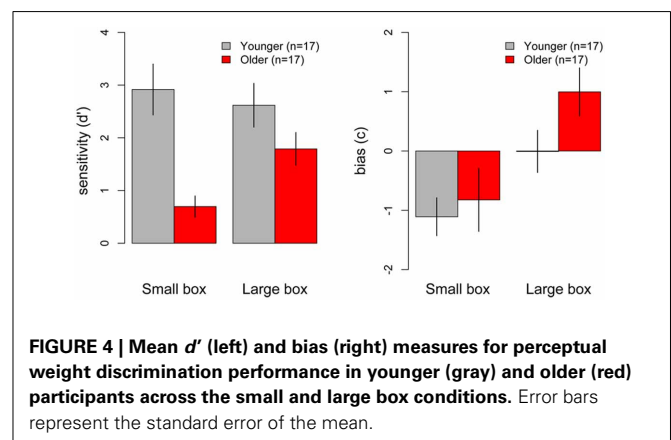


FIGURE 4 | Mean d' (left) and bias (right) measures for perceptual weight discrimination performance in younger (gray) and older (red) participants across the small and large box conditions. Error bars represent the standard error of the mean.

of Age [$F_{(1, 32)} = 18.61, p < 0.001$], with younger participants showing overall higher d' scores than older participants. The main effect of Box Type was also significant [$F_{(1, 32)} = 5.61, p < 0.001$], with overall higher d' scores in the large box compared to the small box condition. The Age \times Box Type interaction was also significant [$F_{(1, 32)} = 4.55, p = 0.04$], indicating that the effect of Age depended on the type of box. To decompose the interaction, simple main effects of Age were analyzed for the small and large box separately. Analyses revealed that older participants showed significantly lower d' scores in the small box condition [$F_{(1, 32)} = 18, p < 0.001$; younger mean = 2.92, older mean = 0.7], but there was no significant difference between d' scores in the two groups in the large box condition [$F_{(1, 32)} = 2.5, p = 0.12$] (see Figure 4). Thus, older participants showed poorer sensitivity than younger participants for discriminating weights in the small box condition, but showed similar performance to younger participants in the large box condition.

Bias analysis

Figure 4 (right) shows the mean response bias (c) scores for younger and older participants in the small and large box conditions. Positive c scores indicate participants' bias for using the upper end of the weight scale (higher weight estimations), negative c scores indicate participants' bias to respond at the lower end of the scale (lower weight estimations), and c scores near zero indicate no response bias for either end of the scale. To test for the presence of response bias, c scores were compared against zero across the small and the large box condition in the younger and older adult groups. In the small box condition, younger participants showed a significant negative bias, with c scores being significantly different from zero [$t_{(16)} = -3.46, p = 0.003$], however, older participants showed no significant bias, as c scores did not differ from zero [$t_{(16)} = -1.55, p = 0.14$]. In the large box condition, the pattern was reversed, such that older participants showed a significant positive bias [$t_{(16)} = 2.46, p = 0.03$], while younger participants showed no response bias, as their c scores did not differ from zero [$t_{(16)} = -0.02, p = 0.1$]. Thus, younger participants preferred to use the lower end of the weight scale in the small box condition only, while older participants preferred to use the upper end of the weight scale in the large box condition only (see **Figure 4**).

Deceptive trials

Linear regression was performed on each individual participant data set for the deceptive trials in order to calculate a slope and an intercept value for the small and the large box condition. Slope and intercept values were analyzed separately using a 2×2 mixed design analysis of variance (ANOVA), with Age (younger or older) as the between subjects factor and Box Type (small or large) as the within subjects factor. For the slope analysis, no significant main effects of Age [$F_{(1, 32)} = 2.05, p = 0.17$]; or Box Type [$F_{(1, 32)} < 1$] were observed. There was no significant interaction between Age and Box Type [$F_{(1, 32)} < 1$]. For the intercept analysis there was no significant effect of Age [$F_{(1, 32)} = 2.78, p = 0.1$]. There was a significant main effect of Box Type [$F_{(1, 32)} = 96.9, p < 0.001$], with lower intercept values for the small box condition. However, there was no evidence for a significant interaction between Age and Box Type [$F_{(1, 32)} = 2.44, p = 0.13$].

Perceived Motor-Efficacy Scale for Older Adults Scores

Table 1 shows the average scores from the Perceived Motor-Efficacy Scale broken down into five subscales validated by Potter and colleagues (Potter et al., 2009). All listed item numbers pertaining to the subscales can be viewed in the Appendix. Higher scores in each subscale indicate greater perceived motor ability, with a maximum score of 10. To examine the relationship between perceived motor-efficacy and perceptual weight judgment performance in the current task, we correlated the scores in each different subscale with slope estimates of the linear regression fits obtained in our experiment. There was a significant negative correlation between the Potter et al. (2009) Confidence Indicator (CI) and the slope of the non-deceptive large box condition ($r = -0.62, p = 0.007$). The CI is a measure of how cautious or confident someone is in their overall motor ability. This correlation suggests those who were more cautious (i.e., lower CI scores) had

Table 1 | Mean scores (standard deviations) for each Perceived Motor-Efficacy subscale administered.

N17 (Older adults)		
Potter subscales	Item numbers	Mean score (SD)
Perceived motor ability in the face of ageing	7; 16; 27; 3; 4	6.11 (2.28)
Perceived ability to perform precise movements	9; 14; 19; 32; 11	7.51 (2.58)
Perceived motor ability in demanding contexts	37; 15; 23; 24; 33	5.49 (2.52)
Perceived manual ability culturally specific	10; 38	9.26 (0.87)
Confidence indicator	12; 21	3.38 (2.25)

higher accuracy in perceptual weight judgments in this condition. There was also a significant positive correlation between the Potter et al. (2009) Perceived Manual Ability (PMA) and the slope of the deceptive small box condition ($r = 0.57, p = 0.02$). PMA reflects self-reported ability to use small tools and perform actions related to the use of the hands. Therefore, those with higher PMA scores performed more accurately in perceptual weight judgments in this condition, with their perceived judgments increasing in line with the physical weight of the object. No other correlations between slope measures and motor efficacy scores were found.

DISCUSSION AND CONCLUSION

Older age brings a number of physical and perceptual changes that can potentially impact older adults' ability to understand other people's actions and the characteristics of the objects acted upon. However, little is known about the effects of ageing on action perception. The present study aimed to fill this gap by using a previously-established paradigm involving weight judgment of objects lifted by an actor (Shim and Carlton, 1997; Bosbach et al., 2005; Hamilton et al., 2007). There are four main findings. First, older participants showed poorer weight estimation than younger participants for all non-deceptive videos, as evidenced by shallower slopes and higher intercepts of the function relating their weight estimates to the physical weight of the object. However, calculating participants' sensitivity (d') for discriminating the different weights revealed that older participants were especially impaired in the small box condition, while performance in the large box condition was equally good in both groups. Thus, light weights were more difficult to discriminate from one another for older adults than for younger adults. Second, we found that response bias differed between older and younger groups, with older participants showing a tendency to use higher weight estimations for weights in the large box condition and younger participants showing a tendency to use lighter weight estimations in the small box condition. Third, younger and older participants showed comparable weight estimation performance in the deceptive small and large box conditions, further indicating that older adults are not impaired in weight estimation when enhanced visual cues are available. Finally, there was a significant positive correlation between two aspects of self-report

motor abilities and weight judgment performance, which indicates a relationship between older adults' judgment of weights based on action observation and their own motor abilities.

One previous study of perceptual weight judgment in Parkinson's disease (PD) patients found that only PD patients showed evidence of poor performance, while younger controls and healthy age-matched controls did not show a significant difference in weight estimation performance (Poliakoff et al., 2010). Our current findings, however, suggest that healthy older adults' performance does differ from younger adults, at least for the small box condition. Older participants in the Poliakoff et al. (2010) study were, on average, younger than in the present study, which may have diminished the possibility of finding age-related differences in performance. Participants in that study were also allowed to lift two weights on either end of the scale prior to the experiment, which may have improved their performance.

VISUAL CUES IN ACTION PERCEPTION

As noted earlier, perceptual weight judgments involve visual analysis of the observed scene and changes in the velocity of movements provide strong diagnostic criteria for accurately deducing the weight of a lifted object (Shim and Carlton, 1997; Hamilton et al., 2007). Overall older adults' performance was worse than that of younger adults, with shallower slopes in weight estimation performance especially in the small box condition, when the weights were light (<1 kg) and the differences between the weights were small (~300 g.). The velocity profiles of the lifting actions in this condition were relatively similar across weights and may have been more challenging for the ageing visual system to exploit. Indeed, motion perception studies have demonstrated that ageing is associated with marked decreases in speed discrimination (Scialfa et al., 1991; Snowden and Kavanagh, 2006). Interestingly, older adults showed similar weight estimation performance to younger adults (as measured by slope estimates) for deceptive trials in the small box condition. This may be due to differences in the available visual cues in the deceptive and non-deceptive videos. In deceptive videos, when the actor is given incorrect information regarding the box weight (e.g., "you are going to lift a light weight" when the weight is heavy), this deceptive information results in online adjustment of the weight lifting behavior. The resulting motion profile increases the ratio of lift phase vs. the reach/grasp phase durations in the deceptive condition relative to the non-deceptive condition (Bosbach et al., 2005). It is possible that older adults are better able to exploit the visual cues in this condition and hence support more efficient performance. However, it also is important to consider that the number of trials in this condition was limited in the current study and these results should be interpreted with caution. Overall, results in the small box conditions suggest that older adults rely heavily on visual cues to judge weight from the actions depicted in the videos.

Consistent with previous studies, perceptual weight sensitivity was greater in the large box compared to the small box condition (Bosbach et al., 2005). For example, Bosbach et al. (2005) demonstrated that participants were more accurate in detecting whether an actor was surprised by the weight of a lifted box in

the large relative to the small box condition. They concluded that there are additional and more salient perceptual cues available when full body motion to heavy weights is employed, possibly leading to a better performance in heavy box condition in their study. These more salient visual cues may relate to the velocity information pertaining to the weight lifted in the action sequences. In this condition, the weights were heavy and differences between the weights was also substantial (changes of 3 kg. between weights). Consequently, the differences in the motion profiles of the lifting sequences may have been more salient. Interestingly, we found no differences in weight discrimination of older and younger participants in the large box condition, suggesting that the visual cues available in this condition were sufficiently salient for older adults to exploit. In addition, previous studies have suggested that older and younger adults may rely more on global form information when processing biological motion (Pilz et al., 2010). Unlike the small box condition, the large box condition contained full body motion, which also may have increased the relative importance of global form information in this condition.

In light of a decline in motor ability, it is possible that older adults may become more dependent on visual analysis of the observed action sequence. Indeed, previous findings suggest that individuals with proprioceptive (Bosbach et al., 2005; Toussaint and Meugnot, 2013) and motor disorders (Poliakoff et al., 2010) may engage a visual strategy for the purpose of action understanding. For example, individuals with short term limb immobilization may rely more on visual analysis for tasks which naturally induce internal motor simulation in a normal population (Toussaint and Meugnot, 2013). Our results similarly suggest that a more visual strategy may be adopted with advancing age. Specifically, we observed that sensitivity in detecting the weight of a lifted object increased as a function of the saliency of the visual cues.

In line with our study, previous research involving action perception in older adults has reported a decline in the ability to mentally represent or simulate actions. Older adults show a decline in the ability to accurately predict the timing of perceived actions, possibly due to a difficulty in building internal forward models, especially when visual cues are not always available (Diersch et al., 2012). Such behavioral changes in action perception are also reflected in the differential neural activity seen in the ageing brain during action observation. Functional brain imaging studies have shown that although a similar, yet less lateralized, action observation network is activated in younger and older adults (Diersch et al., 2013), older adults tend to engage additional cortical regions during action perception (Nedelko et al., 2010; Diersch et al., 2013). For example, in a task involving action prediction, Diersch et al. (2013) demonstrated that even when viewing familiar movements older adults tended to recruit additional visual regions of the brain to carry out the task, compared to younger adults. This suggests an overreliance on visual processing for action perception with increasing age. Similarly, differential neural activation patterns have been observed during motor execution (Seidler et al., 2011). Behaviorally older adults also exhibit an overreliance on visual input in movement tasks (Seidler-Dobrin and Stelmach, 1998; Romero et al., 2003; Barrett

et al., 2013). This overreliance on visual feedback for motor execution may be modulated by functional and structural changes in motor and somatosensory areas of the brain (review see Seidler et al., 2011).

MOTOR SIMULATION IN ACTION PERCEPTION

Although older adults' performance may be modulated to a greater extent than younger adults by the saliency of the visual cues, age-related changes in motor ability may also underlie task performance. Specifically, older adults' difficulty in discriminating between the weights of lifted objects in the small box condition parallels behavioral evidence of marked changes in simple motor behavior (e.g., Romero et al., 2003), possibly arising due to degradation in proprioceptive input with advancing age. Older adults also find it more difficult to detect small differences in the weight of physically lifted objects compared to younger adults (Norman et al., 2009). Weight ratio judgments (i.e., how much lighter is object A compared to object B) become significantly less accurate with ageing (Holmin and Norman, 2012) and the thresholds for accurately detecting such differences are over fifty per cent higher in older, compared to younger adults (Norman et al., 2009). If these behavioral changes are linked to impaired (or imprecise) motor simulation of the same actions, they may at least partly explain the age difference in our task.

We also observed a systematic bias in weight estimation, which may be reflective of the motor system of the observer. Specifically, older adults tended to report that all weights were toward the upper end of the weight scale in the large box condition, but not the small box condition, while younger adults showed a bias to report lighter weights in the small box condition and showed no bias in the large box condition. We can speculate that some form of motor simulation was recruited, as older participants would be expected to experience more difficulty lifting heavier weights, whereas younger participants should be more confident in their abilities with all weights. Interestingly we also observed that older adults' subjective judgment about their action-related skills was reflected in task performance. Specifically, accuracy performance (slope) in the deceptive small box condition correlated positively with older adults' perceived manual ability to use small tools and perform actions related to the use of the hands. Weight estimation in the large box condition was also related to older adults' perceived confidence in movement. Specifically, those who reported being more cautious in carrying out movements, i.e., perceived their own movements to be slower than usual and monitored them more, tended to have better performance in the large box condition than those with higher confidence indicator scores, a score that has been linked previously to physical motor performance (Potter et al., 2009). Potter et al. (2009) noted that higher confidence indicators may be associated with higher minor errors in simple motor execution with advancing age. This may relate to the fact that, while some older adults may experience evolving changes in motor ability, they have yet to reevaluate and integrate such declines into perceived abilities (Potter et al., 2009). Therefore, those who were more aware of their motor abilities showed better performance in the current perceptual weight judgment task across the small and large box conditions. However,

it must be acknowledged that such findings are based on the self reported motor abilities of the older adult participants. The inclusion of more objective measurements of neuropsychological and physical motor capacity would be of benefit to future studies.

CONCLUSION AND FUTURE DIRECTIONS

The current findings advance our understanding of how action perception is affected by the ageing process. Our results strongly suggest that we become increasingly reliant on robust visual cues to interpret the actions of others with advancing age. One possible consequence of this change is that older adults may be compromised in detecting subtle differences between motion profiles in action sequences, which may carry information about the intention of the actor. For example, a recent study showed that older adults were less sensitive to differences in the timing of interactions between two human characters (Roudaia et al., 2013). The timing of events carries important information about causality (Michotte, 1963). When the events involve human movements, the timing of movements carries important social information, such as deception. Due to such changes, it is possible that the ageing brain may use compensatory strategies for action interpretation. For example, the context in which an action is embedded may become essential for older adults to interpret the action, as it has been shown for younger adults in terms of mapping (Iacoboni et al., 2005) and/or inferring the meaning of others' actions, particularly when the observed actions are not encountered on a regular basis (Liepelt et al., 2008). Recent evidence from an object categorization study shows that the effect of context is more pronounced in older than younger adults (Rémy et al., 2013). It may be the case that a similar effect can be found in action understanding with advancing age.

Although the role of visual cues appears to be a plausible account for the present findings, similar to younger adult studies (e.g., Hamilton et al., 2004), the bias found for heavier weights and the correlation between weight judgment and self-perceived action capabilities in older participants suggests that some level of motor engagement may have affected task performance. Future studies should aim to disentangle the relative contribution of declines in physical and perceptual function on action perception with ageing. Finally, examining action understanding at multiple levels of analysis, including *why* an action is performed, may provide further insight into which facets of action perception remain intact or are negatively affected by the ageing process.

ACKNOWLEDGMENTS

We would like to thank Simone Schütz-Bosbach for providing us with the videos used in the present work. Support funding was provided by The Irish Longitudinal Study on Ageing (TILDA).

REFERENCES

- Auvray, M., Hoellinger, T., Hanneton, S., and Roby-Brami, A. (2011). Perceptual weight judgments when viewing one's own and others' movements under minimalist conditions of visual presentation. *Perception* 40, 1081–1103. doi: 10.1068/p6879
- Barrett, M. M., Doherty, E. P., Setti, A., Maguinness, C., Foran, T. G., Kenny, R. A., et al. (2013). Reduced vision selectively impairs spatial updating in fall-prone older adults. *Multisens. Res.* 26, 69–94. doi: 10.1163/22134808-00002412

- Bennett, P. J., Sekuler, R., and Sekuler, A. B. (2007). The effects of aging on motion detection and direction identification. *Vision Res.* 47, 799–809. doi: 10.1016/j.visres.2007.01.001
- Billino, J., Bremmer, F., and Gegenfurtner, K. R. (2008). Differential aging of motion processing mechanisms: evidence against general perceptual decline. *Vision Res.* 48, 1254–1261. doi: 10.1016/j.visres.2008.02.014
- Bosbach, S., Cole, J., Prinz, W., and Knoblich, G. (2005). Inferring another's expectation from action: the role of peripheral sensation. *Nat. Neurosci.* 8, 1295–1297. doi: 10.1038/nn1535
- Brady, N., Maguinness, C., and Ni Choisdealbha, Á. (2011). My hand or yours? Markedly different sensitivity to egocentric and allocentric views in the hand laterality task. *PLoS ONE* 6:e23316. doi: 10.1371/journal.pone.0023316
- Brown, E. C., and Brüne, M. (2012). The role of prediction in social neuroscience. *Front. Hum. Neurosci.* 6:147. doi: 10.3389/fnhum.2012.00147
- Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., et al. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *Eur. J. Neurosci.* 13, 400–404. doi: 10.1111/j.1460-9568.2001.01385.x
- Buccino, G., Vogt, S., Ritzl, A., Fink, G. R., Zilles, K., Freund, H. J., et al. (2004). Neural circuits underlying imitation learning of hand actions: An event-related fMRI study. *Neuron* 42, 323–334. doi: 10.1016/S0896-6273(04)00181-3
- Caçola, P., Roberson, J., and Gabbard, C. (2013). Aging in movement representations for sequential finger movements: a comparison between young, middle-aged, and older adults. *Brain Cogn.* 82, 1–5. doi: 10.1016/j.bandc.2013.02.003
- Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E., and Haggard, P. (2005). Action observation and acquired motor skills: an fMRI study with expert dancers. *Cereb. Cortex* 15, 1243–1249. doi: 10.1093/cercor/bhi007
- Cooke, J. D., Brown, S. H., and Cunningham, D. A. (1989). Kinematics of arm movements in elderly humans. *Neurobiol. Aging* 10, 159–165. doi: 10.1016/0197-4580(89)90025-0
- Csibra, G. (2007). “Action mirroring and action understanding: an alternative account,” in *Sensorimotor Foundations of Higher Cognition: Attention and Performance XXII*, eds P. Haggard, Y. Rosetti, and M. Kawato (Oxford: Oxford University Press), 435–459.
- De Lange, F. P., Spronk, M., Willems, R. M., Toni, I., and Bekkering, H. (2008). Complementary systems for understanding action intentions. *Curr. Biol.* 18, 454–457. doi: 10.1016/j.cub.2008.02.057
- Diersch, N., Cross, E. S., Stadler, W., Schütz-Bosbach, S., and Rieger, M. (2012). Representing others' actions: the role of expertise in the aging mind. *Psychol. Res.* 76, 525–541. doi: 10.1007/s00426-011-0404-x
- Diersch, N., Mueller, K., Cross, E. S., Stadler, W., Rieger, M., and Schütz-Bosbach, S. (2013). Action prediction in younger versus older adults: neural correlates of motor familiarity. *PLoS ONE* 8:e64195. doi: 10.1371/journal.pone.0064195
- Gabbard, C., Caçola, P., and Cordova, A. (2011). Is there an advanced aging effect on the ability to mentally represent action? *Arch. Gerontol. Geriatr.* 53, 206–209. doi: 10.1016/j.archger.2010.10.006
- Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain* 119, 593–609. doi: 10.1093/brain/119.2.593
- Gallese, V., Keysers, C., and Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends Cogn. Sci.* 8, 396–403. doi: 10.1016/j.tics.2004.07.002
- Giese, M. A., and Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nat. Rev. Neurosci.* 4, 179–192. doi: 10.1038/nrn1057
- Grafton, S. T., and Hamilton, A. F. (2007). Evidence for a distributed hierarchy of action representation in the brain. *Hum. Mov. Sci.* 26, 590–616. doi: 10.1016/j.humov.2007.05.009
- Hamilton, A., Wolpert, D., and Frith, U. (2004). Your own action influences how you perceive another person's action. *Curr. Biol.* 14, 493–498. doi: 10.1016/j.cub.2004.03.007
- Hamilton, A. F., Joyce, D. W., Flanagan, J. R., Frith, C. D., and Wolpert, D. M. (2007). Kinematic cues in perceptual weight judgement and their origins in box lifting. *Psychol. Res.* 71, 13–21. doi: 10.1007/s00426-005-0032-4
- Hamilton, A. F., Wolpert, D. M., Frith, U., and Grafton, S. T. (2006). Where does your own action influence your perception of another person's action in the brain? *Neuroimage* 29, 524–535. doi: 10.1016/j.neuroimage.2005.07.037
- Holmin, J. S., and Norman, J. F. (2012). Aging and weight-ratio perception. *PLoS ONE* 7:e47701. doi: 10.1371/journal.pone.0047701
- Hutchinson, C. V., Arena, A., Allen, H. A., and Ledgeway, T. (2012). Psychophysical correlates of global motion processing in the aging visual system: a critical review. *Neurosci. Biobehav. Rev.* 36, 1266–1272. doi: 10.1016/j.neubiorev.2012.02.009
- Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., and Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biol.* 3:e79. doi: 10.1371/journal.pbio.0030079
- Insch, P. M., Bull, R., Phillips, L. H., Allen, R., and Slessor, G. (2012). Adult aging, processing style, and the perception of biological motion. *Exp. Aging Res.* 38, 169–185. doi: 10.1080/0361073X.2012.660030
- Keysers, C., and Gazzola, V. (2007). Integrating simulation and theory of mind: from self to social cognition. *Trends Cogn. Sci.* 11, 194–196. doi: 10.1016/j.tics.2007.02.002
- Knoblich, G., and Sebanz, N. (2006). The Social Nature of Perception and Action. *Curr. Dir. Psychol. Sci.* 15, 99–104. doi: 10.1111/j.0963-7214.2006.00415.x
- Legault, I., Troje, N. F., and Faubert, J. (2012). Healthy older observers cannot use biological-motion point-light information efficiently within 4 m of themselves. *Iperception* 3, 104–111. doi: 10.1068/i0485
- Liepelt, R., Von Cramon, D. Y., and Brass, M. (2008). How do we infer others' goals from non-stereotypic actions? The outcome of context-sensitive inferential processing in right inferior parietal and posterior temporal cortex. *Neuroimage* 43, 784–792. doi: 10.1016/j.neuroimage.2008.08.007
- Limanowski, J., and Blankenburg, F. (2013). Minimal self-models and the free energy principle. *Front. Hum. Neurosci.* 7:547. doi: 10.3389/fnhum.2013.00547
- Macmillan, N. A., and Creelman, C. D. (2005). *Detection Theory: A User's Guide*. 2nd Edn. New Jersey, NJ: Lawrence Erlbaum Associates.
- Maryott, J., and Sekuler, R. (2009). Age-related changes in imitating sequences of observed movements. *Psychol. Aging* 24, 476–486. doi: 10.1037/a0015266
- Marzoli, D., Menditto, S., Lucafò, C., and Tommasi, L. (2013). Imagining others' handedness: visual and motor processes in the attribution of the dominant hand to an imagined agent. *Exp. Brain Res.* 229, 37–46. doi: 10.1007/s00221-013-3587-0
- Marzoli, D., Mitaritonna, A., Moretto, F., Carluccio, P., and Tommasi, L. (2011). The handedness of imagined bodies in action and the role of perspective taking. *Brain Cogn.* 75, 51–59. doi: 10.1016/j.bandc.2010.10.002
- Michotte, A. (1963). *The Perception of Causality*. New York, NY: Basic Books.
- Nedelko, V., Hassa, T., Hamzei, F., Weiller, C., Binkofski, F., Schoenfeld, M. A., et al. (2010). Age-independent activation in areas of the mirror neuron system during action observation and action imagery. A fMRI study. *Restor. Neurol. Neurosci.* 28, 737–747. doi: 10.3233/RNN-2010-0542
- Ni Choisdealbha, Á., Brady, N., and Maguinness, C. (2011). Differing roles for the dominant and non-dominant hands in the hand laterality task. *Exp. Brain Res.* 211, 73–85. doi: 10.1007/s00221-011-2652-9
- Norman, J. F., Norman, H. F., Swindle, J. M., Jennings, L. R., and Bartholomew, A. N. (2009). Aging and the discrimination of object weight. *Perception* 38, 1347–1354. doi: 10.1068/p6367
- Perrett, D. I., Xiao, D., Barraclough, N. E., Keysers, C., and Oram, M. W. (2009). Seeing the future: natural image sequences produce “anticipatory” neuronal activity and bias perceptual report. *Q. J. Exp. Psychol.* 62, 2081–2104. doi: 10.1080/17470210902959279
- Pilz, K. S., Bennett, P. J., and Sekuler, A. B. (2010). Effects of aging on biological motion discrimination. *Vision Res.* 50, 211–219. doi: 10.1016/j.visres.2009.11.014
- Poliakoff, E. (2013). Representation of action in Parkinson's disease: imagining, observing, and naming actions. *J. Neuropsychol.* 7, 241–254. doi: 10.1111/jnp.12005
- Poliakoff, E., Galpin, A. J., Dick, J. P. R., and Tipper, S. P. (2010). Does Parkinson's disease affect judgement about another person's action? *Exp. Brain Res.* 204, 327–331. doi: 10.1007/s00221-009-1976-1
- Potter, L. M., Grealy, M. A., and O'Connor, R. C. (2009). Healthy ageing, perceived motor-efficacy, and performance on cognitively demanding action tasks. *Br. J. Psychol.* 100, 49–70. doi: 10.1348/000712608X304478
- Ramsey, R., and Hamilton, A. F. (2012). How does your own knowledge influence the perception of another person's action in the human brain? *Soc. Cogn. Affect. Neurosci.* 7, 242–251. doi: 10.1093/scan/nsq102
- Rémy, F., Saint-Aubert, L., Bacon-Macé, N., Vayssière, N., Barbeau, E., and Fabre-Thorpe, M. (2013). Object recognition in congruent and incongruent natural scenes: a life-span study. *Vision Res.* 91, 36–44. doi: 10.1016/j.visres.2013.07.006

- Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cogn. Brain Res.* 3, 131–141. doi: 10.1016/0926-6410(95)00038-0
- Romero, D. H., Van Gemmert, A. W. A., Adler, C. H., Bekkering, H., and Stelmach, G. E. (2003). Time delays prior to movement alter the drawing kinematics of elderly adults. *Hum. Mov. Sci.* 22, 207–220. doi: 10.1016/S0167-9457(02)00160-4
- Roudaia, E., Bennett, P. J., Sekuler, A. B., and Pilz, K. S. (2010). Spatiotemporal properties of apparent motion perception and aging. *J. Vis.* 10, 1–15. doi: 10.1167/10.14.5
- Roudaia, E., Hoyet, L., O'Sullivan, C., McGovern, D., and Newell, F. N. (2013). Ageing reduces sensitivity to timing mismatches in the perception of human motion. *Perception* 39, 139. doi: 10.1068/v130621
- Saimpont, A., Pozzo, T., and Papaxanthis, C. (2009). Aging affects the mental rotation of left and right hands. *PLoS ONE* 4:e6714. doi: 10.1371/journal.pone.0006714
- Sialfa, C. T., Guzy, L. T., Leibowitz, H. W., Garvey, P. M., and Tyrrell, R. A. (1991). Age differences in estimating vehicle velocity. *Psychol. Aging* 6, 60–66. doi: 10.1037/0882-7974.6.1.60
- Seidler, R. D., Alberts, J. L., and Stelmach, G. E. (2002). Changes in multi-joint performance with age. *Motor Control* 6, 19–31.
- Seidler, R. D., Bernard, J. A., Burutolu, T. B., Fling, B. W., Gordon, M. T., Gwin, J. T., et al. (2011). Motor control and aging: links to age-related brain structural, functional, and biochemical effects. *Neurosci. Biobehav. Rev.* 34, 721–733. doi: 10.1016/j.neubiorev.2009.10.005
- Seidler-Dobrin, R. D., and Stelmach, G. E. (1998). Persistence in visual feedback control by the elderly. *Exp. Brain Res.* 119, 467–474. doi: 10.1007/s002210050362
- Shim, J., and Carlton, L. G. (1997). Perception of kinematic characteristics in the motion of lifted weight. *J. Mot. Behav.* 29, 131–146. doi: 10.1080/00222899709600828
- Snowden, R. J., and Kavanagh, E. (2006). Motion perception in the ageing visual system: minimum motion, motion coherence, and speed discrimination thresholds. *Perception* 35, 9–24. doi: 10.1068/p5399
- Spunt, R. P., Satpute, A. B., and Lieberman, M. D. (2011). Identifying the what, why, and how of an observed action: an fMRI study of mentalizing and mechanizing during action observation. *J. Cogn. Neurosci.* 23, 63–74. doi: 10.1162/jocn.2010.21446
- Stanislaw, H., and Todorov, N. (1999). Calculation of signal detection theory measures. *Behav. Res. Methods Instruments Comput.* 31, 137–149. doi: 10.3758/BF03207704
- Thioux, M., Gazzola, V., and Keysers, C. (2008). Action understanding: how, what and why. *Curr. Biol.* 18, R431–R434. doi: 10.1016/j.cub.2008.03.018
- Toussaint, L., and Meugnot, A. (2013). Short-term limb immobilization affects cognitive motor processes. *J. Exp. Psychol. Learn. Mem. Cogn.* 39, 623–632. doi: 10.1037/a0028942
- Wallis, G. (1998). Temporal order in human object recognition learning. *J. Biol. Syst.* 6, 299–313. doi: 10.1142/S0218339098000200
- Wallis, G., and Bühlhoff, H. H. (2001). Effects of temporal association on recognition memory. *Proc. Natl. Acad. Sci. U.S.A.* 98, 4800–4804. doi: 10.1073/pnas.071028598

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 12 July 2013; accepted: 31 October 2013; published online: 25 November 2013.

Citation: Maguinness C, Setti A, Roudaia E and Kenny RA (2013) Does that look heavy to you? Perceived weight judgment in lifting actions in younger and older adults. *Front. Hum. Neurosci.* 7:795. doi: 10.3389/fnhum.2013.00795

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2013 Maguinness, Setti, Roudaia and Kenny. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX

Subset of items taken from the Perceived Motor-Efficacy Scale. Underlined items are reverse scored.

3. I usually do not attempt complex movements because I find it difficult to perform them well
4. I rarely avoid certain movements in case I fall
7. I do not feel more anxious than I used to when carrying out certain movements
9. I am not very good at activities involving precise manual movements
10. I am likely to have some difficulty using a knife and fork
11. I feel confident at adjusting movements to improve their accuracy or efficiency
12. I do not have to monitor, or keep an eye on my movements, more than I used to
14. I feel I am good at activities involving hand-to-eye coordination, such as catching a ball
15. I believe I would have no problems running for a bus if I had to
16. I rarely worry about climbing up or down stairs
19. I expect to be able to shift smoothly from one movement to another
21. I feel that my movements are slower than they used to be
23. If I were to trip-up, I am confident that I could prevent myself from falling to the ground
24. I am likely to have difficulty walking to the top of a large flight of stairs
27. I expect to be able to learn new movements within a short time
32. I consider myself to be good at activities requiring the precise timing of actions
33. I am confident in my ability to walk a long distance without any difficulties
37. I am not likely to have difficulties getting about outside in the wind
38. I believe I can easily perform the actions required when using kitchen or bathroom taps



Shared action spaces: a basis function framework for social re-calibration of sensorimotor representations supporting joint action

Giovanni Pezzulo^{1*}, Pierpaolo Iodice¹, Stefano Ferraina² and Klaus Kessler³

¹ Institute of Cognitive Sciences and Technologies, National Research Council, Rome, Italy

² Department of Physiology and Pharmacology, Sapienza University, Rome, Italy

³ School of Life and Health Sciences, Aston Brain Centre, Aston University, Aston Triangle, Birmingham, UK

Edited by:

Sarah H. Creem-Regehr, University of Utah, USA

Reviewed by:

Cristina Becchio, Università Degli Studi di Torino, Italy

Margaret R. Tarampi, University of California, Santa Barbara, USA

*Correspondence:

Giovanni Pezzulo, Institute of Cognitive Sciences and Technologies, National Research Council, Via S. Martino Della Battaglia 44, 00185 Rome, Italy
e-mail: giovanni.pezzulo@istc.cnr.it

The article explores the possibilities of formalizing and explaining the mechanisms that support spatial and social perspective alignment sustained over the duration of a social interaction. The basic proposed principle is that in social contexts the mechanisms for sensorimotor transformations and multisensory integration (learn to) incorporate information relative to the other actor(s), similar to the “re-calibration” of visual receptive fields in response to repeated tool use. This process aligns or merges the co-actors’ spatial representations and creates a “Shared Action Space” (SAS) supporting key computations of social interactions and joint actions; for example, the remapping between the coordinate systems and frames of reference of the co-actors, including perspective taking, the sensorimotor transformations required for lifting jointly an object, and the predictions of the sensory effects of such joint action. The social re-calibration is proposed to be based on common basis function maps (BFMs) and could constitute an optimal solution to sensorimotor transformation and multisensory integration in joint action or more in general social interaction contexts. However, certain situations such as discrepant postural and viewpoint alignment and associated differences in perspectives between the co-actors could constrain the process quite differently. We discuss how alignment is achieved in the first place, and how it is maintained over time, providing a taxonomy of various forms and mechanisms of space alignment and overlap based, for instance, on automaticity vs. control of the transformations between the two agents. Finally, we discuss the link between low-level mechanisms for the sharing of space and high-level mechanisms for the sharing of cognitive representations.

Keywords: joint action, perspective taking, basis function, sensorimotor transformation, spatial alignment, mental alignment, social interaction

INTRODUCTION

Goodale and Milner (1992) proposed a segregation into perception-for-identification (of objects) vs. perception-for-action and empirically corroborated this claim in later work relating the former to the ventral (occipito-temporal) and the latter to the dorsal (occipito-parietal) processing stream, respectively (Milner and Goodale, 2008). While the ventral stream seems to employ relative metrics based on an environment-/object-based frame of reference (FOR), the dorsal perception-for-action stream codes “real” distances within an egocentric FOR (Aglioti et al., 1995; Ganel et al., 2008). This distinction is crucial in the present context, where we will focus primarily on perception-for-action and the properties of the dorsal stream.

The way we organize and neurally represent the space around us in the dorsal stream is functional to action performance and not only to the description of where objects are (Goodale and Milner, 1992; Rizzolatti et al., 1997). During sensorimotor learning, the actions we perform shape our perceptual representations

so that they support efficient sensorimotor transformations such as the calculation of the motor commands required to achieve a goal (e.g., reaching and grasping an object) and the prediction of the sensory consequences of actions (Wolpert et al., 1995; Pouget et al., 2002). These sensorimotor transformations are often (but not exclusively) linked to a brain network that includes the dorsal processing stream, i.e., the posterior parietal cortex (Colby and Goldberg, 1999; Ferraina et al., 2009a), the premotor cortex (Graziano et al., 1994; Rizzolatti and Luppino, 2001), yet, also the cerebellum (Wolpert et al., 1995; Kawato, 1999).

This tight relationship between visuo-spatial representations and actions implies that spatial locations must be encoded in relation to the instantaneous and multisensory internal representation of the agent’s body in order to account for the flexibility and precision of action execution, disregarding other aspects such as the particular body posture and limb locations in relation to the environment (Gross and Graziano, 1995).

This action-based view of visuo-spatial processing in the dorsal stream predicts that the neuronal mechanisms supporting spatial perception and multisensory integration should be dynamic. In this vein, Head and Holmes (1911) have proposed that the brain maintains and continuously updates a multimodal representation of the body: a *body schema*. During movement or learning of new motor skills, the body schema is updated to code where the body parts are located in space and what is their configuration. During development, the body schema is updated to code for new action possibilities due to growth or the acquisition of new motor skills. Furthermore, the body schema should incorporate action-relevant objects and thus be updated when using tools that (for example) extend the reach. For example, the visual response fields of bimodal neurons in monkey intraparietal area (modulated by both somatosensory and visual stimulation) expand as an effect of tool use to include the entire length of the tool (Maravita and Iriki, 2004). In other words, learning to use novel tools stretches the body schema or extends the internal representation of the actor's hand (Arbib et al., 2009). Patients suffering of hemispatial neglect in their near space, as a consequence of parietal cortex lesions, display symptoms in the far space when using a tool to extend their action potentials (Berti and Frassinetti, 2000). Other studies showed that tool use also influences perceptual judgments; for instance, the egocentric distance to a target object is perceived smaller when holding a tool (Witt and Proffitt, 2008). These studies suggest that the dynamic aspects of multisensory receptive fields and perceptual representations depend on the execution of goal-directed actions, consistent with the idea of a common coding of perception and action in *ideomotor* theories (Prinz, 1997).

In this article we extend the principles of the action-based approach to the case of social interactions. We propose that co-actors engaged in social interactions and particularly those having common goals (e.g., lifting together a table, playing beach volleyball as a team) are able to include other-agent's operational spaces in their own space representation.

Numerous studies have shown that co-actors' perception-action loops are not independent but can influence each other (Sebanz et al., 2006). This evidence can be interpreted using a non-representational framework that describes interacting agents as coupled dynamical systems (Kelso et al., 2013). Alternatively, it has been proposed that co-actors continuously use predictive mechanisms (e.g., forward models) to predict both one's own and another's actions, and successively integrate this information to form an action plan (Sebanz and Knoblich, 2009). The prediction of another's action is often described in terms of an *action simulation* that reuses the same internal models as those implied in one's own motor control (Blakemore and Decety, 2001; Wolpert et al., 2003; Jeannerod, 2006; Pezzulo et al., 2007, 2013; Dindo et al., 2011; Pezzulo, 2011a,b). This mechanism is plausibly a costly one, as it requires planning and controlling one's own actions while at the same time simulating the co-actor's (possibly using the same internal models for both control and simulation). Furthermore, simulating another's actions requires an intermediate computational step (i.e., transformation) when the actors are not perfectly aligned in space: an egocentric "shift" from the observer's to the observed FOR, which is often called *perceptual* (Johnson and

Demiris, 2005) or *visuo-spatial perspective taking* (e.g., Zacks and Michelon, 2005).

While not denying the importance of the aforementioned mechanisms based on dynamic coupling and action simulation, we advance a theoretical proposal based on the idea that an agent performing a joint action could benefit from an additional mechanism, a neurally represented "Shared Action Space" (SAS), which directly incorporates information relative to the co-actor in one's own mechanisms for space representation and sensorimotor transformation.

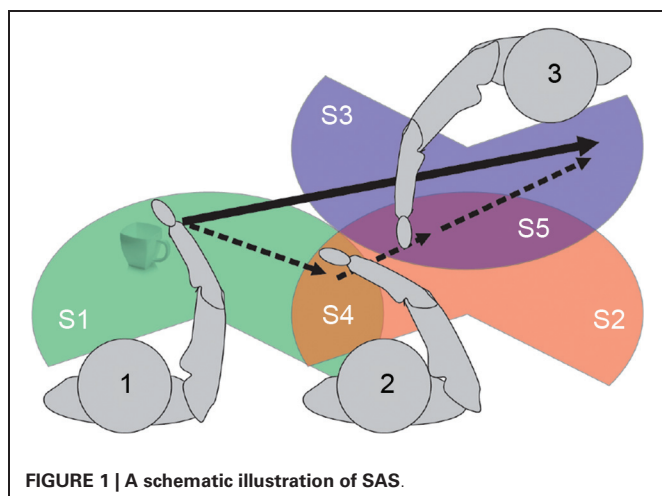
SHARED ACTION SPACES SUPPORT JOINT ACTIONS

The basic proposed principle is that in social contexts the mechanisms for sensorimotor transformations and multisensory integration (learn to) incorporate information relative to the co-actor. As an effect, the mechanisms supporting spatial representations of both agents are re-calibrated, in analogy to the re-calibration of visual receptive fields due to tool use (Maravita and Iriki, 2004). Thus, the co-actors can perceive and act using a SAS (where the word "shared" is chosen in analogy with the idea of sharing cognitive representations during joint actions (Sebanz et al., 2006); see below for a relation between these phenomena).

The social re-calibration provides a useful ground for performing numerous computations required for joint actions; for example, remapping coordinate systems and FORs (e.g., from my-eye-centered FOR to a your-eye-centered FOR or even our-position-centered FOR), sensorimotor transformations (e.g., learning the movements and amount of force necessary to lift an object jointly with another agent), and motor-to-sensory transformations such as forward modeling (e.g., predicting the sensory consequences of a joint action). The social re-calibration might thus constitute an optimal solution to sensorimotor transformation and multisensory integration in joint action or more in general social interaction contexts.

A SAS is usually *extended* compared to the individual action spaces of the co-actors and includes subspaces where actors interact or use other motor potentials. The extension of the operational space supports joint actions requiring both *simultaneous* and *complementary actions*. Consider for example the case of two persons lifting a heavy object together and simultaneously. In this case, the SAS may include *social affordances* (e.g., lifting affordances) that are not available to any of the individuals, who would not be capable of lifting the object by themselves (see also Richardson et al., 2007).

As an example of complementary actions, consider a beach volleyball team of two players. The team can reach the ball everywhere within their half of the field even if each individual player can only reach a part of it; thus the group's SAS is extended compared to the individuals' operational space. **Figure 1** provides a more detailed specification of the latter case. Three agents (1, 2, and 3) have their own operational space (S1, S2, S3, respectively) but also portions overlapping (S4 and S5) where agents could interact. The sum of S1, S2 and S3 represents the group's SAS. Thanks to this space, it is possible for agent 1 to "move" the cup to the left side of agent 3 even if he cannot physically reach such location. To perform this action he has first to pass the cup



to agent 2 (object in S4); subject 2 will then pass the object to agent 3 (object in S5) that, finally, will move the cup in the final position.

Operational spaces are of different significance. S4 and S5 represent “physically” SAS. In S4 both 1 and 2 could physically interact. The same is for S5 where the interaction is among 2 and 3. The extension and use of S4 and S5 depend on the inter-subject distance and relative orientation (both influenced by many factors; see below). However, for each of the subjects, the action space can be extended to the “virtually” SAS, even when direct interaction is not possible; for example, moving objects from S3 to S2 (or from S1 to S2) becomes an available option for all components of the group. If an agent (say, 1) neurally represents the virtually SAS, it can execute a single sensorimotor transformation to (plan to) move the cup from S1 to S3.

This example illustrates that groups such as those shown in **Figure 1** have mixed ownership of space representations. Furthermore, the operational space of group members is extended. We propose that this phenomenon is produced by the neuronal mechanisms that support sensorimotor transformations, which are re-calibrated during social interactions. The re-calibration is similar to the extension of action possibilities due to tool use, except that the skills and action repertoires of the other group members are like “tools” that extend the individual action space into a SAS affording the achievement of individualistic and joint goals.

Note however that being physically close to other persons might not be sufficient to establish a SAS; it depends on the requirements of the situation as well as various social factors how (for example) S3 is merged into the shared space. If the action goal is to simply place the mug on the “far side” of S3 then the shared space would be a merged space as shown in **Figure 1**. If the goal is to place the mug on “left side” of S3 then at least agent 2 would need to represent S3’s left/right axis taking her orientation into account. Different situations might require other kinds of information such as the position, the line of sight, the goals or even the preferences and motor skills of the co-actor. Furthermore, since co-actors are not simple tools with only a passive role, social factors come into play such as the familiarity

and trust of the co-actors in one another, as well the nature of the social interaction (say, cooperative vs. competitive) and the type of social context itself (e.g., informal vs. formal). Overall, then, various task and social requirements affect the way SAS are generated; see Section Prerequisites for Forming Shared Action Spaces and a Proposed Taxonomy.

The rest of the article is organized as follows. Section “Neuro-Computational Mechanisms Supporting Shared Action Spaces” describes the concept of SAS and proposes a neuro-computational mechanism for its implementation. Section “Prerequisites for Forming Shared Action Spaces and a Proposed Taxonomy” discusses the necessary preconditions for forming SAS and advances the idea that different mechanisms, based on automatic motor resonance or on deliberate embodied simulation, could be required depending on spatial relations and angular disparity alignment between the agents. Section “Socio-Cognitive Aspects of the Shared Action Space” discusses the relations between the idea of Shared Action Space and the sharing of cognitive representations and intentions.

NEURO-COMPUTATIONAL MECHANISMS SUPPORTING SHARED ACTION SPACES

The brain of living organisms receives information about the external world (e.g., the position of an object) from different sensory modalities (e.g., visual and auditory) and encodes them using different FORs, for example, eye-centered (i.e., distance between object and eye) for visual information and head-centered (i.e., distance between object and eye) for auditory information (Buneo and Andersen, 2006). Furthermore, information can be encoded in different coordinate systems; for example, the visual modality could encode the distance between object and eye in Cartesian (or polar) coordinates, centered at the eye or at other body’s parts (Lacquaniti et al., 1995). This multimodal information is spread in different brain areas; for example, it has been proposed that the parietal regions could use both eye-centered and hand-centered coordinates (Buneo et al., 2002; Ferraina et al., 2009b) and the premotor cortex could use body-centered representations (Caminiti et al., 1991; Graziano et al., 1994) or intermediate relative-position codes (Pesaran et al., 2006).

This information of the external world can be used to solve different problems in sensorimotor control. A first problem is *multisensory integration*, which consists in integrating information from different modalities to obtain a robust estimate of the position of the object, which in turn could require *coordinate transformation* and the remapping (or combination) of different coordinate frames. Still another problem is *sensorimotor transformation*, such as for example generating motor commands to reach and grasp the object (which in computational motor control is usually linked to *internal inverse models*). Solving this problem often requires coordinate transformations, too, such as when an eye-centered FOR used to visually locate the object has to be transformed in a body-centric or an object-centered FOR (representing the distance between the target object and the hand position and, finally, the effector shape) that could be more appropriate for reaching and grasping it (Jeannerod and Biguer, 1989). The opposite transformation (motor-to-sensory) is often

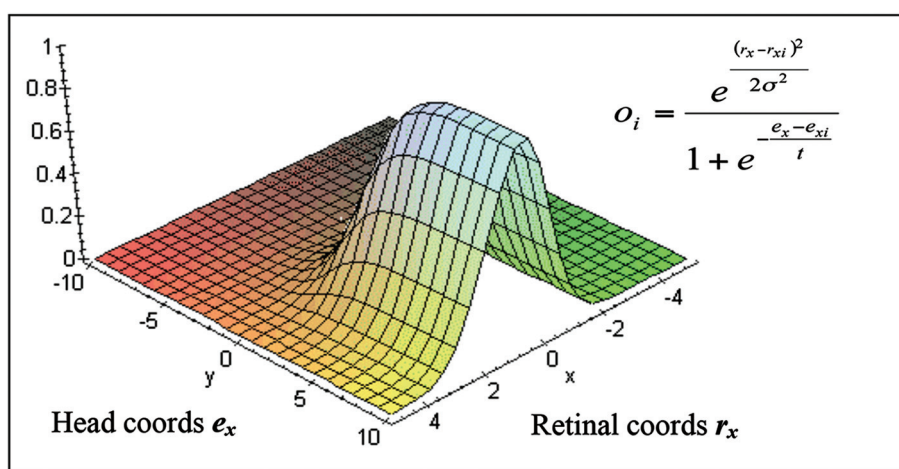
required for the sensory prediction of action consequences, which in computational motor control is often linked to *internal forward models* (Wolpert et al., 1995).

A recent computational theory of how the brain implements multisensory integration and sensorimotor transformations is the “basis functions” framework of Pouget and Snyder (2000) and Pouget et al. (2002). We adopt the “basis functions” framework to formulate our theory of SAS (but note that our theory can also be implemented differently and does not strictly depend on the basis function framework). In the basis function framework, all the streams of information are bi-directionally linked to a common basis function map (BFM; see **Figure 2** Panel A).

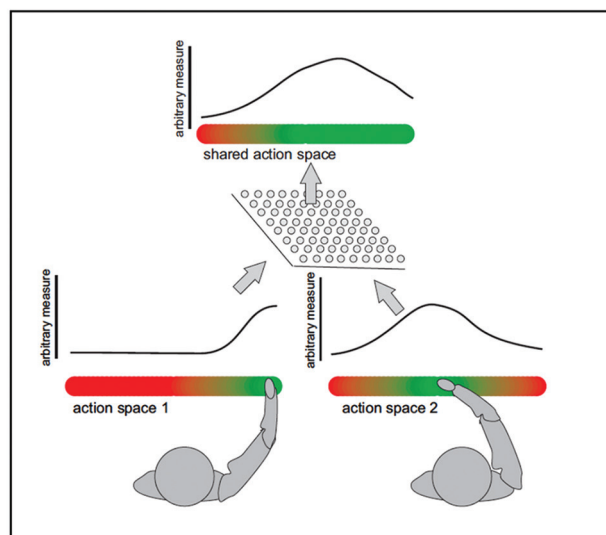
The integration of signals at the level of the BFM (equivalent to an intermediate layer of a multi-layer network) permits solving sensorimotor problems using principles of statistical inference. It permits *coordinate transformation* because the BFM essentially encodes locations in multiple frames of reference simultaneously,

creating a mixed FOR. It permits *multisensory integration* as multiple estimates (say of an object position) obtained by different sensory modalities (e.g., visual and auditory) can be combined in a mixed FOR and weighted by the relative reliability of the sensory modalities (e.g., visual information can be more reliable than auditory information).

There is indeed physiological evidence for such “combined representations” between inputs from different proprioceptive coordinate systems. Andersen and colleagues (reviewed in Andersen, 1994) reported neural populations in the macaque parietal cortex where the preference of specific neurons for a specific retinal location (i.e., the visual signal) was modulated by either head position lateral intraparietal (LIP) area or input from the labyrinth (area 7a). As a whole population such neurons have been proposed to encode combined maps as modeled by Pouget and colleagues (Pouget et al., 2002) as well as Andersen and colleagues (Andersen, 1994). What these results also suggest is that the egocentric perspective of an agent is the result of



Panel A



Panel B

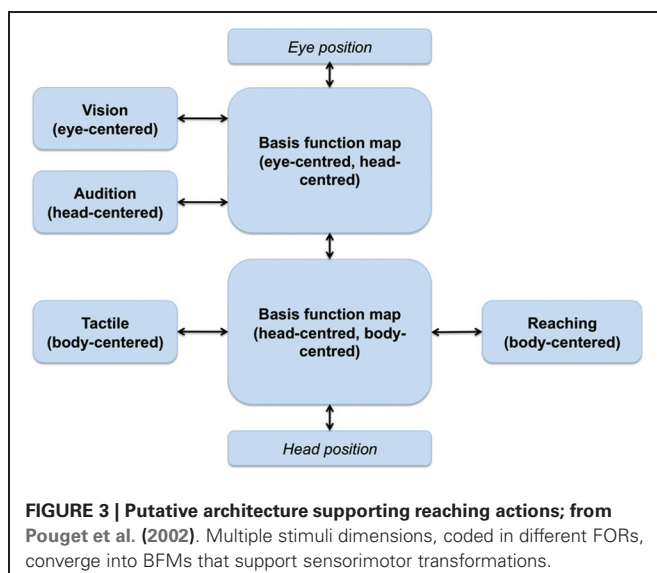
FIGURE 2 | (A) BFMs permit combining different coordinate systems into a unique representation that encodes locations in multiple frames of reference; figure adapted from Kessler (2000). **Panel (B)** an equivalent BFM combining representations of different agents.

the non-linear combination of several proprioceptive FOR that encode locations simultaneously in eye-, head-, and body-related coordinates. For action-related coding limb-relative encoding of spatial locations could be particularly important and has indeed been reported in parietal area 7b of the macaque brain (Gross and Graziano, 1995).

Furthermore a basis function model proposed by Pouget and Sejnowski (1995) and Pouget and Sejnowski (1996) was able to explain a striking modulation of hemispatial neglect reported by Karnath et al. (1993). Karnath et al. showed that a stimulus in the affected hemifield could be perceived much more easily by neglect patients when they turned their body towards the stimulus. This revealed a direct modulation of eye-centred input by proprioceptive information about body posture in neglect, which was elegantly explained by Pouget and Sejnowski's combined basis function model.

The basic architecture shown in **Figure 2**. Panel A also permits implementing efficient *sensorimotor transformations* (say reaching towards the object) not only because it supports the necessary coordinate transformations regardless of the sensory modality (e.g., from eye- or head-centric to body-centric FOR) but also because the BFM serves as an intermediate layer that permits approximating the *nonlinear* sensory-to-motor mapping as a combination of linear problems, see Pouget and Snyder (2000). As the information can flow in any direction (e.g., from sensory to motor but also from motor to sensory inputs), the same network permits also *forward modeling* and the prediction of the sensory consequences of actions.

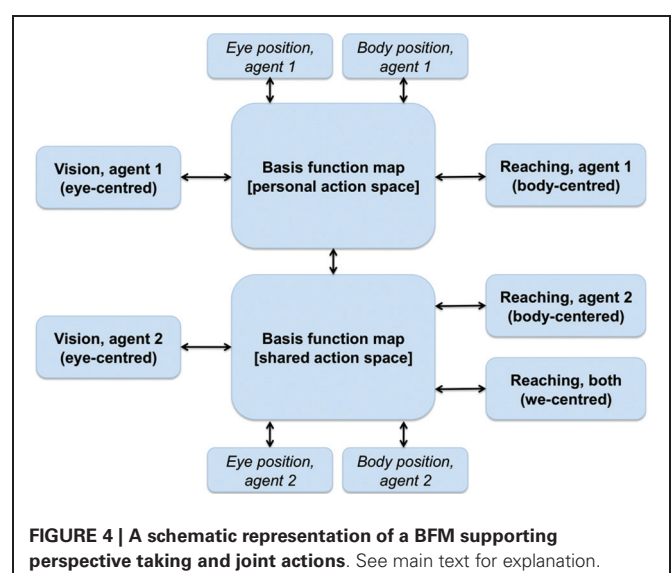
Figure 3 shows a BFN-based neural architecture supporting reaching actions that combines inputs from multiple (sensory and motor) modalities. Due to the bidirectional links, it supports transformations in all directions; for this reason, all the sources of information, either sensory or motor, can be considered both as inputs and outputs depending on the task at hand (e.g., a sensorimotor transformation from vision to action or a prediction from action to vision).



FROM INDIVIDUALISTIC TO INTERACTIVE SENSORIMOTOR TRANSFORMATIONS

We argue that a similar architecture of combined basis functions can support joint action problems and the formation of a SAS between co-actors when information relative to the co-actor (e.g., its position, its actions) is linked to the BFMs. As shown in **Figure 4**, this can be achieved by extending the basis function idea of **Figure 3**. One possibility is that a single BFM can include sensory and motor modalities of oneself and another agent (e.g., one's own and another's eye, head and/or body positions). This map would support "individualistic" sensorimotor transformations (e.g., predict only the consequences of one's own actions) when it only receives input relative to oneself. When it also receives inputs relative to another agent, the same network supports "social" sensorimotor transformations (e.g., predict the combined consequences of own and another's actions). Another possibility, suggested in **Figure 4**, is that two separate BFMs code for individualistic and social sensorimotor transformations. In either way, the BFMs would come to encode a SAS in the sense that it simultaneously encodes the sensorimotor transformations of both agents, and beyond (e.g., actions that they can only do together such as lifting together a heavy object).

It is worth noting that sensorimotor transformations and remapping are predictive processes. For example, Duhamel et al. (1992) showed that receptive fields in LIP shift in the direction of saccades before the eyes have moved, and this mechanism maintains the visual scene stable. Similarly, sensorimotor transformations in the SAS are likely to be predictive processes about a co-actor's future actions and how shared affordances may develop accordingly, which, in turn, is necessary for real-time coordination. In a similar vein, most theories of social interaction and joint action use the concepts of *action simulation* and *forward modeling* to emphasize that predictive processing is necessary for a correct unfolding of the interaction dynamics, see Pezzulo et al. (2013) for a review.



Note that all the associations shown in **Figure 4** between the individual modalities and the BFM are bidirectional. This implies that not only the input modalities influence the BFM, but also vice versa, and so in principle an input can influence backward any other input. In “individualistic” sensorimotor transformations the bidirectionality creates subtle effects (some of which are empirically observed), including the fact that receptive fields linked to a given modality (say, auditory) can “shift” and the amplitude of their response changes when the inputs in any other modality change (e.g., when eyes are moved) (Pouget et al., 2002). This suggests the intriguing possibility that in the presence of SAS the coding of information relative to the others can influence one’s own multisensory coding. This possibility remains to be investigated in the future.

A potential problem with our proposal is that while one’s own body’s sensory and motor information is readily available through sensation and proprioception, the same is not true for information concerning a co-actor. However, several studies show that the boundaries of the body are not fixed and “bodily” representations can generalize and respond for example to the touch of a rubber hand (Botvinick and Cohen, 1998). Furthermore, there are various brain areas that encode “social” information, and which could give access to (at least a part of) a co-actor’s sensory, motor, and affective information, thus providing the kind of inputs required for our model. One possible source of information is the superior temporal sulcus (STS) that is implied in biological motion perception and could encode another’s visual and postural information (Saygin, 2007). Recently, Kessler and Miellet (2013) reported the so-called “embodied body-gestalt” effect (eBG), where the instantaneous posture of the observer directly impacts on how efficiently occluded bodies of other people are integrated into a body gestalt. This seems to suggest that proprioceptive information, i.e., the own body schema, directly impacts on the perception of another’s posture and actions, which could be mediated by combined representations in form of basis function networks. In extension of the eBG, physiological evidence exists for combined representations in the perception of space in relation to another’s body in form of visuo-tactile neurons that are sensitive to visual stimuli linked to another’s body (Ishida et al., 2010); see also Thomas et al. (2006).

Furthermore, mirror neurons could give access to information relative to another’s actions and their goals (Rizzolatti and Craighero, 2004). Mirror responses are sensitive to the operational space of perceived agents (Caggiano et al., 2009) and so could therefore signal the potentialities for interaction and the utility of integrating another’s actions into one’s own sensorimotor transformations (for example, for executing complementary actions). Mirror neurons are part of a wider “action observation network (AON)” that includes parietal, premotor, and occipitotemporal regions within the (human) brain and processes various kinds of information relative to other agents (Kessler et al., 2006; Biermann-Ruben et al., 2008; Grafton, 2009; Neal and Kilner, 2010). All this information is potentially relevant as an input dimension for forming the SAS (i.e., as one of the peripheral boxes of **Figure 4**). Furthermore, an intriguing possibility is that (portions of) the AON might constitute a proper part of the SAS itself rather than providing one of its inputs. If this is true, social

resonance, mirror responses, and the body-gestalt effect could be reflections of the such combined representations (formalized here as BFMs and networks). Finally, resonance mechanisms (e.g., empathy for pain, Avenanti et al., 2005) could give access to another’s affective states that could be useful to modulate the sensorimotor interaction, see Section Problems and Open Issues of the Current Proposal.

It is worth noting that all the aforementioned processes act largely automatically. However, social cognition is supported by a range of deliberate mechanisms, too, which are often referred to as a “mentalizing” network (Frith and Frith, 2008). Although these mechanisms are typically associated with high-level information (e.g., inferring the beliefs of other agents) there are various demonstrations that they can influence social perception and ongoing action simulations, see Pezzulo et al. (2013) for a review. This suggests that an additional input can be provided by deliberate forms of perspective taking and embodied simulations that differ substantially from automatic effects. In Section Prerequisites for Forming Shared Action Spaces and a Proposed Taxonomy we elaborate on the idea that different kinds of spatial arrangements between the co-actors make some inputs, but not others, available, determining different characteristics of the SAS.

Overall, the mechanism shown in **Figure 4** can integrate various aspects of the co-actor’s sensory, motor, and goal information (at least after proper training, see later). Although this information cannot be as reliable as one’s own proprioception, it could suffice to support efficient sensorimotor interactions and joint actions.

HOW JOINT ACTION PROBLEMS ARE RESOLVED WITHIN A SHARED ACTION SPACE

The SAS exemplified in **Figure 4** provides a neuronal substrate permitting actors to co-represent the other agent(s) and to support joint actions (or more generally social interactions) efficiently. For example, it could permit perspective taking and the remapping of egocentric eye-centered coordinates between the co-actors (providing that an estimate of the co-actor’s position can be obtained). It could permit taking another’s movements into consideration when planning an action, which is useful for avoiding collisions but also for modulating one’s actions so that the combined effect with the co-actor’s actions is appropriate (say, when lifting a table together, the table remains stable and horizontal), or for calculating the combined operational space of the co-actors, as in the beach volley team example before. Below we discuss in detail how the SAS permits solving a few selected problems of joint actions and sensorimotor interactions.

Extending the operational space; multisensory aspects

As we have discussed before, experiments on tool use show that multisensory representations remap when new skills are acquired, suggesting that they code for an “operational space” that depends on action possibilities (e.g., how far I can reach) rather than absolute position of objects in space. The same multisensory remapping could occur as a consequence of the formation of a SAS, in which the action possibilities of co-actors (or more generally of agents engaged in social interactions) extend. For example,

a somatosensory remapping could occur as a consequence of the extended operational space of a team of beach volleyball players; somatic and visual responses could be elicited that are linked to parts of the space that can be reached by any of the team. **Figure 1** provides a schematic illustration of an extended operational space.

In analogy with the aforementioned evidence on tool use, it can be argued that every player sees the other players as “tools” that extend their bodies and action possibilities; for example, stretching the space that can be reached. A study conducted by Thomas et al. (2006) shows that sensory events can be elicited that are associated by the body of another person. The authors propose that such “interpersonal body representations” could be elicited automatically when seeing another person (thus, engaging in a joint action is not necessary).

The multisensory remapping could profoundly change the way we organize the space around us. A common distinction in spatial cognition is between *peripersonal* and *extrapersonal* space (Previc, 1998). Although different sub-divisions have been proposed, they are often described in terms of what actions they support (e.g., grasping space, ambient extrapersonal space as the space where visual inputs can be collected), that is in terms of operational space; see Rizzolatti and Luppino (2001). This conceptualization suggests the possibility that what is considered a peripersonal or an extrapersonal space changes as a function of social interactions; for example, the peripersonal space of a team of beach volleyball players could combine the individual peripersonal spaces with mixed ownership. In this case, the extended operational space consists of two peripersonal spaces with overlapping parts. Similarly, the extrapersonal space that normally is mapped by visual or acoustic modalities (but also olfactory; Koulakov and Rinberg, 2011) should be influenced by social interaction. A portion of the visual space hidden by an obstacle could be re-integrated in the internal representation of the extrapersonal space using information provided by co-actors.

Extending the operational space; motor aspects

Up to the moment we have discussed somatosensory remapping. However, extending the operational space also changes what affordances and action possibilities are available. Twenty years of research on mirror mechanisms have shown that monkeys and humans code for goal-directed actions performed by other agents in a flexible way (Rizzolatti and Craighero, 2004) and can consider several details including the operational space of the agents (Caggiano et al., 2009) and the possibility of complementary actions (Newman-Norlund et al., 2007; for review see Kessler and Garrod, 2013). Other studies suggest that humans can code for the action possibilities of other agents, too, and that objects can activate affordances both when they are in one's own and another's reaching space (Costantini et al., 2010, 2011a,b). This evidence can be linked to the idea of a SAS that is extended compared to the individual action space. The SAS sketched in **Figure 4** is modulated by both one's actions and another's actions, and one's affordances and another's affordances.

This information, once coded in the SAS, can be used for performing joint actions. For example, a beach volley player

can use the model shown in **Figure 4** to predict whether or not a teammate will catch the ball and so prepare in advance a complementary action.

Note that in the beach volleyball example the operational space is the combination of the individual operational spaces. There are other cases in which the presence of two or more co-actors creates truly novel possibilities for action. Consider for example an agent facing the problem of producing the necessary actions (including body and arms posture, force, etc.) to lift a heavy object together with a co-actor. The object cannot be lifted by any of the agents, but can be lifted if both combine their efforts. A problem is how an individual agent can form a motor plan or predict the consequences of a joint action. If she can only use her internal models (e.g., forward models) without taking into consideration her co-actors actions, she cannot generate the sensory prediction that the heavy object will be lifted. However, if her sensorimotor transformations are based on a SAS, her/their forward model can consider the combined effects of her and the co-actor's actions, and predict effects that cannot be produced by individual actions. In a similar way, a SAS could permit an agent to incorporate another's motor acts (e.g., the force that she will apply) into his own plans and mesh them for more accurate control and prediction.

It is important to distinguish between action goals that are congruent between the agents (e.g., imitation of martial arts movements during practice), that are complementary between the agents (e.g., during standard dance), and that are competitive (e.g., during martial arts competition). For instance, these goals may directly influence how information about another's action space is integrated into the egocentric basis-function map(s). That is, one could think of another modulation in form of a basis function (e.g., sigmoid as in **Figure 2A**) that would reflect space/action selection likelihood, thus, resulting in dynamically augmented vs. inhibited spaces and actions. These space/action landscapes could dramatically differ depending on goals that are congruent, complementary, or competitive. For example, when imitation of a movement is required the basis function would augment the same action as expected/observed in the other agent. For a complementary or a competitive joint action the identical action expected/observed in another agent would be suppressed while an appropriate complementary action (that could be defensive or aggressive in the competitive case) would be augmented. These examples illustrate that the functioning and even the coding of BFM are highly task- and goal-dependent.

Multisensory integration

The mechanism shown in **Figure 4** permits combining the action space of two (or more) individuals. In turn, this permits integrating perceptual and motor streams of two or more individuals, which might prove useful for example for state estimation. Consider the problem of estimating the position or trajectory of an object lying between two persons (say a ball in beach volleyball). An actor's eye/head coordinates of the ball are mapped onto her body/hand coordinates for action. At the same time, these are combined with the action space of the other person forming a SAS. Within the shared space, sensory and motor information of the other person can be integrated as well that might help forming

a more robust estimation of the ball trajectory or position. For example, an actor can use the co-actor's movements (e.g., if she moves towards the ball or not) as an additional source of evidence for estimating the ball (actual and future) position.

Perceptual perspective taking and the remapping between frames of reference

As mentioned in the previous sections, the social context itself as input could have a direct modulatory effect on the combined representations in the basis function network(s) triggering a transition from an individualistic to a social or SAS. This may result in a combined operational space (a BFM of higher complexity, cf. **Figure 4**) or in a full switch to another action-guiding FOR; in other words *perspective taking*. In Section Prerequisites for Forming Shared Action Spaces and a Proposed Taxonomy we will describe in detail the spatial conditions under which perspective taking becomes necessary, while it is essential at this stage to point out the importance of the social context. In specific social situations, e.g., in a formal or hierarchical context such as a job interview, it is more likely that we adopt the other's FOR (i.e., the interviewer's perspective) than when chatting to a friend. Kessler (2000) proposed that such a direct influence of social context could also be represented as a combination of basis functions, where the likelihood of adopting the other's FOR (or any other non-egocentric FOR) increases with the formality/hierarchy (see Tversky and Hard, 2009, for other dimensions) of the social context (cf. the eye/head model by Pouget and Sejnowski (1995, 1996) shown in **Figure 2A**, where "formality of social context" would be quantified on the y-axis and "FOR orientation" on the x-axis).

While social context could mediate the likelihood for adopting another's FOR, the transformation process between the egocentric and the other's FOR is a somewhat different matter. We propose that under specific circumstances, i.e., when people are spatially aligned the transformation between the egocentric FOR and the other's FOR could be computationally equivalent to the usual re-mappings of coordinate frames (say from eye- to hand-centered) necessary for the individual to plan and control reaching and grasping actions (see next sections for details). Evidence indicates that such egocentric-to-egocentric remapping can give access to sources of evidence that are unavailable to any of the two original perspectives (Becchio et al., 2013).

In contrast to the case when people's viewpoints are aligned, when their viewpoints are mis-aligned their operational spaces cannot be easily merged and an action-guiding FOR must be chosen or negotiated (see next sections for details). This could be the FOR of one of the agents but some joint actions could benefit from adopting a common allocentric (e.g., object-centered) FOR, where it could be easier to exert detailed control over the combined effects of actions (e.g., ensuring that a lifted table remains horizontal). Although it remains largely unknown what coordinate frames are used during joint action, evidence indicates that joint attention can change the FOR from an egocentric to an allocentric one (Bockler et al., 2011).

In either case the transformation of the egocentric into a mis-aligned target FOR (either the other person's or an allocentric FOR) is not easily described by means of combined

basis functions. However, recent evidence suggests that this transformation process could be a gradual transformation within the body schema map(s) of the perspective taker (Kessler and Rutherford, 2010; Kessler and Thomson, 2010; Kessler and Wang, 2012) that can be described as a shift within basis function networks. Kessler (2000) proposed a network model that used shifter circuits (Van Essen and Anderson, 1990) to shift the egocentric FOR orientation via intermediate orientations into the target orientation congruent to a simulated body rotation (Kessler and Thomson, 2010), which would be equivalent to the use of sensorimotor basis function networks in a "simulation mode". That is, the anticipated sensorimotor and visuo-spatial outcomes are generated within the (individualistic) operational space by gradual orientation shifts without actually executing the usually associated movement. The result would be a spatially updated operational space with a simulated (egocentric) viewpoint as origin that would be spatially aligned with an allocentric or the other agent's FOR.

PROBLEMS AND OPEN ISSUES OF THE CURRENT PROPOSAL

Despite its attractiveness, the basis function framework is computationally complex and prone to scalability problems; these problems could be magnified in social domains. Below we shortly discuss potential problems and open issues linked to our proposal.

An open issue is specifying how the computations linked to the SAS (e.g., the basis functions in the BFMs) are learned in the first place. In parietal cortex, mechanisms supporting sensorimotor transformations only arise after training and can be flexibly modified by new experience. In the same way, we propose that the SAS and in particular the basis functions required for the sensorimotor transformations are formed through learning. Humans and other social species often learn sensorimotor skills (e.g., lifting objects together with somebody else, playing volleyball) while engaged in social interactions and could acquire SAS as part of the sensorimotor learning. Of course the quality of the social skills and SAS depend also on the nature of the training; sensorimotor transformations can be more or less reliable when we play volleyball with our usual partners or when we interact with a stranger (the differences are also due to the success or failure of other mechanisms such as mindreading). Given that the computations of the basis function framework are hard even in individual domains, it is unclear if and how it can scale up to "social" sensorimotor skills. A scheme that is often used for scalability is making the architecture more modular. In this sense, it is possible to hypothesize that the formation of a SAS could require forming new BFMs specialized for social interactions rather than (or in addition to) reusing and extending existing ones. Testing these possibilities empirically is an interesting direction for future research.

Another open issue is what is the better FOR for performing joint actions such as lifting an object together or passing on an object. In some cases, a natural FOR can be the body position/orientation of one of the two actors (e.g., the actor who receives the object) (Tversky and Hard, 2009). This FOR permits controlling the action from the point of view of the receiving agent so that for example the "end-state comfort" (Rosenbaum et al.,

2001) of the receiving agent can be optimized; as an example, the giver agent can pass an object to the receiver agent so that she grasps it comfortably (e.g., grasps a cup from the handle). In other cases, such as for example in symmetric joint actions (e.g., lifting an object together), an allocentric (object-centered) FOR can be used. Still another intriguing possibility is that joint actions benefit from creating novel “we-centered” frames of references, for instance a FOR that is centered between my body and your body, and novel metrics such as “relative to the distance between you and me” and “the sum of my force and your force”. The peculiarity of these metrics is that they are modified by the actions of both actors (e.g., the distance between you and me changes as an effect of my actions and your actions). They could be particularly efficacious for formulating some joint control problems, such as for example monitoring the distance between two volleyball players while performing a defence (Pezzulo, 2013). The fact that social groups (or teams) are hierarchically organized could further influence the form and extension of the SAS. A related problem is that it remains unclear so far, how different forms of spatial alignments and social requirements affect the selection or merging of individualistic FORs for establishing a common action space. This issue will be addressed in the next section where we propose a taxonomy of SAS.

In the present model we are assuming that during social interactions agents perform with similar motivations. This is often untrue. One of the two volleyball players, in our example, could be more/less motivated during the match because of a larger/smaller expected personal reward. As a consequence, his influence on actions produced in the SAS will have more/less strength and the partner has-to/could adapt for optimal performance. Neural modulation for self and other reward outcome expectation/monitoring has been shown in different areas of the frontal lobe of primates (Chang et al., 2013) and the estimate of self/other motivational variables have been proposed to act as a gain modulation during common FOR generation (Chang, 2013). In this respect, a related issue to be considered is the level of each agent’s altruism, strongly influencing behavior, as revealed by all neurobiological studies exploiting game theory based approaches to decision making (Tankersley et al., 2007; Lee, 2008; Waytz et al., 2012). Because of these and other important factors influencing social interaction, the amount of shared space used by each individual and the number and contribution of actions to common goals are expected to be negotiable and more dynamic than what we are describing with our over simplification.

Finally, both the present model and most of the studies that explored action space of individuals and joint actions all dealt with agents unmoving. However, during a beach volley match every player changes his position continuously and so do the teammates. The same argument could be valid for describing synergic actions directed to objects that will change their position in space as a consequence of the cooperation. In all these cases, the SAS is dynamically updated in extension and boundaries in a non-easily predictable way. In this situation, a body-centered FOR of the action space could facilitate this continuous update of the representation of overlapped portions of the space more than an object-based or extra-personal FOR. Thus, our model

is partial for describing all possible sources and forms of action space sharing and will require further aspects to be included in the future.

PREREQUISITES FOR FORMING SHARED ACTION SPACES AND A PROPOSED TAXONOMY

Up to now we discussed basic forms of integrating individualistic action spaces and hinted that different forms or mechanisms could be employed depending on social and spatial factors. One important distinction was made in relation to different action goals. We distinguished between action goals that are *congruent* between agents (e.g., imitation of martial arts movements during practice), that are *complementary* between agents (e.g., during standard dance), and that are *competitive* (e.g., during martial arts competition). These goals directly influence how information about another’s action space is integrated into basis-function maps, resulting in dynamically augmented vs. inhibited spaces and actions. While the goals differ, all these operations assume that the two action spaces can be directly merged into a shared space. However, direct merging might not always be appropriate and in the current section we will elaborate on the different *mechanisms* for combining spaces that define different types of SAS. Note however, that all shared spaces and combinatory mechanisms can be explained within the proposed basis function framework.

We propose a taxonomy that distinguishes between “merged” vs. “aligned” shared spaces, based on different social requirements and spatial characteristics of the interaction. This distinction is based on two main dimensions that characterize a joint action situation: (i) the social sophistication of the joint goal(s) and action requirements, in contrast to (ii) the spatial orientation/viewpoint difference between the two agents. The first dimension determines how much complexity and sophistication is required for one agent to represent the other’s experience of the world and their potential actions therein. The second dimension determines what mechanisms an agent can employ for mentally sharing an action space with the other (self-other mapping) depending primarily on the spatial layout between the two agents and their FORs (i.e., orientation difference) as well as other available FORs in the environment.

ACTION REQUIREMENTS OF A SITUATION

It is important to distinguish between situations with low-level requirements for co-representation where individualistic action spaces can be combined via automatic resonance mechanisms (i.e., mirroring, e.g., Kessler and Garrod, 2013) or low level viewpoint matching, from situations with high-level requirements, where more explicit and controlled mental alignment is required (Kessler and Rutherford, 2010; Kessler and Miellet, 2013).

Low-level requirements (and Level-1 perspective taking)

As described in relation to **Figure 1**, the three agents might simply need to represent the overlap between their individualistic action spaces for placing the cup within “easy reach” of another agent. In general, situations like these would only require superimposing the egocentric and the other agents’ action spaces within a shared space, identifying areas of overlap. Another agent’s position,

viewpoint or orientation in space matters only to the extent that it shapes their region of direct influence in relation to the egocentric space and those of any other agents. In these cases the individualistic action spaces can be directly merged according to the basis-function framework proposed above.

It is important to note that such low-level requirements and the associated merging of action spaces are also proposed to apply to the simplest form of perspective taking. Typically, perspective taking is regarded as a high-level, deliberate process of social cognition, yet, two different forms or levels of complexity have been identified (Flavell et al., 1981; Michelon and Zacks, 2006) and should be considered here. Level-1 perspective taking refers to understanding *what* another person perceives or not (e.g., what is visible to them or not), while Level-2 perspective taking refers to a deeper understanding of *how* another person experiences the world. The distinction is evidenced by different developmental onset ages (Level-1 ~2 years; Level-2 ~4–5 years) and cross-species differences, where certain forms of Level-1 perspective taking seem to be shared with other species, whereas Level-2 has so far been only conclusively identified in humans (Tomasello et al., 2005; Bräuer et al., 2007; Emery and Clayton, 2009).

This highlights the differences in complexity between the two levels, bolstering our argument that in situations where Level-1 perspective taking can resolve viewpoint/orientation differences, individualistic action spaces can be directly merged into a shared space. In the visual domain Level-1 perspective taking seems to be based on a mechanism that infers the line-of-sight of another agent based on their gaze information (Michelon and Zacks, 2006). In the present context and based on Pouget's basis-function framework such a representation could be easily and directly transformed into body-related rather than head/eye-centred coordinates, allowing for judgments of "reachability" in addition to visibility. For instance, in a situation where it is only necessary to team off and grasp objects that are hidden from the other person's view, then it is only important to represent the other's line-of-sight to determine which actions will have to be performed by ourselves and which the other agent has available (Michelon and Zacks, 2006; Kessler and Rutherford, 2010). These two action spaces could be directly merged as no transformation is required beyond representing the others' action space in relation to their body orientation and gaze direction; see Seyama and Nagayama (2005) for the integration between body orientation and gaze direction perceived in others. In general, coordinating actions that refer to very simple spatial relationships between agents and potential target objects will allow for direct merging of the agents' action-spaces.

High-level requirements (and Level-2 perspective taking)

In contrast other social goals require more a sophisticated combination of action-spaces in form of alignment. This is the case for instance, when the spatial inter-relationship between agents and/or objects, such as "visibility", are not enough but specific directional information (e.g., left vs. right) in relation to a particular origin or FOR is required. Specific mental transformations of the egocentric FOR of one agent into another are necessary in order to achieve such alignment (e.g., Kessler and Rutherford, 2010; Kessler and Thomson, 2010). The higher

cognitive effort allows for more differentiated SAS where origin-specific directions can be distinguished and where the other's body laterality is represented. For instance, one could directly determine if the other person uses their right or left hand/foot for an action. The default neurocomputational mechanism for the required transformation could be a simulated rotation of orientation in multiple basis-function maps, i.e., in multiple combined sensorimotor representations that constitute the internal body schema (e.g., Andersen, 1994; Pouget and Snyder, 2000).

Furthermore, if one agent mentally adopts another agent's viewpoint for a more complex representational alignment, then this process can be congruent to Level-2 spatial perspective taking (Kessler and Rutherford, 2010). However, agents could also choose/negotiate to use neither of their FORs but a third, "allocentric" FOR instead and where both agents would have to accomplish a mental transformation into that FOR. Such a FOR could be in relation to a fronted object (e.g., the left or right side of a car), also called intrinsic allocentric or in relation to more absolute features of the environment (such as "north"), also called absolute allocentric (see Figure 5). For instance, volleyball players might not only represent a SAS relative to each other but in relation to the allocentric alignment of the playing field, thus, optimizing their SAS relative to the purpose of the game (i.e., they are typically facing the net and their adversaries). All these processes are usually strongly influenced by learning, after including in the own representation all potential sources of information useful for common goals. The transformation can be mechanistically congruent for alignment with another person or with an allocentric FOR and has been characterized as an embodied simulation of a body rotation. However, the social goals may substantially differ: alignment with an allocentric FOR would pursue the goal of imagining the self in that virtual perspective, in contrast to the goal of imagining another's visuo-spatial experience in the case of alignment with another person's FOR (see Figure 5).

Finally, disregarding which FOR is chosen in a given context, an embodied mental transformation into that FOR's orientation only becomes necessary when the difference in orientation between the egocentric and the target FOR surpasses a certain angular disparity. This is where our second taxonomic dimension regarding spatial orientation differences ties in with our considerations so far.

SPATIAL ORIENTATION/VIEWPOINT DIFFERENCES BETWEEN AGENTS (AND FORS)

The spatial/physical orientation difference between two agents can be crucial for how easy their action spaces can be merged. Merging refers to the direct integration of action spaces in the proposed basis function framework. If the two agents stand/sit next to each other, sharing a viewpoint, then their action-spaces can be easily merged disregarding the complexity of their joint goal—at all levels of complexity the mapping of their individualistic spaces into a shared space will be a direct merging operation. Nevertheless the complexity of the goal may determine what aspects of the action-space are represented at all (e.g., mere visibility vs. more sophisticated laterality). We propose to identify this case as the

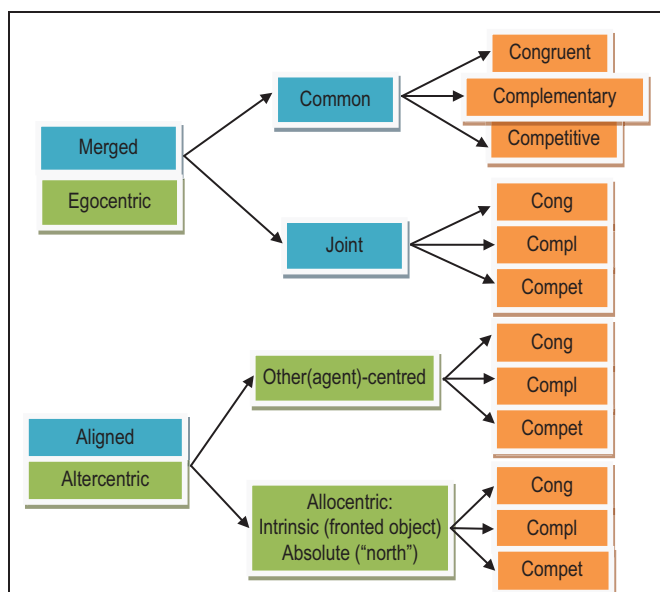


FIGURE 5 | Taxonomy for SAS divided into types (blue), origins (green), and action goals (orange). SAS can be of the “aligned” or “merged” type, with the latter subdivided into “common” and “joint” subtypes. The origin of an SAS can remain within the egocentric FOR or can be transformed into another origin (“altercentric”) imposed by another agent (“other(agent)-centred”), an intrinsically fronted object (“intrinsic allocentric”), or an absolute feature of the environment such as “north” (“absolute allocentric”). Goals can be “congruent,” “complementary,” or “competitive.” Further explanations are provided in the text.

“common” shared space subtype of “merged” action spaces (see Figure 5).

If the angular disparity between the agents increases, then the effort of combining their action-spaces increases as well. Typically there is a discrete jump in cognitive effort (e.g., response times) at around 60–90° where the overlap between the two FORs diminishes (Kessler and Thomson, 2010; Janczyk, 2013). However, this increase in effort is *only* the case for joint action goals that require sophisticated spatial alignment (Kessler and Rutherford, 2010). In the case of simple goals, individualistic action spaces can still be directly merged, disregarding orientation differences, since actions are only constrained by origin-independent spatial relationships between agents and objects such as “visibility” and “reachability” (see Section Low-Level Requirements (and Level-1 Perspective Taking)). That is, action spaces can be merged directly even for agents being positioned face to face (=180° angular disparity). Figure 1 exemplifies this in form of S5 that defines the reachability overlap between Persons 2 and 3. As described earlier (see previous sections) merging operations are likely to rely on resonance mechanisms that automatically map the observer’s body repertoire (actions, postures) and instantaneous body schema onto an observed person (Kessler and Miellet, 2013). We propose to label this type of shared space as “joint” action space. The individualistic action spaces are merged, yet in contrast to a *common* action space, the agent’s viewpoints and orientations are not physically aligned.

In the case of complex goals, the two agents would have to settle on a particular FOR and mentally align their egocentric

FOR with it to establish an “aligned” SAS. As proposed above, the default neurocomputational mechanism for the required transformation could be a simulated rotation of orientation in multiple basis-function maps, hence, the transformation can be resolved within the proposed framework. Once such a transformation into a common FOR has been accomplished there are at least two options for how this may affect the SAS. Note that we propose that a particular transformation indeed only needs to be conducted once for establishing the transition into the dominant FOR, but subsequently this FOR-dependent action-space will either replace the initially egocentric one or induce a specification of additional subspaces in a merged “joint” egocentric action space (e.g., my “left” is their “right” and vice versa) conform to the proposed basis function framework. Alternatively, however, several SAS might co-exist in concordance to the observation that several FORs can be simultaneously represented in typical (Furlanetto et al., 2013) as well as atypical neuro-cognitive processing (i.e., in heautoscopy, Brugger et al., 1997; Blanke and Mohr, 2005; Braithwaite et al., 2013). These are clearly hypothetical statements and further research is needed.

One exception to embodied transformation being the default mechanism at higher angular disparities (for complex requirements) may occur when the two agents are positioned face to face (=180° angular disparity). In this particular configuration agents may employ a different strategy by simply reversing their own egocentric space, for instance, “my left is your right” (Kessler and Wang, 2012). Again, this may feed into the specification of subspaces in a “joint” egocentric action space.

In summary we propose that socially shared space is not unitary and the following main features of the social and spatial configuration have to be taken into consideration for the way individualistic action spaces are combined into a shared space: (1) Below 60–90° of angular disparity between agents, merging into a SAS with a common egocentric FOR could occur directly, disregarding complexity of social requirements; (2) Angular disparities above 60–90° together with low-level requirements (e.g., “reachability”) may still be based on direct merging into a joint egocentric action space; yet, this egocentric FOR is not in common with the other agent; (3) Angular disparities above 60–90° together with high-level requirements (e.g., precise left/right distinctions) necessitate a transformation of the egocentric body schema into the orientation of another agent or into a common allocentric FOR in order to achieve an aligned action space with a common FOR. Strategies other than embodied transformation are possible, e.g., mental calculation (“my left is your right”) at 180°.

FINALIZING AND EXEMPLIFYING THE TAXONOMY

We initially distinguished between action goals that are congruent between agents (e.g., imitation of martial arts movements during practice), that are complementary between agents (e.g., during standard dance), and that are competitive (e.g., during martial arts competition). These goals directly influence how information about another’s action space is integrated into basis-function maps, resulting in dynamically augmented vs. inhibited spaces and actions. Based on the above considerations we propose the following taxonomy of SAS. Primarily, we suggest distinguishing between merged vs. aligned shared spaces. While merged action

spaces remain basically egocentric but are extended to incorporate the other's action space, aligned spaces require a mental transformation into another FOR. In addition, for merged spaces we propose two further sub-types, resulting in three types overall (see **Figure 5**).

Firstly, a merging process may result in a common action space, which is the likely outcome when the agents are spatially/physically aligned (i.e., identical viewpoints). The resulting SAS can be directly described within the proposed basis-function framework. Common action-spaces could easily represent simple as well as sophisticated action requirements (e.g., place the cup into another's "visible" vs. "right" space) since there is little or no discrepancy between individualistic FORs.

Secondly, joint action-spaces could be classified as spaces that have been directly merged despite strong orientation/viewpoint differences between the agents and their FORs. This is only possible with rather simplistic joint goals that only require determining "reachability", "visibility" or other simple agent-to-object and agent-to-agent relationships (e.g., Level-1 perspective taking). Joint spaces can be directly represented within the proposed basis-function framework (cf. previous sections).

Thirdly, we propose that aligned action-spaces should denote combined spaces that have not been merged in a strict sense, but where, for instance, a dominant target FOR has been negotiated, which is shared between the agents (either one of the agents' FORs or another intrinsic- or absolute-allocentric FOR). These action-spaces are likely to emerge in relation to sophisticated goals and interactions, requiring complex co-representation of another agent's experience of the world and their potential actions therein (i.e., Level-2 perspective taking). The transformation into alignment is effortful and has been characterized (Kessler, 2000) as a simulated change of orientation within multiple combined sensorimotor representations (i.e., networks of basis-function maps) identified as the body schema that constitutes the egocentric FOR (Andersen, 1994; Pouget and Sejnowski, 1995, 1996). Thus, aligned action spaces can also be described within the basis-function framework; albeit, as a transformation- rather than a merging operation. Also note that after establishing FOR alignment, the resulting sub-space characterization could be used as input to specify a joint action-space within the basis-function framework, thus, not requiring further effortful transformation. Hence, it may well be that aspects of all three types of shared spaces could dynamically contribute to a single interaction, especially if more than two agents are involved (cf. **Figure 1**). To reiterate, there is also the very interesting possibility that several shared space representations may co-exist simultaneously (e.g., joint and aligned) according to the observation, for instance in heautoscopy (for reviews see Blanke and Mohr, 2005; Furlanetto et al., 2013), that several perspectives or FORs may be represented in parallel.

Accordingly, the agents' configuration in **Figure 1** can be interpreted in different ways. Firstly, if the joint goal is to simply transfer the cup to the far side of Agent 3, then all three action spaces S1-S3 could be directly merged into a SAS. Note, however that each person would represent the other two in different ways, thus their shared space representations will differ, yet, for successful completion a few important aspects would be "meta-shared"

(meaning that two or more agents have congruent representations in this respect), such as the overlapping action spaces (S4, S5). In this particular example Agent 1 would only need to (represent and) place the cup into S4, then Agent 2 would need to (represent and) take the cup from S4 and (represent and) pass it into S5, where Agent 3 (represents and) takes the cup, finally placing it into her egocentric left subspace of S3. Note however, that Agents 1 and 2 share their orientation, so their merged action space (including the overlapping subspace S4) is a *common* space, while Agents 2 and 3 merge their action spaces into a *joint* space as they are oriented face to face. Thus, their individual representation of the joint action space will have different origins, based within each agent's egocentric orientation, however, this would not affect actions in relation to the overlapping space S5 as long as the joint action requirement remains simple (e.g., "placing the cup within reach").

Secondly, for more sophisticated inter- and joint-actions the individualistic action spaces would have to be merged or combined in more sophisticated ways that specify more detail about subspaces. Agents 1 and 2 are physically aligned in space and would therefore generate a common shared space for substantial parts of the space surrounding them: The left side of S1 is to the left of both agents while the right side of S2 is also to the right of both agents. However, the quite crucial space in-between the two agents, S4, is ambiguous with respect to left/right labeling. The agents would have to determine that this subspace has opposite labels for each agent (i.e., "right" for S4 vs. S1, but "left" for S4 vs. S2) and include these into the shared space. According to our taxonomy the resulting shared space would then be a mix between a common and a joint space.

Similar considerations apply to Agents 2 and 3. Here the orientations differ dramatically (180°), so their entire action spaces (S2 vs. S3) will have opposite left/right labels. Again a mental calculation could quickly determine this opposite labeling and include these as subspace specifications within a joint action space ("left" within S2 is "right" within S3 and vice versa). Alternatively, at greater expense, one of the agents (e.g., Agent 2) could adopt the other's perspective (Agent 3) and mentally align her action space with the other's egocentric FOR. This would result in an aligned action space with the same origin for both agents (centred on Agent 3) and with identical left/right labels for both individualistic action spaces (S2 and S3). Such abstract considerations become highly relevant in particular social contexts. For instance, if Agent 3 is a child who is not yet very skilled in grasping a cup and/or the content is hot, then Agent 2 (e.g., the mother) might anticipate more precisely where and how to place the cup within S5: Placing the cup in the child's "right" space with the handle turned towards the child's right hand, would significantly facilitate the child's task, yet, make it considerably harder for the mother in terms of specifying the child's "right" subspace (which is actually the mother's egocentric "left") within a joint or an aligned SAS.

SOCIO-COGNITIVE ASPECTS OF THE SHARED ACTION SPACE

It may require extensive practice to generate SAS that lead to successful execution of joint actions. The mother and child example may only require the mother's ability to conduct perspective

taking and the child's ability to grasp a handle for maximising the chances of success. Other joint actions require shared action plans that have to be extensively practiced alongside individual skills in order to maximise success. This would be the case for a beach volleyball team where the two players would learn to represent the other's action space in relation to their own and to the playing field. Furthermore the two SAS representations that each player generates would need to have substantial features in common for avoiding misunderstandings, collisions, etc. Hence, practice will have to improve their individual playing skills, their representation of the other's actions in a SAS, as well as the compatibility of their SAS at meta-level.

Up to now we have primarily considered perceptual, spatial and action-related determinants of SAS, such as the relative position of the two actors. Additional aspects such as the exact goals, action requirements and the social context play a crucial role but have only been assumed so far. However, it is likely that the formation and use of the SAS depend on socio-cognitive determinants such as the level of trust between the agents, group membership (in-group vs. out-group), etc. For example, recent evidence indicates that social exclusion is a determinant of action co-representation (Ambrosini et al., 2013; Costantini and Ferri, 2013).

Which FOR is chosen as the common, action-guiding FOR of a SAS can therefore depend on a variety of context factors such as the social relationship between the agents (e.g., hierarchy), the bodily ability for action (e.g., skill level, injury), or general characteristics of the social situation (e.g., "formality" of the situation as described in previous sections). Resuming our beach volleyball example, the SAS would differ if both players were equally good compared to when one player would clearly be the lead player, or if one player was a child or a learner, or if one player had suffered injury, e.g., was playing with an incapacitated arm affecting their action space on one particular side. It is also quite easy to imagine that SAS in this example would be quite different if it was a competitive game compared to more leisurely play.

In social structures with strong hierarchy, subjects tend to asymmetrically use their peripersonal/personal space. In military interactions, high-ranking agents use (move in) relatively more space than low-ranking agents (Dean et al., 1975). Signal integration is also influenced by social interaction. Heed et al. (2010) showed that the level of multisensory integration in peripersonal space is influenced by others actions in the same space and use of sensory signals.

We have used the example of tool-use to introduce the augmented space representation that usually follows agent-to-agent interaction. However, we are aware that a tool could only assume a passive role; there is no level of cooperation or interaction that could be described in tool-use. Thus, an important difference to tool-based extensions of an action space is that in agent-to-agent interaction spaces continuity of "force transmission" is not always important. In other words, tools but not other agents need to be physically manipulated. In the example of **Figure 1**, Agent 1 and Agent 3 have a common goal and successfully collaborate, both accessing the motor repertoire of Agent 2, without physically sharing parts of their peripersonal space. Recent research suggest

that action co-representation between agents (Sebanz et al., 2003) also emerges when actors are positioned in different rooms but believe they are collaborating (Tsai et al., 2006). Thus, physical interaction may not be a necessary condition for social collaboration and SAS but it seems to facilitate sharing in specific experimental scenarios, see e.g., Guagnano et al. (2010). Finally, it should be noted that social information has various levels of complexity and some subjects could only be able to share some of it. Autistic subjects have difficulties with sharing high-level social information; in particular with all functions included in so-called "theory of mind" (Baron-Cohen, 2000), however they display normal access to low-level social information (Sebanz et al., 2005).

CONCLUSIONS

Although it is well known that agents can have their abilities augmented by acting together with others, it is unclear how the brain mechanistically implements this process. Several mechanisms have been proposed that include entrainment, mutual prediction (Wilson and Knoblich, 2005), the sharing of representations (Sebanz et al., 2006), and a collective, we-mode of representation (Gallotti and Frith, 2013). In this article we argue that (at least some forms of) social interaction and social cognition (including cooperative and competitive ones) might be supported by the "social" re-use and re-calibration of the neuronal mechanisms for sensorimotor transformations and multisensory integration (Pouget et al., 2002).

We propose a *basis function* framework for social recalibration of sensorimotor representations; the resulting SAS are an embodied basis for joint action and sustained spatial and social perspective alignment. Coding the extended operational space and the social affordances created by the presence of co-actors in terms of *basis functions* for one's own, another's and joint actions could constitute a parsimonious solution to most interaction problems. This is especially evident if one assumes an ideomotor theory in which actions are coded in terms of their distal effects (Hommel et al., 2001). Co-actors sharing or merging their operational spaces can plausibly better plan, achieve, and monitor their joint goals. Future research would have to empirically assess our claims and in particular the proposed neuronal coding supporting "social" sensorimotor transformations that we have putatively identified as BFM.

A second important direction for future research is understanding if and how the mechanisms that we described for *spatial* perspective taking can be considered as an example from which we can extrapolate to other, more complex forms of perspective taking and social cognition. Indeed, there are various demonstrations that during social interactions and in particular joint actions co-actors share representations and "align" at multiple levels, besides purely spatial alignments; some examples are mimicry of behavior, sharing of cognitive representations and formation of a linguistic common ground (e.g., during linguistic exchanges) (Clark, 1996; Bargh and Chartrand, 1999; Sebanz et al., 2006; Garrod and Pickering, 2009).

Spatial and cognitive forms of alignment have several similarities and could use similar computational principles (although not necessarily the same neuronal mechanisms). For example, a key

aspect of the SAS is that it can be used for both planning one's own actions and predicting another's actions. The same feature is usually attributed to the common ground that is established during linguistic conversations (Clark, 1996) and to shared representations during joint actions (Sebanz et al., 2006).

Furthermore, we have emphasized that the SAS supports the alignment of individual FORs; one example would be the selection of a FOR centred on the body of the receiving person during a handout action. In a similar but more simplistic way, automatic mechanisms of resonance and mutual emulation are often advocated for the alignment of behavior (Bargh and Chartrand, 1999; Kessler and Miellet, 2013) and other forms of sharing and alignment (Garrod and Pickering, 2009), which in turn facilitate coordination.

In addition to automatic mechanisms co-actors can also adopt intentional strategies to form or calibrate a SAS. For example, co-actors (or a teacher and a student) can align spatially so that their operational spaces optimally overlap and the sensorimotor transformation does not require a complex rotation. In a similar way, intentional strategies of *signaling* help aligning the individualistic representations-for-action and “negotiating” a common plan for action (Pezzulo and Dindo, 2011); for example, a volleyball player can exaggerate her movements to signal a teammate that she is doing a left pass. Common ground formation during linguistic exchanges can help aligning the interlocutors' situation models, which in turn facilitate the interaction. Future studies would be needed to assess if all these processes link to our proposal of SAS.

This discussion suggests that spatial and cognitive forms of perspective taking are not disconnected but rather have bidirectional influences. In this vein, Spivey (2012) has argued that the spatial intersection of individuals is always also an intersection of minds, because portions of shared space are occupied by another cognitive agent whose cognitive states can intersect with one's own.

However, it is still unclear what are the mechanisms regulating the interactions between sharing action space and sharing cognitive representations. One simple explanation is that the mechanisms regulating spatial alignment and those regulating cognitive and affective evaluation (e.g., beliefs, liking, trust) of other persons are both regulated along the similar “positive—negative” dimension, and this can create positive feedback loops. For example, sharing a spatial operational space (or performing a joint action) can improve the positive beliefs (or affective reactions) and increase the trust in another person. In turn, because the persons now trust more one another, they come closer to one another and this in turn facilitates the sharing of their action space. The same mechanism can produce distrust and prevent the sharing of action spaces in other (e.g., competitive) situations. The plausibility of this hypothesis remains to be assessed empirically.

REFERENCES

- Aglioti, S., DeSouza, J. F., and Goodale, M. A. (1995). Size-contrast illusions deceive the eye but not the hand. *Curr. Biol.* 5, 679–685. doi: 10.1016/s0960-9822(95)00133-3
- Ambrosini, E., Blomberg, O., Mandrigin, A., and Costantini, M. (2013). Social exclusion modulates pre-reflective interpersonal body representation. *Psychol. Res.* doi: 10.1007/s00426-012-0476-2. [Epub ahead of print].
- Andersen, R. A. (1994). “Coordinate transformations and motor planning in posterior parietal cortex,” in *The Cognitive Neurosciences*, ed M. Gazzaniga (Cambridge, MA: The MIT Press), 519–532.
- Arbib, M., Bonaiuto, J., Jacobs, S., and Frey, S. (2009). Tool use and the distalization of the end-effector. *Psychol. Res.* 73, 441–462. doi: 10.1007/s00426-009-0242-2
- Avenanti, A., Buetti, D., Galati, G., and Aglioti, S. M. (2005). Transcranial magnetic stimulation highlights the sensorimotor side of empathy for pain. *Nat. Neurosci.* 8, 955–960. doi: 10.1038/nn1481
- Bargh, J. A., and Chartrand, T. L. (1999). The unbearable automaticity of being. *Am. Psychol.* 54, 462–479. doi: 10.1037/0003-066x.54.7.462
- Baron-Cohen, S. (2000). “Theory of mind and autism: a fifteen year review,” in *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience*, eds S. Baron-Cohen, H. Tager-Flusberg and D. J. Cohen (Oxford: Oxford University Press), 3–20.
- Becchio, C., Del Giudice, M., Dal Monte, O., Latini-Corazzini, L., and Pia, L. (2013). In your place: neuropsychological evidence for altercentric remapping in embodied perspective taking. *Soc. Cogn. Affect. Neurosci.* 8, 165–170. doi: 10.1093/scan/nsr083
- Berti, A., and Frassinetti, F. (2000). When far becomes near: remapping of space by tool use. *J. Cogn. Neurosci.* 12, 415–420. doi: 10.1162/089892900562237
- Biermann-Ruben, K., Kessler, K., Jonas, M., Siebner, H. R., Bäumer, T., Münchau, A., et al. (2008). Right hemisphere contributions to imitation tasks. *Eur. J. Neurosci.* 27, 1843–1855. doi: 10.1111/j.1460-9568.2008.06146.x
- Blakemore, S.-J., and Decety, J. (2001). From the perception of action to the understanding of intention. *Nat. Rev. Neurosci.* 2, 561–567. doi: 10.1038/35086023
- Blanke, O., and Mohr, C. (2005). Out-of-body experience, heautoscopy, and autoscopic hallucination of neurological origin implications for neurocognitive mechanisms of corporeal awareness and self-consciousness. *Brain Res. Brain Res. Rev.* 50, 184–199. doi: 10.1016/j.brainresrev.2005.05.008
- Bockler, A., Knoblich, G., and Sebanz, N. (2011). Giving a helping hand: effects of joint attention on mental rotation of body parts. *Exp. Brain Res.* 211, 531–545. doi: 10.1007/s00221-011-2625-z
- Botvinick, M., and Cohen, J. (1998). Rubber hands ‘feel’ touch that eyes see. *Nature* 391, 756. doi: 10.1038/35784
- Braithwaite, J. J., James, K., Dewe, H., Medford, N., Takahashi, C., and Kessler, K. (2013). Fractionating the unitary notion of dissociation: disembodied but not embodied dissociative experiences are associated with exocentric perspective-taking. *Front. Hum. Neurosci.* 7:719. doi: 10.3389/fnhum.2013.00719
- Bräuer, J., Call, J., and Tomasello, M. (2007). Chimpanzees really know what others can see in a competitive situation. *Anim. Cogn.* 10, 439–448. doi: 10.1007/s10071-007-0088-1
- Brugger, P., Regard, M., and Landis, T. (1997). Illusory reduplication of one's own body: phenomenology and classification of Autoscopic Phenomena. *Cognit. Neuropsychiatry* 2, 19–38. doi: 10.1080/135468097396397
- Buneo, C. A., and Andersen, R. A. (2006). The posterior parietal cortex: sensorimotor interface for the planning and online control of visually guided movements. *Neuropsychologia* 44, 2594–2606. doi: 10.1016/j.neuropsychologia.2005.10.011
- Buneo, C. A., Jarvis, M. R., Batista, A. P., and Andersen, R. A. (2002). Direct visuomotor transformations for reaching. *Nature* 416, 632–636. doi: 10.1038/416632a
- Caggiano, V., Fogassi, L., Rizzolatti, G., Thier, P., and Casile, A. (2009). Mirror neurons differentially encode the peripersonal and extrapersonal space of monkeys. *Science* 324, 403–406. doi: 10.1126/science.1166818
- Caminiti, R., Johnson, P. B., Galli, C., Ferraina, S., and Burnod, Y. (1991). Making arm movements within different parts of space: the premotor and motor cortical representation of a coordinate system for reaching to visual targets. *J. Neurosci.* 11, 1182–1197.
- Chang, S. W. C. (2013). Coordinate transformation approach to social interactions. *Front. Neurosci.* 7:147. doi: 10.3389/fnins.2013.00147
- Chang, S. W. C., Gariépy, J.-F., and Platt, M. L. (2013). Neuronal reference frames for social decisions in primate frontal cortex. *Nat. Neurosci.* 16, 243–250. doi: 10.1038/nn.3287
- Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.
- Colby, C. L., and Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annu. Rev. Neurosci.* 22, 319–349. doi: 10.1146/annurev.neuro.22.1.319
- Costantini, M., Ambrosini, E., Scorolli, C., and Borghi, A. M. (2011a). When objects are close to me: affordances in the peripersonal space. *Psychon. Bull. Rev.* 18, 302–308. doi: 10.3758/s13423-011-0054-4

- Costantini, M., Ambrosini, E., Tieri, G., Sinigaglia, C., and Committeri, G. (2010). Where does an object trigger an action? An investigation about affordances in space. *Exp. Brain Res.* 207, 95–103. doi: 10.1007/s00221-010-2435-8
- Costantini, M., Committeri, G., and Sinigaglia, C. (2011b). Ready both to your and to my hands: mapping the action space of others. *PLoS One* 6:e17923. doi: 10.1371/journal.pone.0017923
- Costantini, M., and Ferri, F. (2013). Action co-representation and social exclusion. *Exp. Brain Res.* 227, 85–92. doi: 10.1007/s00221-013-3487-3
- Dean, L. M., Willis, F. N., and Hewitt, J. (1975). Initial interaction distance among individuals equal and unequal in military rank. *J. Pers. Soc. Psychol.* 32, 294–299. doi: 10.1037/0022-3514.32.2.294
- Dindo, H., Zambuto, D., and Pezzulo, G. (2011). “Motor simulation via coupled internal models using sequential Monte Carlo,” in *Proceedings of IJCAI*, 2113–2119.
- Duhamel, J. R., Colby, C. L., and Goldberg, M. E. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science* 255, 90–92. doi: 10.1126/science.1553535
- Emery, N. J., and Clayton, N. S. (2009). Comparative social cognition. *Annu. Rev. Psychol.* 60, 87–113. doi: 10.1146/annurev.psych.60.110707.163526
- Ferraina, S., Battaglia-Mayer, A., Genovesio, A., Archambault, P., and Caminiti, R. (2009a). Parietal encoding of action in depth. *Neuropsychologia* 47, 1409–1420. doi: 10.1016/j.neuropsychologia.2008.12.028
- Ferraina, S., Brunamonti, E., Giusti, M. A., Costa, S., Genovesio, A., and Caminiti, R., et al. (2009b). Reaching in depth: hand position dominates over binocular eye position in the rostral superior parietal lobule. *J. Neurosci.* 29, 11461–11470. doi: 10.1523/jneurosci.1305-09.2009
- Flavell, J. H., Everett, B. A., Croft, K., and Flavell, E. R. (1981). Young children's knowledge about visual perception: further evidence for the Level 1–Level 2 distinction. *Dev. Psychol.* 17, 99–103. doi: 10.1037/0012-1649.17.1.99
- Frith, C. D., and Frith, U. (2008). Implicit and explicit processes in social cognition. *Neuron* 60, 503–510. doi: 10.1016/j.neuron.2008.10.032
- Furlanetto, T., Bertone, C., and Becchio, C. (2013). The bilocated mind: new perspectives on self-localization and self-identification. *Front. Hum. Neurosci.* 7:71. doi: 10.3389/fnhum.2013.00071
- Gallotti, M., and Frith, C. D. (2013). Social cognition in the we-mode. *Trends Cogn. Sci.* 17, 160–165. doi: 10.1016/j.tics.2013.02.002
- Ganel, T., Tanzer, M., and Goodale, M. A. (2008). A double dissociation between action and perception in the context of visual illusions: opposite effects of real and illusory size. *Psychol. Sci.* 19, 221–225. doi: 10.1111/j.1467-9280.2008.02071.x
- Garrod, S., and Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Top. Cogn. Sci.* 1, 292–304. doi: 10.1111/j.1756-8765.2009.01020.x
- Goodale, M. A., and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends Neurosci.* 15, 20–25. doi: 10.1016/0166-2236(92)90344-8
- Grafton, S. T. (2009). Embodied cognition and the simulation of action to understand others. *Ann. N Y Acad. Sci.* 1156, 97–117. doi: 10.1111/j.1749-6632.2009.04425.x
- Graziano, M. S., Yap, G. S., and Gross, C. G. (1994). Coding of visual space by premotor neurons. *Science* 266, 1054–1057. doi: 10.1126/science.7973661
- Gross, C. G., and Graziano, M. S. A. (1995). Multiple representations of space in the brain. *Neuroscientist* 1, 43–50. doi: 10.1177/107385849500100107
- Guagnano, D., Rusconi, E., and Umiltà, C. A. (2010). Sharing a task or sharing space? On the effect of the confederate in action coding in a detection task. *Cognition* 114, 348–355. doi: 10.1016/j.cognition.2009.10.008
- Head, H., and Holmes, G. (1911). Sensory disturbances from cerebral lesions. *Brain* 34, 102–254. doi: 10.1093/brain/34.2-3.102
- Heed, T., Habets, B., Sebanz, N., and Knoblich, G. (2010). Others' actions reduce crossmodal integration in peripersonal space. *Curr. Biol.* 20, 1345–1349. doi: 10.1016/j.cub.2010.05.068
- Hommel, B., Musseler, J., Aschersleben, G., and Prinz, W. (2001). The theory of event coding (TEC): a framework for perception and action planning. *Behav. Brain Sci.* 24, 849–878; discussion 878–937. doi: 10.1017/S0140525X01000103
- Ishida, H., Nakajima, K., Inase, M., and Murata, A. (2010). Shared mapping of own and others' bodies in visuotactile bimodal area of monkey parietal cortex. *J. Cogn. Neurosci.* 22, 83–96. doi: 10.1162/jocn.2009.21185
- Janczyk, M. (2013). Level 2 perspective taking entails two processes: evidence from PRP experiments. *J. Exp. Psychol. Learn. Mem. Cogn.* doi: 10.1162/jocn.2009.21185. [Epub ahead of print].
- Jeannerod, M. (2006). *Motor Cognition*. Oxford: Oxford University Press.
- Jeannerod, M., and Biguer, B. (1989). [Egocentric reference and represented space]. *Rev. Neurol.* 145, 635–639.
- Johnson, M., and Demiris, Y. (2005). Perceptual perspective taking and action recognition. *Int. J. Adv. Robot. Syst.* 2, 301–309. doi: 10.5772/5775
- Karnath, H. O., Christ, K., and Hartje, W. (1993). Decrease of contralateral neglect by neck muscle vibration and spatial orientation of trunk midline. *Brain* 116, 383–396. doi: 10.1093/brain/116.2.383
- Kawato, M. (1999). Internal models for motor control and trajectory planning. *Curr. Opin. Neurobiol.* 9, 718–727. doi: 10.1016/s0959-4388(99)00028-8
- Kelso, J. A. S., Dumas, G., and Tognoli, E. (2013). Outline of a general theory of behavior and brain coordination. *Neural Netw.* 37, 120–131. doi: 10.1016/j.neunet.2012.09.003
- Kessler, K. (2000). *Spatial Cognition and Verbal Localisations: A Connectionist Model for the Interpretation of Spatial Prepositions*. Wiesbaden: Deutscher Universitäts-Verlag.
- Kessler, K., Biermann-Ruben, K., Jonas, M., Roman Siebner, H., Bäumer, T., Münchau, A., et al. (2006). Investigating the human mirror neuron system by means of cortical synchronization during the imitation of biological movements. *Neuroimage* 33, 227–238. doi: 10.1016/j.neuroimage.2006.06.014
- Kessler, K., and Garrod, S. (2013). Editorial: cortex discussion forum on “The meaning of mirror neurons”. *Cortex* (in press). doi: 10.1016/j.cortex.2013.09.009
- Kessler, K., and Miellet, S. (2013). Perceiving conspecifics as integrated body-gestalts is an embodied process. *J. Exp. Psychol. Gen.* 142, 774–790. doi: 10.1037/a0029617
- Kessler, K., and Rutherford, H. (2010). The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions. *Front. Psychol.* 1:213. doi: 10.3389/fpsyh.2010.00213
- Kessler, K., and Thomson, L. A. (2010). The embodied nature of spatial perspective taking: embodied transformation versus sensorimotor interference. *Cognition* 114, 72–88. doi: 10.1016/j.cognition.2009.08.015
- Kessler, K., and Wang, H. (2012). Spatial perspective taking is an embodied process, but not for everyone in the same way: differences predicted by sex and social skills score. *Spat. Cogn. Comput.* 12, 133–158. doi: 10.1080/13875868.2011.634533
- Koulovskov, A. A., and Rinberg, D. (2011). In search of the structure of human olfactory space. *Front. Syst. Neurosci.* 5:65. doi: 10.3389/fnsys.2011.00065
- Lacquaniti, F., Guigon, E., Bianchi, L., Ferraina, S., and Caminiti, R. (1995). Representing spatial information for limb movement: role of area 5 in the monkey. *Cereb. Cortex* 5, 391–409. doi: 10.1093/cercor/5.5.391
- Larry, M., and Dean, F. N. W. (1975). Initial interaction distance among individuals equal and unequal in military rank. *J. Pers. Soc. Psychol.* 32, 294–299. doi: 10.1037/0022-3514.32.2.294
- Lee, D. (2008). Game theory and neural basis of social decision making. *Nat. Neurosci.* 11, 404–409. doi: 10.1038/nn2065
- Maravita, A., and Iriki, A. (2004). Tools for the body (schema). *Trends Cogn. Sci.* 8, 79–86. doi: 10.1016/j.tics.2003.12.008
- Michelon, P., and Zacks, J. M. (2006). Two kinds of visual perspective taking. *Percept. Psychophys.* 68, 327–337. doi: 10.3758/bf03193680
- Milner, A. D., and Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia* 46, 774–785. doi: 10.1016/j.neuropsychologia.2007.10.005
- Neal, A., and Kilner, J. M. (2010). What is simulated in the action observation network when we observe actions? *Eur. J. Neurosci.* 32, 1765–1770. doi: 10.1111/j.1460-9568.2010.07435.x
- Newman-Norlund, R. D., van Schie, H. T., van Zuijlen, A. M. J., and Bekkering, H. (2007). The mirror neuron system is more active during complementary compared with imitative action. *Nat. Neurosci.* 10, 817–818. doi: 10.1038/nn1911
- Pesaran, B., Nelson, M. J., and Andersen, R. A. (2006). Dorsal premotor neurons encode the relative position of the hand, eye, and goal during reach planning. *Neuron* 51, 125–134. doi: 10.1016/j.neuron.2006.05.025
- Pezzulo, G. (2013). Studying mirror mechanisms within generative and predictive architectures for joint action. *Cortex* doi: 10.1016/j.cortex.2013.06.008. [Epub ahead of print].
- Pezzulo, G. (2011a). Grounding procedural and declarative knowledge in sensorimotor anticipation. *Mind Lang.* 26, 78–114. doi: 10.1111/j.1468-0017.2010.01411.x

- Pezzulo, G. (2011b). Shared representations as coordination tools for interactions. *Rev. Philos. Psychol.* 2, 303–333. doi: 10.1007/s13164-011-0060-5
- Pezzulo, G., and Dindo, H. (2011). What should I do next? Using shared representations to solve interaction problems. *Exp. Brain Res.* 211, 613–630. doi: 10.1007/s00221-011-2712-1
- Pezzulo, G., Candidi, M., Dindo, H., and Barca, L. (2013). Action simulation in the human brain: twelve questions. *New Ideas Psychol.* 31, 270–290. doi: 10.1016/j.newideapsych.2013.01.004
- Pezzulo, G., Hoffmann, J., and Falcone, R. (2007). Anticipation and anticipatory behavior. *Cogn. Process.* 8, 67–70. doi: 10.1007/s10339-007-0173-z
- Pouget, A., Deneve, S., and Duhamel, J.-R. (2002). A computational perspective on the neural basis of multisensory spatial representations. *Nat. Rev. Neurosci.* 3, 741–747. doi: 10.1038/nrn914
- Pouget, A., and Sejnowski, T. J. (1996). “A model of spatial representations in parietal cortex explains hemineglect,” in *Advances in Neural Information Processing Systems*, Volume 8, eds D. S. Touretzky, M. Mozer and M. Hasselmo (Cambridge, MA: MIT Press), 10–16.
- Pouget, A., and Sejnowski, T. J. (1995). “Spatial representations in the parietal cortex may use basis functions,” in *Advances in Neural Information Processing Systems*, Volume 7, eds G. Tesauro, D. S. Touretzky and T. Leen (Cambridge, MA: MIT Press), 157–164.
- Pouget, A., and Snyder, L. H. (2000). Computational approaches to sensorimotor transformations. *Nat. Neurosci.* 3 Suppl., 1192–1198. doi: 10.1038/81469
- Previc, F. H. (1998). The neuropsychology of 3-D space. *Psychol. Bull.* 124, 123–164. doi: 10.1037/0033-2909.124.2.123
- Prinz, W. (1997). Perception and action planning. *Eur. J. Cogn. Psychol.* 9, 129–154. doi: 10.1080/713752551
- Richardson, M. J., Marsh, K. L., and Baron, R. M. (2007). Judging and actualizing intrapersonal and interpersonal affordances. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 845–859. doi: 10.1037/0096-1523.33.4.845
- Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192. doi: 10.1146/annurev.neuro.27.070203.144230
- Rizzolatti, G., Fadiga, L., Fogassi, L., and Gallese, V. (1997). The space around us. *Science* 277, 190–191. doi: 10.1126/science.277.5323.190
- Rizzolatti, G., and Luppino, G. (2001). The cortical motor system. *Neuron* 31, 889–901. doi: 10.1016/s0896-6273(01)00423-8
- Rosenbaum, D. A., Meulenbroek, R. J., and Vaughan, J. (2001). Planning reaching and grasping movements: theoretical premises and practical implications. *Motor Control* 2, 99–115.
- Saygin, A. P. (2007). Superior temporal and premotor brain areas necessary for biological motion perception. *Brain* 130, 2452–2461. doi: 10.1093/brain/awm162
- Sebanz, N., Bekkering, H., and Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends Cogn. Sci.* 10, 70–76. doi: 10.1016/j.tics.2005.12.009
- Sebanz, N., and Knoblich, G. (2009). Prediction in joint action: what, when, and where. *Top. Cogn. Sci.* 1, 353–367. doi: 10.1111/j.1756-8765.2009.01024.x
- Sebanz, N., Knoblich, G., and Prinz, W. (2003). Representing others’ actions: just like one’s own? *Cognition* 88, B11–B21. doi: 10.1016/S0010-0277(03)00043-X
- Sebanz, N., Knoblich, G., Stumpf, L., and Prinz, W. (2005). Far from action-blind: representation of others’ actions in individuals with Autism. *Cogn. Neuropsychol.* 22, 433–454. doi: 10.1080/02643290442000121
- Seyama, J., and Nagayama, R. (2005). The effect of torso direction on the judgement of eye direction. *Vis. Cogn.* 12, 103–116. doi: 10.1080/13506280444000111
- Spivey, M. J. (2012). The spatial intersection of minds. *Cogn. Process.* 13(Suppl. 1), S343–S346. doi: 10.1007/s10339-012-0520-6
- Tankersley, D., Stowe, C. J., and Huettel, S. A. (2007). Altruism is associated with an increased neural response to agency. *Nat. Neurosci.* 10, 150–151. doi: 10.1038/nn1833
- Thomas, R., Press, C., and Haggard, P. (2006). Shared representations in body perception. *Acta Psychol. (Amst.)* 121, 317–330. doi: 10.1016/j.actpsy.2005.08.002
- Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* 28, 675–691; discussion 691–735. doi: 10.1017/s0140525x05000129
- Tsai, C.-C., Kuo, W.-J., Jing, J.-T., Hung, D. L., and Tzeng, O. J.-L. (2006). A common coding framework in self-other interaction: evidence from joint action task. *Exp. Brain Res.* 175, 353–362. doi: 10.1007/s00221-006-0557-9
- Tversky, B., and Hard, B. M. (2009). Embodied and disembodied cognition: spatial perspective-taking. *Cognition* 110, 124–129. doi: 10.1016/j.cognition.2008.10.008
- Van Essen, D. C., and Anderson, C. H. (1990). “Reference frames and dynamic remapping processes in vision,” in *Computational Neuroscience*, ed E. Schwartz (Cambridge, MA: MIT Press), 278–294.
- Waytz, A., Zaki, J., and Mitchell, J. P. (2012). Response of dorsomedial prefrontal cortex predicts altruistic behavior. *J. Neurosci.* 32, 7646–7650. doi: 10.1523/jneurosci.6193-11.2012
- Wilson, M., and Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychol. Bull.* 131, 460–473. doi: 10.1037/0033-2909.131.3.460
- Witt, J. K., and Proffitt, D. R. (2008). Action-specific influences on distance perception: a role for motor simulation. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 1479–1492. doi: 10.1037/a0010781
- Wolpert, D. M., Doya, K., and Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 358, 593–602. doi: 10.1098/rstb.2002.1238
- Wolpert, D. M., Gharamani, Z., and Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science* 269, 1880–1882. doi: 10.1126/science.7569931
- Zacks, J., and Michelon, P. (2005). Transformations of visuospatial images. *Behav. Cogn. Neurosci. Rev.* 4, 96–118. doi: 10.1177/1534582305281085

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 12 July 2013; accepted: 03 November 2013; published online: 26 November 2013.

Citation: Pezzulo G, Iodice P, Ferraina S and Kessler K (2013) Shared action spaces: a basis function framework for social re-calibration of sensorimotor representations supporting joint action. *Front. Hum. Neurosci.* 7:800. doi: 10.3389/fnhum.2013.00800 This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2013 Pezzulo, Iodice, Ferraina and Kessler. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Through your eyes: incongruence of gaze and action increases spontaneous perspective taking

Tiziano Furlanetto¹, Andrea Cavallo¹, Valeria Manera^{1,2}, Barbara Tversky² and Cristina Becchio^{1*}

¹ Department of Psychology, Center for Cognitive Science, University of Torino, Torino, Italy

² Department of Psychology, Stanford University, Stanford, CA, USA

Edited by:

Klaus Kessler, University of Glasgow, UK

Reviewed by:

Andrew Bayliss, University of East Anglia, UK

Dana Samson, Université Catholique de Louvain, Belgium

*Correspondence:

Cristina Becchio, Department of Psychology, University of Turin; Via Po, 14, 10123 Turin, Italy
e-mail: cristina.becchio@unito.it

What makes people spontaneously adopt the perspective of others? Previous work suggested that perspective taking can serve understanding the actions of others. Two studies corroborate and extend that interpretation. The first study varied cues to intentionality of eye gaze and action, and found that the more the actor was perceived as potentially interacting with the objects, the stronger the tendency to take his perspective. The second study investigated how manipulations of gaze affect the tendency to adopt the perspective of another reaching for an object. Eliminating gaze cues by blurring the actor's face did not reduce perspective-taking, suggesting that in the absence of gaze information, observers rely entirely on the action. Intriguingly, perspective-taking was higher when gaze and action did not signal the same intention, suggesting that in presence of ambiguous behavioral intention, people are more likely take the other's perspective to try to understand the action.

Keywords: spontaneous perspective taking, agency, action, gaze, incongruous cues, ambiguous intention

INTRODUCTION

Near/far, above/below, right/left presuppose a referential center of orientation. Because we cannot separate ourselves from our bodies, it is natural to think that this center of orientation is the body. As Husserl put it, "the 'far' is far from me, from my Body; the 'to the right' refers to the right side of my Body" (1952/1989). But what happens in presence of others? Are there circumstances where "to the left" with respect to another's body is preferred to "to the right" with respect to my own?

Evidence that the presence of others may change our own coding of spatial locations of objects is provided by recent studies investigating spatial judgment (Tversky and Hard, 2009; Zwickel, 2009; Zwickel and Müller, 2010). In a typical experiment, participants viewed a photograph of two objects on a table. When participants were asked to describe the location of one object relative to another, the dominant response was to adopt their own spatial perspective. If, however, the scene included a person looking or reaching for one of the objects, almost one third of participants spontaneously adopted the other person's perspective, describing the locations from the other's right or left (Tversky and Hard, 2009). These findings indicate that the presence of another person may encourage participants to spontaneously take that person spatial perspective, and describe the locations of the objects from her right or left. Similarly, studies investigating spontaneous visual perspective taking found that observers were slower to make self-perspective judgments when the scene includes a person looking at the scene from a different visual perspective, suggesting that even when the other person's perspective is irrelevant to the task, observers cannot prevent computing the other's perspective (Samson et al., 2010).

What makes people spontaneously take others' perspectives despite the very real presence of their own? The "mere presence"

of a human body does not seem sufficient to elicit spontaneous perspective taking (Mazzarella et al., 2012). People adopt the perspective of another person who acts (Frischen et al., 2009; Thiriaux et al., 2010) or is positioned to act on objects and even more so when attention is drawn to the person's potential for action, for instance, by phrasing the query about spatial relations in terms of action (e.g., "In relation to the bottle, where does he place the book?" Tversky and Hard, 2009). What is more, people even adopt the perspective of simple geometric shapes when the actions of the shapes appear intentional (Zwickel, 2009).

Together, the research suggests that spontaneous perspective taking may be related to understanding and anticipating another's action rather than to the mere presence of a human body. If so, perspective taking should increase when the perceived intention to act increases. This prediction was tested in the first of two experiments. Participants were presented with brief videos (rather than still photographs) depicting two objects, a milk carton and a glass full of milk, on a table, with or without a person behind (see Figure 1). Because looking at an object often signals intention to act on the object (Allison et al., 2000; Mennie et al., 2007; Becchio et al., 2008; Pierno et al., 2008; Sartori et al., 2009; Innocenti et al., 2012), the tendency to take the actor's perspective should be stronger when the actor looks at one of the objects and even stronger when the actor reaches toward the object.

Findings concerning the contribution of gaze cues to spontaneous perspective taking have not been consistent. Using static photographs, Tversky and Hard (2009) found no significant difference in the proportion of description from the other's point of view for looking and looking-and-reaching scenes, suggesting that gaze shifts and overt hand actions have similar effects on perspective taking. In contrast, however, Mazzarella et al. (2012) found that the actor's hand action, but not the actor's gaze,

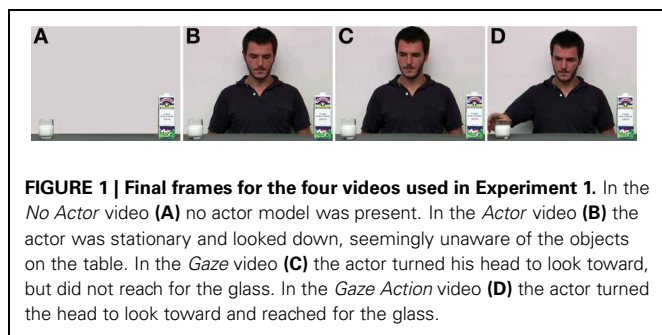


FIGURE 1 | Final frames for the four videos used in Experiment 1. In the *No Actor* video (A) no actor model was present. In the *Actor* video (B) the actor was stationary and looked down, seemingly unaware of the objects on the table. In the *Gaze* video (C) the actor turned his head to look toward, but did not reach for the glass. In the *Gaze Action* video (D) the actor turned the head to look toward and reached for the glass.

modulated the tendency to adopt his perspective. The question remains therefore open as to whether gaze contributes to perspective taking. To address this issue, in a second experiment we manipulated the congruency of gaze and action cues. Gaze cues can be informative but also produce ambiguity with respect to others' actions and behavioral intentions. For instance, football and basketball players often "fake" to fool their opponents, by looking in one direction and acting in another. We predicted that if perspective taking is related to understanding another's action, then, by making the agent's intention ambiguous, incongruous gaze would increase perspective taking.

STUDY 1: SPONTANEOUS PERSPECTIVE TAKING INCREASES AS PERCEIVED INTENTIONALITY INCREASES

Study 1 was designed to test whether the perceived potential for interaction with objects increases spontaneous perspective taking. We predicted that the more a person is perceived as potentially acting on an object, the greater the need to understand the action, hence in the scene, the stronger the tendency to spatially represent the locations of the objects from the actor's perspective.

METHODS

Participants

One hundred and twenty undergraduate students (53 male and 67 female; mean age: 23.5 ± 3.3 , range 18–37 years) from the University of Turin volunteered to take part in the experiment. All had normal or corrected-to-normal vision, were right handed, and were naïve with respect to the purpose of the study.

Materials and procedures

Participants were presented with one of four videos depicting two objects, a milk cartoon and a glass full of milk, on a table. Scene information was manipulated by introducing an actor model and by varying the actor's gaze and action (see **Figure 1**). In the *No Actor* video ($n = 30$) no actor model was present. The other three videos included an actor. The *Actor* video ($n = 30$) showed the actor stationary, looking down, seemingly unaware of the objects on the table. In the *Gaze* video ($n = 30$) the actor turned his head to look toward the glass, but did not reach it. In the *Gaze Action* video ($n = 30$) the actor turned the head to look toward and reached for the glass. Videos including the actor started with the actor looking down for 2 s. The actor then turned his head to look at the object (in the *Gaze* and the *Gaze Action* video) and, after 1 s, reached for the object (in the *Gaze Action* video). Each video lasted 4.15 s. The question "In relation to the glass, where

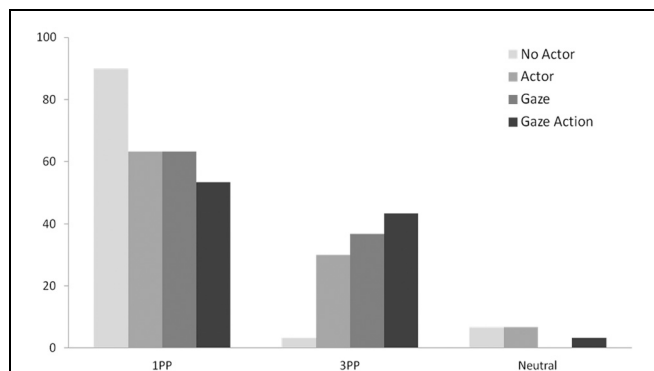


FIGURE 2 | Percentages of 1PP, 3PP, and neutral responses in Experiment 1.

is the milk cartoon?" was displayed below the last frame of each video and remained visible until response or until 9 s elapsed. Participants' verbal responses were recorded by the experimenter who was sitting behind the participant.

DATA ANALYSIS AND RESULTS

The responses were scored as 1PP (first person perspective) if the answer was from the participant's point of view, 3PP (third person perspective) if the answer was from the actor's viewpoint, and neutral if the answer gave spatial information from neither perspective (e.g., "next to," "to the side," "on the table"). Examples of responses scored as 1PP include: "right," "on the right," "to the right from my perspective." Examples of responses scored as 3PP include: "left," "to his left," "to the left from his perspective." Scored responses were converted into three binary variables for analysis: one variable was coded 1 if the response was 3PP and 0 if it was not; the second variable was coded 1 if the response was 1PP and 0 if it was not; the third variable was coded 1 if the response was neutral and 0 if it was not. To assess the influence of agency cues on spontaneous perspective taking, separate binary logistic regression analyses were conducted on 3PP, 1PP, and neutral responses. The type of video (*No Actor*, *Actor*, *Gaze*, and *Gaze Action*) was entered as independent variable of interest.

In line with predictions, binary logistic regression analysis on 3PP responses yielded a significant linear effect of agency (Wald $\chi^2 = 10.903$, $df = 1$, odd ratio = 1.968, CI = 1.317–2.941, $p = 0.001$). The percentage of 3PP responses was highest for *Gaze Action* video (43.3%), lower for the *Gaze* video (36.7%), and even lower for the *Actor* video (30%, see **Figure 2**). For the *No actor* video, only one participant adopted the 3PP perspective. Similarly, the percentage of 1PP responses was affected by agency cues (Wald $\chi^2 = 7.872$, $df = 1$, odd ratio = 0.591, CI = 0.409–0.853, $p = 0.005$). The percentage of 1PP responses was highest for the *No Actor* video (90%), lower for the *Actor* video (63.3%) and the *Gaze* (63.3%) video, and lowest for the *Gaze Action* video (53.3%, see **Figure 2**). The percentage of neutral responses was not affected by agency cues (Wald $\chi^2 = 0.993$, odd ratio = 0.645, CI = 0.272–1.528, $p = 0.319$).

Together, these findings corroborate and extend the idea that increased potential for interaction enhances spontaneous perspective taking.

STUDY 2: SPONTANEOUS PERSPECTIVE TAKING INCREASES AS INCONGRUITY OF INTENTION INCREASES

Gaze is an important source of information about others' intentions and actions (Allison et al., 2000; Mennie et al., 2007; Becchio et al., 2008; Pierno et al., 2008; Sartori et al., 2009; Innocenti et al., 2012). From the gaze of another person, we can infer what the person is interested in, what she might desire, and, consequently, what she will do next (Pierno et al., 2006). Gaze direction, however, can also produce ambiguity with respect to the other's intention. This can occur, when gaze conveys conflicting information with respect to the behavioral intention of the agent (Hudson and Jellema, 2011). In this situation, the agent's action can be perceived as ambiguous and observers might be encouraged to adopt the perspective of the other person to understand her intention. Spontaneous perspective taking might thus be expected to be even stronger when gaze is incongruous than when gaze and action signal the same, and therefore unambiguous, intention.

To test this prediction, in Study 2, we presented participants with videos of an actor reaching for a glass in presence of a milk cartoon. The actor either looked toward the glass before reaching (*Gaze Action*) or reached without looking (*Ambiguous Gaze Action*). We predicted that the absence of a shift of gaze in the direction of action would make the action harder to understand and therefore increase the likelihood of adopting the actor perspective. In contrast, no increase in perspective taking should be expected when access to the actor's gaze during reaching is prevented by blurring the actor's face (*Blurred Gaze Action*). This is because, in this situation, the absence of gaze cues does not render the agent's behavioral intention ambiguous.

METHODS

Participants

Based on the prevalence of 3PP/1PP responses for the *Gaze Action* scene compared to the *Actor* scene and the *Gaze* scene in Experiment 1 (9.7%), we estimated that we would need 135 participants in each condition to evaluate the effect of gaze manipulations on 3PP and 1PP responses (see Supplementary Material). Four hundred and five undergraduate students (191 male and 214 female; mean age: 23.3 ± 3.3 ; range 18–48 years) from the University of Turin were thus recruited to take part in Experiment 2. All had normal or corrected-to-normal vision, were right handed, and were naïve with respect to the purpose of the study.

Materials and procedures

Procedures were the same as those in Study 1, except that participants were presented with one of three videos depicting an actor reaching for one of two objects—a milk cartoon and a glass full of milk—on a table (see Figure 3). In the *Gaze Action* video ($n = 135$) the actor turned his head, looked toward and reached for the glass (see Study 1). In the *Blurred Gaze Action* video ($n = 135$) the actor turned his head, looked toward and reached for the glass as in the *Gaze Action* video. Access to the actor's gaze direction was, however, prevented by blurring the actor's face. In the *Ambiguous Gaze Action* video ($n = 135$) the actor reached for the glass without looking at it.



FIGURE 3 | Final frames for videos in Experiment 2. In the *Gaze Action* video (A) the actor turned the head to look toward and reached for the glass. In the *Blurred Gaze Action* video (B) the actor turned his head, looked toward and reached for the glass but participant's access to the actor's gaze direction was prevented by blurring the actor's face. In the *Ambiguous Gaze Action* video (C), the actor reached the glass without looking toward it.

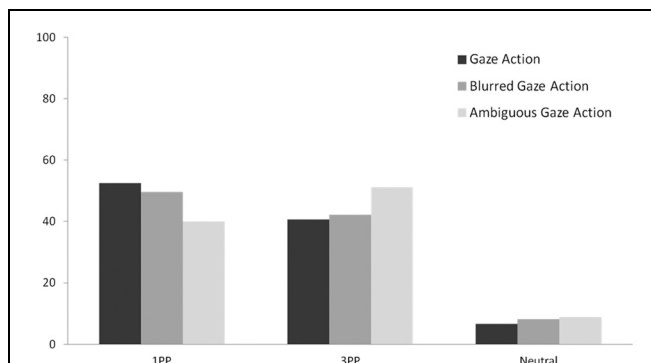


FIGURE 4 | Percentages of 1PP, 3PP and neutral responses in Experiment 2.

DATA ANALYSIS AND RESULTS

As in Study 1, the responses were scored as 1PP if the answer was from the participant's point of view, 3PP if the answer was from the actor's viewpoint, and neutral if the answer gave spatial information from neither perspective. Separate chi square analyses were conducted to compare observed frequencies of 3PP (vs. 1PP and neutral responses), 1PP (vs. 3PP and neutral responses), and neutral responses (vs. 1PP and 3PP responses) for the *Ambiguous Gaze Action* and for the *Blurred Gaze Action* scenes with expected frequencies for the *Gaze Action* scene.

Strikingly, when the actor reached without looking, 51.1% of the participants adopted his perspective (see Figure 4). Chi-square analysis revealed a marginally significant increase in 3PP responses for the *Ambiguous Gaze Action* scene compared to the *Gaze Action* scene (51.1% vs. 40.7%; $\chi^2 = 3.713$, $df = 1$, $p = 0.054$, $r = 0.166$). Conversely, 1PP responses were significantly lower for videos in which the actor reached without looking than for videos in which reaching was preceded by looking (40% vs. 52.6%; $\chi^2 = 8.586$, $df = 1$, $p = 0.003$, $r = 0.252$). Taken together, these findings suggest that perspective taking was increased for the *Ambiguous Gaze Action* scene compared to the *Gaze Action* scene. As predicted, *Blurred Gaze Action* and *Gaze Action* videos yielded equivalent percentages of 3PP (42.2% vs. 40.7%; $\chi^2 = 0.112$, $df = 1$, $p = 0.738$, $r = 0.029$) and 1PP responses (52.5% vs. 49.6%; $\chi^2 = 2.478$, $df = 1$, $p = 0.115$, $r = 0.029$). Neutral responses were neither affected by the ambiguity of actor's intentions (6.6% vs. 8.8%; $\chi^2 = 1.071$, $df = 1$,

$p = 0.301$, $r = 0.089$) nor by the gaze blurring (6.6% vs. 8.1%; $\chi^2 = 0.476$, $df = 1$, $p = 0.490$, $r = 0.059$).

GENERAL DISCUSSION

Converging evidence from social neuroscience suggests that people use knowledge of their own bodies to understand other people's behavior (Grafton, 2009). Accordingly, understanding of others' actions, intentions, and emotions has been proposed to rely on mechanism of embodied simulation (e.g., Becchio et al., 2012). Together with previous research (e.g., Tversky and Hard, 2009), the present results suggest that, in the service of action understanding, people may also embody others' location, spatially representing the world from others' point of view rather than from their own.

AGENCY CUES IN VIDEO DISPLAYS

The mere presence of another person in a position to act on objects encouraged about 30% of respondents to take the other person's perspective. Critically, as demonstrated in Study 1, the tendency to take the actor's perspective increased when the actor looked at one of the objects (36.7%) and became even stronger when the actor reached for one object (43.3%). This corroborates the interpretation that perspective taking increases to the extent that the person is perceived to be potentially interacting with the objects.

Previous studies investigating spontaneous perspective taking have reported considerably lower percentages of third-person responses for looking and reaching scenes than those reported here (e.g., 22 and 29%, respectively; Tversky and Hard, 2009). One aspect of the present study that is likely to have contributed to increase perspective taking is the use of videos instead of photographs. Videos provide dynamic cues to action not available in static displays. As human observers are particularly sensitive to human body movements (Blake and Shiffrar, 2007), it is plausible that the gradual unfolding of action emphasizes and draws attention to action, thereby increasing perspective taking.

A question for future research is whether perspective taking is further encouraged by the observation of actions potentially directed at the observer. Social cognition has been proposed to be substantially different when we are in interaction with others (second-person interaction) rather than merely observing them (third-person interaction; Schilbach et al., 2013). Second-person interaction modulates emphatic brain responses (Singer et al., 2006) and there is evidence that simulation of another person's action, as reflected in the activation of the observer motor system, gets stronger the more the other is perceived as an interaction partner (Kourtis et al., 2010). In terms of perspective taking, observation of the actions of a potentially interacting partner might thus be expected to elicit stronger perspective compared to observation of the actions of a third party we do not interact with.

WHEN LOOKING IS AMBIGUOUS

Mazzarella et al. (2012) reported that action triggered perspective-taking, but gaze cues did not. They suggest that this may be because eye gaze is not critically relevant, as grasping

is, to understanding what an actor is currently doing. However, other research has shown that other gaze direction is informative not only about *future intentions* but also about *present intentions* and *motor intentions* and can change the way current actions are perceived (Pierno et al., 2008). Reaching is typically guided by the eyes. Gaze leads the hand to the object to be grasped and supports predictive motor control in manipulation (Johansson et al., 2001). Observing a person grasping without looking may thus be perceived as ambiguous. What is he planning to do? Why is he not looking at the object he is reaching for? In Experiment 2 we found that compared to a situation in which gaze and action signaled the same intention, perspective taking increased for reaching without looking, apparently in an effort to understand the intended action in the face of conflicting cues. In contrast, we observed no increase in perspective taking when looking cues were eliminated by blurring the eyes, suggesting that when there was no conflict, observers used the direction of reaching as a cue to understand the intention.

Allocation of attention to gaze cues is a flexible process that depends in part on the perceived ambiguity of an agent's intentions. Observers do not attend to an agent's gaze direction automatically, but rather do so when other social cues are insufficient to determine the immediate course or goal of the action (Hudson and Jellema, 2011). The present findings suggest that similarly to attention, perspective taking may not be triggered directly by the perceptual properties of gaze stimuli, but may depend on gaze intentional significance in the overall context. When gaze and action cues convey the same information, gaze processing adds little to action in terms of intention attribution. Eliminating gaze cues has thus no influence on perspective taking. However, when gaze and action convey incongruous information making the agent's intention ambiguous, gaze direction becomes relevant and may increase spontaneous perspective taking. These findings may help to reconcile inconsistent findings concerning the relative contributions of gaze and action cues to perspective taking (e.g., Tversky and Hard, 2009; Mazzarella et al., 2012) by showing that, rather than depending on specific bodily cues (and not others), perspective taking is influenced by the attribution of intentions to others.

CONCLUSIONS

Here, participants watched videos of two objects on a table under varying conditions. They were asked to report the spatial relations between the two objects. When only the objects were in the scene, participants responded from their own viewpoint. However, when the scene included an actor in the position to act on the objects, participants frequently took the actor's perspective. The first study showed that the more the actor was perceived as potentially interacting with the objects, the stronger the tendency to take his perspective. The second study investigated how manipulations of gaze affect the tendency to adopt the perspective of another reaching for an object and found that perspective-taking increased when gaze and reaching information was incongruous making the agent's behavioral intention ambiguous. These findings add further support to the idea that spontaneous perspective taking is in the service of action understanding. When the action is more difficult to understand,

there is more perspective taking. It is as if observers are putting themselves in the place of the actor to understand what he is intending to do.

But why would someone spontaneously take the spatial perspective of another when the other appears to be engaged in action? Interacting with others, understanding what they are doing and what they are likely to do next all require some comprehension of what the world looks like to them. As suggested previously (Tversky and Hard, 2009), taking the perspective of the other may be effective for planning a response to others' actions, but also for learning by observation. What makes the current results surprising is that the action was mundane—so no need to learn by observation—and required no complementary action in response. Even more surprising is that the perspective was expressed in language, in the especially confusable terms, “left”

and “right,” which are well-known to take more time to produce and to produce more errors than other directional terms like “front” and “back.” Despite this, when the agent's intention was ambiguous, the majority of participants spontaneously adopted the agent's perspective rather than their own.

ACKNOWLEDGMENTS

This work was supported by a grant from the Regione Piemonte, bando Scienze Umane e Sociali 2008, L.R. n.4/2006 to Cristina Becchio.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/Human_Neuroscience/10.3389/fnhum.2013.00455/abstract

REFERENCES

- Allison, T., Puce, A., and McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends Cogn. Sci.* 4, 267–278. doi: 10.1016/S1364-6613(00)01501-1
- Becchio, C., Bertone, C., and Castiello, U. (2008). How the gaze of others influences object processing. *Trends Cogn. Sci.* 12, 254–258. doi: 10.1016/j.tics.2008.04.005
- Becchio, C., Manera, V., Sartori, L., Cavallo, A., and Castiello, U. (2012). Grasping intentions: from thought experiments to empirical evidence. *Front. Hum. Neurosci.* 6:117. doi: 10.3389/fnhum.2012.00117
- Blake, R., and Shiffrar, M. (2007). Perception of human motion. *Annu. Rev. Psychol.* 58, 47–73. doi: 10.1146/annurev.psych.57.102904.190152
- Frischen, A., Loach, D., and Tipper, S. (2009). Seeing the world through another person's eyes: simulating selective attention via action observation. *Cognition* 111, 211–218. doi: 10.1016/j.cognition.2009.02.003
- Grafton, T. (2009). Embodied cognition and the simulation of action to understand others. *Ann. N.Y. Acad. Sci.* 1156, 97–117. doi: 10.1111/j.1749-6632.2009.04425.x
- Hudson, M., and Jellema, T. (2011). Resolving ambiguous behavioural intentions by means of involuntary prioritisation of gaze processing. *Emotion* 11, 681–686. doi: 10.1037/a0023264
- Husserl, E. (1952/1989). *Ideen zu einer Reinen Phänomenologie und Phänomenologischer Philosophie. Zweites Buch: Phänomenologische Untersuchungen zur Konstitution*. The Hague: Martin Nijhoff.
- Innocenti, A., De Stefani, E., Bernardi, N. F., Campione, G. C., and Gentilucci, M. (2012). Gaze direction and request gesture in social interactions. *PLoS ONE* 7:e36390. doi: 10.1371/journal.pone.0036390
- Johansson, R. S., Westling, G., Bäckström, A., and Flanagan, J. R. (2001). Eye-hand coordination in object manipulation. *J. Neurosci.* 21, 6917–6932.
- Kourtis, D., Sebanz, N., and Knoblich, G. (2010). Favouritism in the motor system: social interaction modulates action simulation. *Biol. Lett.* 23, 758–761. doi: 10.1098/rsbl.2010.0478
- Mazzarella, E., Hamilton, A., Troiano, L., Mastroianni, B., and Conson, M. (2012). Observation of other's action but not eye gaze triggers allocentric visual perspective. *Q. J. Exp. Psychol.* 65, 2447–2460. doi: 10.1080/17470218.2012.697905
- Mennie, N., Hayhoe, M., and Sullivan, B. (2007). Look-ahead fixations: anticipatory eye movements in natural tasks. *Exp. Brain Res.* 179, 427–442. doi: 10.1007/s00221-006-0804-0
- Pierno, A. C., Becchio, C., Turella, L., Tubaldi, F., and Castiello, U. (2008). Observing social interactions: the effect of gaze. *Soc. Neurosci.* 3, 51–59. doi: 10.1080/17470910701563269
- Pierno, A. C., Becchio, C., Wall, M. B., Smith, A. T., Turella, L., and Castiello, U. (2006). When gaze turns into grasp. *J. Cogn. Neurosci.* 18, 2130–2137. doi: 10.1162/jocn.2006.18.12.2130
- Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., and Bodley Scott, S. E. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people see. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 1255–1266. doi: 10.1037/a0018729
- Sartori, L., Becchio, C., Bulgheroni, M., and Castiello, U. (2009). Modulation of the action control system by social intention: unexpected social requests override preplanned action. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 1490–1500. doi: 10.1037/a0015777
- Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., et al. (2013). Toward a second-person neuroscience. *Behav. Brain Sci.* 36, 441–462. doi: 10.1017/S0140525X12000660
- Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., and Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature* 439, 466–469.
- Thirioux, B., Mercier, M. R., Jorland, G., Berthoz, A., and Blanke, O. (2010). Mental imagery of self-location during spontaneous and active self-other interactions: an electrical neuroimaging study. *J. Neurosci.* 30, 7202–7214. doi: 10.1523/JNEUROSCI.3403-09.2010
- Tversky, B., and Hard, B. M. (2009). Embodied and disembodied cognition: spatial perspective-taking. *Cognition* 110, 124–129. doi: 10.1016/j.cognition.2008.10.008
- Zwikel, J. (2009). Agency attribution and visuospatial perspective taking. *Psychon. Bull. Rev.* 16, 1089–1093. doi: 10.3758/PBR.16.6.1089
- Zwikel, J., and Müller, H. J. (2010). Observing fearful faces leads to visuospatial perspective taking. *Cognition* 117, 101–105. doi: 10.1016/j.cognition.2010.07.004

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 17 May 2013; accepted: 22 July 2013; published online: 12 August 2013.
Citation: Furlanetto T, Cavallo A, Manera V, Tversky B and Becchio C (2013) Through your eyes: incongruence of gaze and action increases spontaneous perspective taking. *Front. Hum. Neurosci.* 7:455. doi: 10.3389/fnhum.2013.00455

Copyright © 2013 Furlanetto, Cavallo, Manera, Tversky and Becchio. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Psychological influences on distance estimation in a virtual reality environment

Kohske Takahashi^{1*†}, Tobias Meilinger^{1,2†}, Katsumi Watanabe¹ and Heinrich H. Bülthoff^{2,3*}

¹ Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan

² Department of Human Perception, Cognition and Action, Max Planck Institute for Biological Cybernetics, Tübingen, Germany

³ Department of Brain and Cognitive Engineering, Korea University, Seoul, Korea

Edited by:

Sarah H. Creem-Regehr, University of Utah, USA

Reviewed by:

Sarah H. Creem-Regehr, University of Utah, USA

Matthew R. Longo, Birkbeck, University of London, UK

*Correspondence:

Kohske Takahashi, Research Center for Advanced Science and Technology, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan
e-mail: ktakahashi@fennel.rcast.u-tokyo.ac.jp;

Heinrich H. Bülthoff, Max Planck Institute for Biological Cybernetics, Spemannstrasse 38, 72076 Tübingen, Germany
e-mail: heinrich.buelthoff@tuebingen.mpg.de

[†]These authors have contributed equally to this work.

Studies of embodied perception have revealed that social, psychological, and physiological factors influence space perception. While many of these influences were observed with real or highly realistic stimuli, the present work showed that even the orientation of abstract geometric objects in a non-realistic virtual environment could influence distance perception. Observers wore a head mounted display and watched virtual cones moving within an invisible cube for 5 s with their head movement recorded. Subsequently, the observers estimated the distance to the cones or evaluated their friendliness. The cones either faced the observer, a target behind the cones, or were oriented randomly. The average viewing distance to the cones varied between 1.2 and 2.0 m. At a viewing distance of 1.6 m, the observers perceived the cones facing them as closer than the cones facing a target in the opposite direction, or those oriented randomly. Furthermore, irrespective of the viewing distance, observers moved their head away from the cones more strongly and evaluated the cones as less friendly when the cones faced the observers. Similar distance estimation results were obtained with a 3-dimensional projection onto a large screen, although the effective viewing distances were farther away. These results suggest that factors other than physical distance influenced distance perception even with non-realistic geometric objects in a virtual environment. Furthermore, the distance perception modulation was accompanied by changes in subjective impression and avoidance movement. We propose that cones facing an observer are perceived as socially discomforting or threatening, and potentially violate an observer's personal space, which might influence the perceived distance of cones.

Keywords: distance perception, spatial perception, virtual reality environment, personal space, object geometry

INTRODUCTION

Perceived space is not necessarily veridical, as demonstrated by many optical illusions (e.g., Müller-Lyer illusion, Ponzo illusion, and the Ebbinghaus illusion). Apart from illusions, spatial perception is susceptible to the influences of observer's psychological and physiological states. Hills appear steeper after a 1-h run (Proffitt et al., 1995; Proffitt, 2006), and a glass of water looks larger when observers feel thirsty (Velthkamp et al., 2008). These studies support the notion of embodied perception, according to which observers' mental and bodily states modify spatial perception (Proffitt, 2006).

Of our particular interest is distance perception. Distance modulates, explicitly and implicitly, the way we behave in the real world (e.g., personal space, Liberman et al., 2007). Recently, many studies have examined how factors other than physical distance influence distance perception. For example, desired objects are felt as nearer or are seen as closer (Balcetis and Dunning, 2010; Alter and Balcetis, 2011). Wearing a backpack or throwing a heavy ball results in larger subsequent distance estimations compared with wearing no backpack or throwing a light ball (Proffitt et al., 2003; Witt et al., 2004). Threatening objects (e.g., a living

tarantula) are perceived as closer (Cole et al., 2012). A location related to a rival group (e.g., Fenway Park for a Yankees fan) is imagined as nearer when accompanied by a feeling of threat (Xiao and Van Bavel, 2012). These studies imply that distance perception reflects more than physical distance, namely, social, psychological, and physiological aspects.

Thus far, the influence of the social, psychological, or physiological factors on distance perception have been tested primarily in real world situations with semantically meaningful stimuli, in line with the notion of embodied perception (Proffitt, 2006). These situations evoke associations between the presented stimuli and expected reward or punishment (e.g., a tarantula might hurt us at a closer distance). It would be plausible to argue that the expectations of reward or punishment (i.e., prospect and threat) influence distance perception by modulating psychological states. In the present study, we simplified the situation so that visual stimuli no longer afforded realistic rewards or punishments and observers were aware that the affective values associated with the visual stimuli, if any, were not real. For this purpose, we investigated the modulation of distance perception using a virtual environment and meaningless geometric objects. A virtual

environment is an experimental tool used increasingly in a wide range of contexts from navigation behavior (e.g., Frankenstein et al., 2012) to social phenomena (e.g., personal space, Bailenson et al., 2003). In virtual environments, objects are typically not real, which enables us to examine situations where observers know that the objects are *not* associated with realistic rewards or punishments (e.g., a tarantula in a virtual environment will not hurt us even at the closest distance). We presented cone-shaped objects and manipulated the orientation of the cones. The tips of the cones faced an observer, faced another location in a virtual environment, or were oriented randomly. We expected the psychological reaction to vary depending on cone orientation. In particular, the cone tips that faced the observer might induce threat or unfriendliness as in the real world situations, wherein some people develop aichmophobia, an excessive fear of sharp or pointy objects such as needles (Morse and Cohen, 1983; Shabani and Fisher, 2006).

EXPERIMENT 1

MATERIALS AND METHODS

Observers

Fourteen paid volunteers (3 females, age 19–53 years) participated in the experiment after giving written informed consent. The experimental setup was approved by the local ethics committee.

Apparatus

During the experiment, the observers stood behind a horizontal bar and grabbed a gamepad that was attached to the bar. We controlled stimulus generation and data acquisition using MATLAB with the Psychtoolbox extension (Brainard, 1997; Pelli, 1997). Visual stimuli were presented through a stereoscopic head mounted display (Kaiser SR80) with a field of view of 63° (horizontal) \times 53° (vertical), a resolution of 1280×1024 pixels for each eye, with 100% overlap, and a 60 Hz refresh rate. We fixed the inter-pupil distance for the stereo projection at 6 cm for all observers. The observers' head movements (i.e., translation and rotation) were monitored by four high-speed motion capture cameras (Vicon® MX 13) with a 120 Hz sampling rate; they were used for online stereo projection and offline head movement analysis. The stereo presentation setup allowed the observers to feel immersed in the virtual environment.

Stimuli

The visual stimulus consisted of 50 cone-shaped 3-dimensional (3-D) objects (Figure 1). The cones were of 7 cm radius and 30 cm height and moved inside an imaginary cube ($200 \times 200 \times 200$ cm) located at the observer's eye height. The mean viewing distance, from the observer to the center of the imaginary cube, was 120 cm (the cone distances ranged from 20 to 220 cm), 160 cm (the cone distances ranged from 60 to 260 cm), or 200 cm (the cone distances ranged from 100 to 300 cm). The cones moved at 75 cm/s in a direction randomly determined for each cone; each cone's direction changed every 333 ms. When the center of the cone's mass (i.e., three quarters of the middle line down from the vertex) reached a cube wall, it was reflected from the wall. The cones had a direction (i.e., based on the orientation of their tips, Figure 1B). In the ME condition, the cones were

pointing toward the observer's chest. In the TAR condition, the cones were pointing toward an invisible target placed 340 cm away from the observer. In the RND condition, the cones were pointing in pseudo-random directions. The cone directions in the RND condition were determined based on an algorithm similar to that used by Gao et al. (2010). The tip of a cone directed off a virtual line from the center of the cone to the center of the imaginary cube by a specific degree ranging from -90° to 90° . The deviation amount was fixed for each cone and was randomly determined. This made the motion profile of each cone as similar as possible for the different cone direction conditions.

Procedure

A button press started the trial. After viewing a blank screen for 0.5 s, the observer viewed the moving cones for 5 s. After the visual stimuli disappeared, a probe circle appeared at a random distance (110–210 cm away) from the observer. The observer indicated the center of the imaginary cube by moving the probe along an invisible line, which was extended horizontally through the middle of the imaginary cube, and pressing a button (Figure 1A).

We used a 3×3 within-subjects design. The factors were viewing distance (3 levels: 120, 160, and 200 cm), and cone direction (3 levels: ME, TAR, and RND). Each condition was repeated six times resulting in 54 trials, which were presented in a random order. Before the experiment, the observers were allowed to practice as long as they wanted.

RESULTS

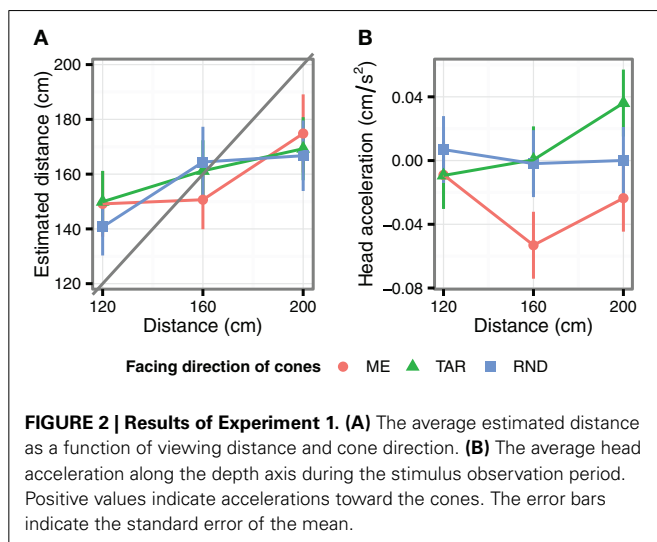
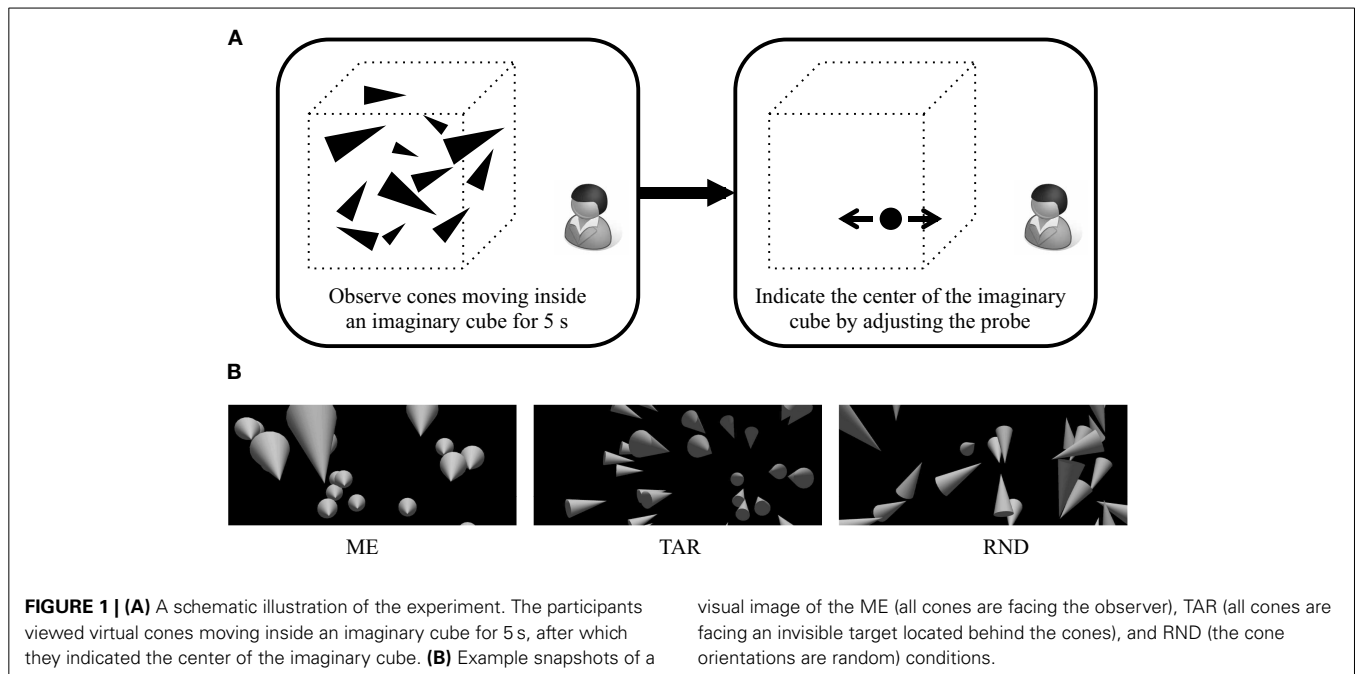
After removing values that deviated greater than 3 standard deviations from the overall mean computed for all observations (i.e., the outlier observations), the data were submitted to a mixed model analysis with the within-subjects factors of viewing distance and cone direction. We also reported partial eta squared (η_p^2) values derived from the data aggregated for each observer and for each of the conditions.

Distance estimations

Figure 2A shows the distance estimation results. As expected, the estimated distances differed depending on the viewing distance [$F_{(2, 708)} = 40.0$, $p < 0.001$, $\eta_p^2 = 0.67$]. Thus, the observers could distinguish between the different viewing distances. Note the slopes were shallower than the veridical estimation¹. That is, the observers underestimated the distance at 200 cm [$F_{(1, 13)} = 4.99$, $p = 0.044$, $\eta_p^2 = 0.28$], and overestimated the distance at 120 cm [$F_{(1, 13)} = 5.95$, $p = 0.030$, $\eta_p^2 = 0.31$]. No deviation from the actual distance was found in the 160 cm condition ($F < 1$).

The cone directions did not significantly bias the distance estimations [$F_{(2, 708)} = 0.61$, $p = 0.544$, $\eta_p^2 = 0.07$]. However, we found a significant interaction between viewing distance and cone direction [$F_{(4, 708)} = 3.38$, $p = 0.009$, $\eta_p^2 = 0.26$]. At the

¹Distance perception in virtual environments is known to be distorted relative to real world distance perception (Loomis and Knapp, 2003; Thompson et al., 2004). Therefore, we were not surprised to observe effects in this direction. However, this was beyond the focus of the present study and we will not speculate further about the potential reasons for these effects.



160 cm viewing distance, the estimated distance was significantly shorter in the ME condition compared to the TAR condition [$F_{(1, 151)} = 4.37, p = 0.038, \eta_p^2 = 0.28$] and the RND condition [$F_{(1, 149)} = 6.72, p = 0.010, \eta_p^2 = 0.23$]. On the other hand, at 120 cm, the estimated distances in the RND condition were significantly shorter than in the TAR condition [$F_{(1, 150)} = 6, p = 0.016, \eta_p^2 = 0.63$] and tended to be shorter than in the ME condition [$F_{(1, 149)} = 3.81, p = 0.053, \eta_p^2 = 0.22$]². At the viewing distance of 200 cm, we did not observe an effect on cone direction ($F < 1$).

²It is unclear why the cones of random directions were perceived as closer at the short viewing distance. At the short distance, the cones of random directions might have appeared as more crowded and less organized. Consequently, they might have “overwhelmed” the observers to a greater extent than the

Head movements

We examined head movements along the depth axis. **Figure 2B** shows the observers’ head accelerations in each of the conditions, averaged over 5 s. The head acceleration differed significantly between the cone direction conditions [$F_{(2, 700.2)} = 3.78, p = 0.023, \eta_p^2 = 0.28$], irrespective of viewing distance (i.e., no interaction between viewing distance and cone direction, [$F_{(4, 700.4)} = 1.14, p = 0.336, \eta_p^2 = 0.10$]). In the ME condition, the observers accelerated their heads more strongly away from the cones than in the TAR condition [$F_{(1, 462.4)} = 6.43, p = 0.012, \eta_p^2 = 0.41$], or in the RND condition [$F_{(1, 463.2)} = 4.19, p = 0.041, \eta_p^2 = 0.33$]. For the velocity of the head movements, the effect of viewing distance, cone direction, and the interaction were not statistically significant ($F_s < 1.32, p_s > 0.26$).

DISCUSSION

The results of Experiment 1 suggested that the cone direction modulated the distance estimations; this modulation effect depended on the viewing distance. The cones facing toward the observers were perceived as closer when they appeared in the viewing distance of 160 cm. Furthermore, the observers moved away from the cones more strongly when the cones faced them, which implies avoidance behavior. A post-experiment questionnaire also suggested that the observers experienced the cones facing them as more negative (less friendly or more threatening) than those facing the other directions. Thus, the modulation of the distance estimations might be related to the observers’ negative impressions of the cones. In Experiment 2, therefore,

cones of the more ordered conditions, and were thus perceived as closer. Because our primary interest was the effect of the cones facing toward the observers, we did not provide an in-depth discussion of the comparisons between the RND vs. the ME and the TAR conditions.

we directly tested whether cone direction affected the subjective impression of the cones.

EXPERIMENT 2

MATERIALS AND METHODS

Nine paid volunteers (2 females, age 19–24 years) participated after giving a written informed consent. The material and methods were identical to those of Experiment 1 except for the following. The observers sat on a chair and wore a different head mount display (Sony HMZ-T2). Their head movement was not monitored. After the 5-s stimulus presentation, the observers rated the “friendliness” of the cones on a 7-point scale with the poles labeled as “hostile” and “friendly” by mouse clicking.

RESULTS

Figure 3 shows the results of Experiment 2. The cone direction affected the friendliness ratings of the cones [$F_{(2, 469)} = 88.0$, $p = 0.001$, $\eta_p^2 = 0.62$]. The cones were rated as less friendly (or more hostile) in the ME condition than in the TAR condition [$F_{(1, 310)} = 49.7$, $p < 0.001$, $\eta_p^2 = 0.49$], which were in turn rated as less friendly than in the RND condition [$F_{(1, 310)} = 45.2$, $p < 0.001$, $\eta_p^2 = 0.58$]. In addition, viewing distance significantly modulated the friendliness ratings [$F_{(2, 496)} = 4.49$, $p = 0.012$, $\eta_p^2 = 0.28$], which increased linearly with the viewing distance [$F_{(1, 472)} = 9.02$, $p = 0.003$, $\eta_p^2 = 0.31$]. The interaction between viewing distance and cone direction was not statistically significant ($F < 1$).

DISCUSSION

The cones were rated as friendlier when they were facing the target rather than the observer, and when they were further away. The effect of distance estimations depended on viewing distance. In contrast, the effect of cone direction on the friendliness ratings did not depend on viewing distance. Although the decreased perceived friendliness of the cones that faced the observer is one factor that affected the distance estimations, other factors affected the pattern of distance estimates.

EXPERIMENT 3

First, we tested whether the observed effects were specific to a virtual reality setup using a head mounted display which allowed head movements. To do so, we used a projection screen with shutter glasses to emulate 3-D vision with the head fixed. Second, we wanted to clearly identify the distance at which the effect was observed. Therefore, we concentrated on the distance difference between the ME and TAR conditions and examined the effect of cone direction at seven different viewing distances. Last, we wanted to determine the within-participant relationship between the distance estimates and the friendliness ratings.

MATERIALS AND METHODS

Nineteen paid volunteers (12 females, age 19–27 years) participated after giving a written informed consent. The visual stimuli were presented on a large screen by a 3D stereo projector (Sight 3D, Solidray Co. Ltd.) and stereo shutter glasses (3D Vision, NVIDIA). The refresh rate of the projection was 120 Hz (i.e., 60 Hz for each eye). The screen was 133.5 cm (height) \times 178 cm

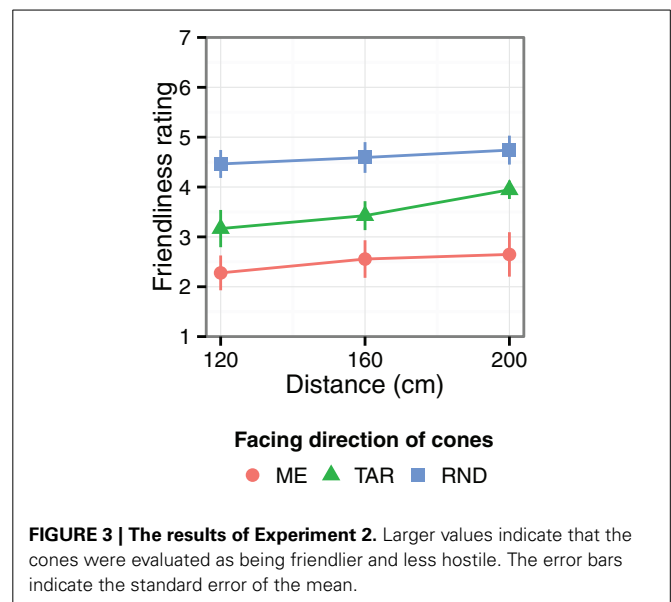


FIGURE 3 | The results of Experiment 2. Larger values indicate that the cones were evaluated as being friendlier and less hostile. The error bars indicate the standard error of the mean.

(width). The height of the center of the screen above the ground was 144 cm. The observers sat on a chair in front of the screen with their head fixed on a chin rest. The distance from the screen to the observer’s eye position was 213.5 cm. The visual stimuli were identical to those used in Experiment 1 except for the following. Since the visual angle (field of view) was smaller than that in Experiment 1, the size of the imaginary cube was 100 \times 100 \times 100 cm and the number of cones was 25. In the ME and TAR conditions, we used 7 different distances from 100 to 220 cm at 20 cm intervals. We omitted the RND condition. The observers first engaged in the distance estimation procedure as in Experiment 1. Next, they rated the cones’ friendliness as in Experiment 2. For the distance estimations, each stimulus was repeated six times. Hence, there were 84 trials. For the subjective ratings, each stimulus was presented once.

RESULTS

Distance estimations

The results are shown in **Figure 4A**. We found statistically significant effects for cone direction [$F_{(1, 1532)} = 16.2$, $p < 0.001$, $\eta_p^2 = 0.33$], and for viewing distance [$F_{(6, 1532)} = 2.7$, $p < 0.001$, $\eta_p^2 = 0.83$]. There was a significant interaction effect found between cone direction and viewing distance [$F_{(6, 1532)} = 2.41$, $p = 0.025$, $\eta_p^2 = 0.15$]. The distance estimations were significantly closer in the ME condition compared to the TAR condition only at the 200 cm and 220 cm viewing distances ($F_s > 15.0$, $p_s < 0.001$). No differences were found at the other viewing distances ($F_s < 1.95$, $p_s > 0.165$). Notably, the distance estimations were more veridical in Experiment 3 than in Experiment 1. We found no significant differences between the presented distances and the estimated distances at any of the viewing distances ($F_s < 1.79$, $p_s > 0.198$).

Friendliness ratings

Figure 4B shows the average score of the friendliness ratings. The results were generally consistent with those of Experiment 2. The

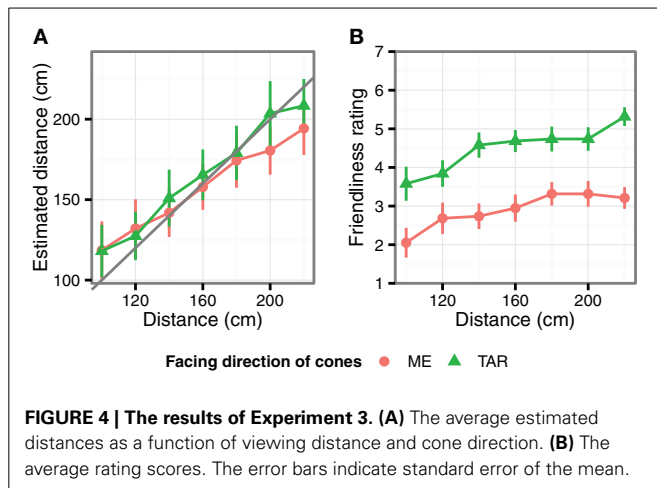


FIGURE 4 | The results of Experiment 3. (A) The average estimated distances as a function of viewing distance and cone direction. **(B)** The average rating scores. The error bars indicate standard error of the mean.

cones in the ME condition were rated as less friendly than the cones in the TAR condition [$F_{(1, 234)} = 84.9$, $p < 0.001$, $\eta_p^2 = 0.51$]. The effect of distance was also significant [$F_{(1, 234)} = 4.78$, $p < 0.001$, $\eta_p^2 = 0.24$]. The friendliness ratings almost monotonically increased with the viewing distance. The viewing distance by cone direction interaction was not statistically significant ($F < 1$). Furthermore, the overall within-participant correlation between the distance estimations and the friendliness ratings was statistically significant ($r = 0.13$, $p = 0.036$). The positive correlation observed supports the relation between these two measures.

DISCUSSION

The results of Experiments 1 and 2 were at least partially replicated in Experiment 3, which used a different virtual reality presentation method. Consistent with Experiment 2, the cones facing the observer were rated as less friendly, irrespective of the viewing distance. Furthermore, the cone direction modulated the distance estimates; similar to the results of Experiment 1, this modulation depended on the viewing distance. We compared the results of Experiments 1 and 3. In both experiments, the cones facing toward the observers were perceived as closer. This effect depended on the viewing distances, and only the effective viewing distance was different between Experiments 1 and 3. This point is further addressed in the General Discussion section.

GENERAL DISCUSSION

The present study examined whether factors other than physical distance—the orientation of simple geometric objects—influenced distance perception in a virtual environment. In Experiment 1, the observers estimated the virtual cones facing them as closer than the cones facing other directions, when the cones were presented at a certain distance (i.e., 60–260 cm away). Furthermore, the cone direction affected the observers' head movements (Experiment 1) and the subjective impression of the friendliness of the cones (Experiment 2). When the tips of the cones faced the observers, the observers moved away from the cones and rated the cones as less friendly (more hostile). These effects were observed irrespective of the viewing distance. Experiment 3 replicated the distance estimation and friendliness

rating results, although the effective viewing distance was further away (greater than 150–270 cm).

The effect on distance perception could not be a direct effect of geometric factors. As a cone is a 3-D object, the position of the cone is somewhat ambiguous when referred to by a point-shaped probe. For example, if the tips of a cone served as a representative point, the cones facing toward the observers would be estimated as closer than the cones facing the opposite direction, when the center of the cones was located at the same position, as was the case in the present experiment. If these geometric factors played a role in the modulation of distance perception, then cone direction would have influenced distance perception irrespective of viewing distance. We found, however, that the modulation of distance perception due to the cone direction depended on the viewing distance, which is not consistent with an account based on a direct effect of geometric factors. Rather, the geometric factor—whether cones faced toward the observers or not—would affect the distance perception through mediating psychological factors, such as experienced unfriendliness or perception of a threat.

In Experiments 1 and 3, the observers perceived the cones facing toward them as closer, when the cones were presented at certain viewing distances. The effective viewing distances were, however, not the same; they were from 60 to 250 cm and 150 to 270 cm in Experiments 1 and 3, respectively. The visual stimuli differences—the size of the imaginary cube and the number of cones—might have caused the difference in the effective viewing distances. However, at the moment, we speculate that differences in the devices used for the 3-D stereo presentation were responsible for the dependency on the different viewing distance. Distance in virtual environments is not necessarily veridical, but sometimes distorted compared with real spaces (Loomis and Knapp, 2003; Thompson et al., 2004). The amount of distortion depends on the setup used. The lower slope of the estimated distance against the presented distance in Experiment 1 (Figure 2A) suggests that the presented distances might be mapped onto a smaller subjective range. This was less of a factor in Experiment 3, in which the presented and estimated distances matched more closely (Figure 4A). Consequently, the subjective ranges within which cone orientation influenced distance perception might have been even more similar than suggested by the presented distances.

Several psychological factors are known to influence distance perception. Many studies have examined such influences in real world situations with meaningful stimuli (Proffitt et al., 2003; Witt et al., 2004; Balcetis and Dunning, 2010; Alter and Balcetis, 2011; Cole et al., 2012; Xiao and Van Bavel, 2012). In contrast, the aim of the present study was to examine distance perception in a virtual environment with simple visual stimuli. The virtual cones could not physically hurt the observers (and the observers knew this); nevertheless, the observers perceived the virtual cones facing toward them as closer when they were presented at a specific viewing distance. Moreover, distance perception modulation was accompanied by observers' avoidance behavior and negative subjective impression of the cones (i.e., they were rated as less friendly or more hostile). Thus, distance perception modulation was observed even when the observers were aware that the reward or punishment was virtual.

The cones facing the observers at a specific viewing distance were perceived as closer. Perhaps the modulation of distance perception was mediated by the perception of an emotional threat and/or social discomfort. One possibility is related to the fact that pointy objects tend to evoke aversion (Morse and Cohen, 1983; Shabani and Fisher, 2006); the aversion evoked might have been stronger when the cone tips faced the observers. This fits with the avoidance behavior indicated by the head tracking data as well as the less friendly ratings. The cones might also trigger social processing related to the regulation of interpersonal distance. The observers' backwards movements when faced by the cone tips could be also considered as the signature of implicit avoidance behavior. This is consistent with the finding that observers in a virtual environment keep the larger distance with a virtual avatar facing toward them (Bailenson et al., 2003). Previous studies suggested that socially or emotionally negative targets in the real world were felt or perceived as closer (Cole et al., 2012; Xiao and Van Bavel, 2012). Another possibility to explain the effect of the cones' direction on the distance estimations is that the cones facing toward the observers might be perceived as potentially approaching them. Many studies suggested that approaching objects lead to specific (negative in most cases) perceptual and social states (Mühlberger et al., 2008; Tajadura-Jiménez et al., 2010).

Although the cones facing toward the observers were rated as less friendly regardless of the viewing distance, they were perceived as being closer at specific viewing distances. Therefore, even if perceived friendliness was related to the modulation of distance estimation, it would not be a direct cause of the modulation. The distance perception modulation might be related to the violation of personal space (Liberman et al., 2007). Wilcox et al. (2006) showed that objects in a virtual environment were felt as intrusive, when the viewing distance was less than 100 cm. The modulation of distance perception by the cones' direction might take place only when they appear near the intrusiveness boundary (i.e., the personal space boundary). Objects much closer than the boundary would be perceived as violating personal space, irrespective of their perceived friendliness, while objects far away from the boundary would be perceived as not violating personal space. On the other hand, when the objects were close to the boundary, the perception of them as intrusive and violating personal space might depend on their friendliness. At this specific viewing

distance, the cones facing toward the observers that resulted in the negative reactions (i.e., avoidance behavior and less friendly ratings) might be felt as intrusive and violating; if they were facing another direction, they might be perceived as non-intrusive. A recent study demonstrated that the representation of personal space is sensitive to social factors (Teneggi et al., 2013). According to this view, the intrusive cones would be perceived as closer since they violated the observers' personal space (Schnall, 2011). Although these accounts are speculative, they warrant further investigation by combining a virtual environment, personal space, and distance perception.

In sum, the present study showed that the orientation of simple geometric objects in a virtual environment could influence their perceived distance from observers, their perceived friendliness, and implicit avoidance behavior. Several issues concerning distance perception in virtual environments remain open. For example, a direct comparison of distance perception using the same stimuli in a real and a virtual environment would help us understand how the disconnection from the real world (real rewards and punishments) affect our distance perception. The results of the present study suggest that the perception of an object as closer and having a negative impression about that object co-occur. In contrast, in the real world, we form positive and negative impressions of objects that are perceived as closer. For instance, desired objects are perceived as closer (Balcetis and Dunning, 2010). How virtual rewards affect distance perception, warrants further investigation. Our study also implies that distance estimation may serve as an objective measure for the strength of psychological reactions in the social domain in virtual environments. There has been an increase in the combination of social communication with virtual environments. Examining distance perception in virtual environments with an emphasis on psychological and social aspects will lead to the development and application of user-friendly technologies.

ACKNOWLEDGMENTS

This research was supported by JSPS KAKENHI (23240034, 25700013, 12F02779, 09J06479), JST CREST, and the National Research Foundation of Korea's World Class University program (Grant R31-10008). We would like to thank Nadine Simon for help in data collection.

REFERENCES

- Alter, A. L., and Balcetis, E. (2011). Fondness makes the distance grow shorter: desired locations seem closer because they seem more vivid. *J. Exp. Soc. Psychol.* 47, 16–21. doi: 10.1016/j.jesp.2010.07.018
- Bailenson, J. N., Blascovich, J., Beall, A. C., and Loomis, J. M. (2003). Interpersonal distance in immersive virtual environments. *Pers. Soc. Psychol. Bull.* 29, 819–833. doi: 10.1177/0146167203029007002
- Balcetis, E., and Dunning, D. (2010). Wishful seeing: more desired objects are seen as closer. *Psychol. Sci.* 21, 147–152. doi: 10.1177/0956797609356283
- Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436. doi: 10.1163/156856897X00357
- Cole, S., Balcetis, E., and Dunning, D. (2012). Affective signals of threat increase perceived proximity. *Psychol. Sci.* 24, 34–40. doi: 10.1177/0956797612446953
- Frankenstein, J., Mohler, B. J., Bühlhoff, H. H., and Meilinger, T. (2012). Is the map in our head oriented north. *Psychol. Sci.* 23, 120–125. doi: 10.1177/0956797611429467
- Gao, T., McCarthy, G., and Scholl, B. J. (2010). The wolfpack effect. Perception of animacy irresistibly influences interactive behavior. *Psychol. Sci.* 21, 1845–1853. doi: 10.1177/0956797610388814
- Liberman, N., Trope, Y., and Stephan, E. (2007). "Psychological distance," in *Social Psychology: Handbook of Basic Principles*, eds A. W. Kruglanski and E. T. Higgins (New York, NY: Guilford Press), 353–383.
- Loomis, J. M., and Knapp, J. M. (2003). "Visual perception of egocentric distance in real and virtual environments," in *Virtual and Adaptive Environments*, eds L. J. Hettinger and M. W. Haas (Mahwah, NJ: Erlbaum), 21–46.
- Morse, D. R., and Cohen, B. B. (1983). Desensitization using meditation-hypnosis to control "needle" phobia in two dental patients. *Anesth. Prog.* 30, 83–85.
- Mühlberger, A., Neumann, R., Wieser, M. J., and Pauli, P. (2008). The impact of changes in spatial distance on emotional responses. *Emotion* 8, 192–198. doi: 10.1037/1528-3542.8.2.192
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics:

- transforming numbers into movies. *Spat. Vis.* 10, 437–442. doi: 10.1163/156856897X00366
- Proffitt, D. R. (2006). Embodied perception and the economy of action. *Perspect. Psychol. Sci.* 1, 110–122. doi: 10.1111/j.1745-6916.2006.00008.x
- Proffitt, D. R., Bhalla, M., Gossweiler, R., and Midgett, J. (1995). Perceiving geographical slant. *Psychon. Bull. Rev.* 2, 409–428. doi: 10.1037/0096-1523.30.5.811
- Proffitt, D. R., Stefanucci, J., Banton, T., and Epstein, W. (2003). The role of effort in perceiving distance. *Psychol. Sci.* 14, 106–112. doi: 10.1111/1467-9280.t01-1-01427
- Schnall, S. (2011). “Embodiment in affective space: social influences on the perception of spatial layout,” in *Spatial Dimensions of Social Thought*, eds A. Maas and T. Schubert (Berlin: De Gruyter), 129–152.
- Shabani, D. B., and Fisher, W. W. (2006). Stimulus fading and differential reinforcement for the treatment of needle phobia in a youth with autism. *J. Appl. Behav. Anal.* 39, 449–452. doi: 10.1901/jaba.2006.30-05
- Tajadura-Jiménez, A., Väljamäe, A., Asutay, E., and Västfjäll, D. (2010). Embodied auditory perception: the emotional impact of approaching and receding sound sources. *Emotion* 10, 216–229. doi: 10.1037/a0018422
- Teneggi, C., Canzoneri, E., di Pellegrino, G., and Serino, A. (2013). Social modulation of peripersonal space boundaries. *Curr. Biol.* 23, 406–411. doi: 10.1016/j.cub.2013.01.043
- Thompson, W. B., Willemsen, P., Gooch, A. A., Creem-Regehr, S. H., Loomis, J. M., and Beall, A. C. (2004). Does the quality of the computer graphics matter when judging distances in visually immersive environments. *Presence Teleoper. Virtual Environ.* 13, 560–571. doi: 10.1162/1054746042545292
- Veltkamp, M., Aarts, H., and Custers, R. (2008). Perception in the service of goal pursuit: motivation to attain goals enhances the perceived size of goal-instrumental objects. *Soc. Cogn.* 26, 720–736. doi: 10.1521/soco.2008.26.6.720
- Wilcox, L. M., Allison, R. S., Elfassy, S., and Grelik, C. (2006). Personal space in virtual reality. *ACM Trans. Appl. Percept.* 3, 412–428. doi: 10.1145/1190036.1190041
- Witt, J. K., Proffitt, D. R., and Epstein, W. (2004). Perceiving distance: a role of effort and intent. *Perception* 33, 577–590. doi: 10.1068/p5090
- Xiao, Y. J., and Van Bavel, J. J. (2012). See your friends close and your enemies closer: social identity and identity threat shape the representation of physical distance. *Pers. Soc. Psychol. Bull.* 38, 959–972. doi: 10.1177/0146167212442228
- that could be construed as a potential conflict of interest.

Received: 14 June 2013; accepted: 28 August 2013; published online: 18 September 2013.

Citation: Takahashi K, Meilinger T, Watanabe K and Bühlhoff HH (2013) Psychological influences on distance estimation in a virtual reality environment. *Front. Hum. Neurosci.* 7:580. doi: 10.3389/fnhum.2013.00580

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2013 Takahashi, Meilinger, Watanabe and Bühlhoff. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships



The role of potential agents in making spatial perspective taking social

Amy M. Clements-Stephens^{1,2,3}, Katarina Vasiljevic³, Alexandra J. Murray³ and Amy L. Shelton^{1,2,3*}

¹ School of Education, Johns Hopkins University, Baltimore, MD, USA

² Center for Talented Youth, Johns Hopkins University, Baltimore, MD, USA

³ Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, MD, USA

Edited by:

Antonia Hamilton, University of Nottingham, UK

Reviewed by:

Leonhard Schilbach, University

Hospital Cologne, Germany

Dana Samson, Université Catholique

de Louvain, Belgium

*Correspondence:

Amy L. Shelton, School of

Education, Johns Hopkins

University, 3400 North Charles

Street, Baltimore, MD 21218, USA

e-mail: ashelton@jhu.edu

A striking relationship between visual spatial perspective taking (VSPT) and social skills has been demonstrated for perspective-taking tasks in which the target of the imagined or inferred perspective is a potential agent, suggesting that the presence of a potential agent may create a social context for the seemingly spatial task of imagining a novel visual perspective. In a series of studies, we set out to investigate how and when a target might be viewed as sufficiently agent-like to incur a social influence on VSPT performance. By varying the perceptual and conceptual features that defined the targets as potential agents, we find that even something as simple as suggesting animacy for a simple wooden block may be sufficient. More critically, we found that experience with one potential agent influenced the performance with subsequent targets, either by inducing or eliminating the influence of social skills on VSPT performance. These carryover effects suggest that the relationship between social skills and VSPT performance is mediated by a complex relationship that includes the task, the target, and the context in which that target is perceived. These findings highlight potential problems that arise when identifying a task as belonging exclusively to a single cognitive domain and stress instead the highly interactive nature of cognitive domains and their susceptibility to cross-domain individual differences.

Keywords: perspective taking, social skills, agency, individual differences, spatial cognition

The ability to imagine the world from the point of view of another person comes in a variety of forms, from understanding another person's opinion on a discussion topic to literally imagining what the visual world would look like from their perspective. The latter, termed visual-spatial perspective taking (VSPT), has traditionally been considered a form of spatial problem solving. However, over the course of the last decade, there has been a growing body of research supporting a relationship between one's social abilities and the ease with which they are able to engage in this more visually driven form of perspective taking (Brunyé et al., 2012; Kessler and Wang, 2012; Shelton et al., 2012), highlighting the role of VSPT for everyday social interactions. Impairment on tasks that require adopting another's perspective—be it to judge if an object is visible from another viewpoint (Level 1 VSPT) or to represent what a spatial layout might look like from another viewpoint (Level 2 VSPT)—is a hallmark feature of Autism Spectrum Disorders (ASD; Baron-Cohen, 1992; Best et al., 2008). A vast majority of the research examining this relationship between social and VSPT abilities tends to come in two forms: either investigation of how/when VSPT abilities are impaired or preserved in individuals with ASD due to their known deficits in social skills (Hobson, 1984; David et al., 2006; Hamilton et al., 2009; Gould et al., 2011; Zwickel et al., 2011; Schilbach et al., 2012) or investigations of the natural variability that is observed in more typically-developing populations (Brunyé et al., 2012; Shelton et al., 2012).

One approach to understanding how social abilities might influence VSPT is through the investigation of the role that agents play in cognitive tasks. A wide range of literature on embodied cognition has investigated the conditions and tasks that appear to be sensitive to the presence of a human agent (Eppel et al., 1983; Schober, 1998; Ruby and Decety, 2001; Ames et al., 2008; David et al., 2008; Tversky and Hard, 2009; Kessler and Rutherford, 2010; Kessler and Thomson, 2010; Schilbach et al., 2012). For example, Schober (1998) has demonstrated that people will make an effort to adopt a listener's perspective when describing a spatial display, whereas they use their own perspective or neutral statements such as cardinal directions when asked to simply describe the display (no human listener). Similarly, Tversky and Hard (2009) asked individuals to describe spatial events depicted in scenes with or without the presence of another person in the scene. They found that the scene descriptions differed such that the participants spontaneously adopted the perspective of a person in a scene, even when such perspective taking was not relevant to the task. Moreover, they had no direct contact with the agent, suggesting that it was the mere presence and not any interactive requirement that motivated spontaneous perspective taking. Additionally, Schilbach et al. (2012) showed a sensitivity toward face-like stimuli that had a modulatory effect on performance when completing a gaze-mediated stimulus-response compatibility paradigm. When the social context of the stimuli was manipulated (face, face-like, or object stimulus), there was a

reduction in the observed congruency effects for faces as compared to objects, suggesting that there was an effect of social context on action control. Based on these results, it is clear that human participants are sensitive to the presence of other human agents in ways that affect performance.

An alternative line of work has explored how agency might be attributed to objects (Zwicker, 2009; Zwicker et al., 2011; Zwicker and Müller, 2013). In a series of studies, Zwicker and colleagues set out to better understand VSPT when non-human entities were used as the target of perspective taking, both in typically-developing individuals (Zwicker, 2009; Zwicker and Müller, 2013) and those with ASD (Zwicker et al., 2011). This was first accomplished by examining whether individuals would adopt the perspective of geometrical shapes if the movement of the shapes appeared intentional. Intentionality was manipulated by using movements that implied interactions between the shapes. For example, when two triangles were moving about each other, they might evoke descriptions that reflect theory of mind (ToM) such as, “The small triangle surprised the large triangle.” Zwicker (2009) presented individuals with either the systematic ToM movement or random movement and found that individuals spontaneously adopted the perspective of the probed triangle when the movement implied agency but not when the movement was random. Follow-up work on individuals with ASD showed that although they were able to understand that the triangles were interacting in one case (ToM condition) and not in the other (random condition), the additional attribution of agency did not occur as evidenced by less appropriate descriptions of the animations (Zwicker et al., 2011). Therefore, it was concluded that in the case of individuals with ASD, although the perceptual cues (type of movement) were sufficient to invoke intentionality, they did not imbue the triangles with agency. These studies highlight one possible feature, intentional movement, which may be necessary for non-human targets to be perceived as potential agents. However, it is unknown what the potential boundary conditions are associated with perceived agency and the minimal requirements needed for stimuli to evoke VSPT when the stimuli are static.

In addition to understanding how agency attribution impacts VSPT, an additional line of research has been focused more on investigating the direct relationship between social abilities and VSPT. In particular, Brunyé et al. (2012) assessed whether gender and sub-clinical autistic traits were not only predictive of VSPT, but could differentiate between the levels of VSPT. The VSPT tasks used in this study required participants to either determine whether a light was visible from the perspective of an avatar (Level 1) or whether the light was to the left/right of the avatar (Level 2) and participants completed the autism quotient (AQ; Baron-Cohen et al., 2001). The overall score on the AQ was used with higher scores being indicative of greater autistic-like traits. Results from this study found slowed reaction times for the Level 2 VSPT task in males and females with relatively high AQ scores, suggesting that individuals with more autistic-like traits had greater difficulty taking the perspective of the avatar. This relationship was not seen for the Level 1 VSPT task. Taken together, these findings suggest that even in sub-clinical healthy populations, having more autistic-like traits influences one’s ability to engage

in perspective taking when the judgment to be made goes beyond asking whether something might be visible in the alternative perspective and requires a more complex set of judgments to be made about the spatial properties of the visual scene from an alternative perspective.

In a similar manner, Shelton et al. (2012) focused more specifically on social skills by using a combined score derived from the AQ social and communication subscales. As such, a lower score on this combined AQ score would be associated with individuals who have strong social skills whereas a higher score would be more associated with individuals who are less socially savvy¹. In this experiment, participants were seated in front of a display of three buildings with seven different colored potential targets of perspective taking oriented around the display at 45° intervals. Participants were presented with an image and were asked to identify which viewpoint was being displayed, whether it be their own or one of the potential targets. Agency was manipulated by having each participant complete three different conditions: artist figures, triangles, and cameras. It was hypothesized that artist figures would be more human-like than either the triangles or cameras, with triangles as clear inanimate objects and cameras as potential intermediaries of perspective (people look through them). A striking relationship was found. Participants with lower AQ combo scores (more social) were more accurate at taking the perspective of the artist figures ($r = -0.584$) than those with higher AQ combo scores (less social), whereas no such relationship was found for either the triangles ($r = -0.084$) or the cameras ($r = -0.053$) conditions. It should be acknowledged that all of the conditions used objects, but the relative amount of potential agency conveyed varied across the different conditions. As such, these findings point to another potential requirement for perceived agency, especially with respect to static images; that is, it may be necessary for the potential “agents” to possess some human-like qualities.

Not only do these studies provide indications as to what it means for an object to be perceived as a potential agent, they also introduce a framework for distinguishing when VSPT includes a social component or not. That is, VSPT may be primarily spatial and remain so when targets do not evoke the suggestion of social engagement, but VSPT may become more dependent on interactions with social skills when targets are more agent-like, allowing one’s social skills to influence behavior for better or for worse. This offers a method for assessing what kinds of targets might make VSPT more or less social. In particular, we can use the correlation between VSPT performance and social skills as a measure of when VSPT is or is not incurring social skill influence. If a target is motivating the task to be a “socially relevant” form of VSPT, we expect a relationship between measures of social skill and VSPT performance. However, if a target is not incurring the agency necessary to motivate social relevance, we expect to

¹One important note on social skills in this context is that the relevant dimension is likely to be one’s understanding and appreciation of social attributes and situations rather than social-seeking behavior or extroversion. Throughout, we use this broader definition, suggesting a form of social intelligence or savvy rather than more strictly whether someone engages in more or less social activity.

see a “non-social” form of VSPT such that social skills are not correlated with VSPT performance. Critically, we do not predict opposite relationships for socially relevant and non-social VSPT, but rather suggest that the task can either be sensitive to social influence or not (see **Figure 1**).

Using this approach to assess social influence, we can begin to ask deeper questions about what targets, target features, and conditions change the way an individual approaches the VSPT task. First, we can test how changing basic features might influence the degree to which a target seems to acquire agency that brings social skills to bear. In the previous study (Shelton et al., 2012), we used brute force differences (human form vs. inanimate objects), but more subtle information can also be manipulated. For example, we can manipulate the presence or absence of very basic facial features. Second, we can ask whether and how experience with one type of target might modulate the perceived agency of another target, which we refer to as experiential context. In the previous study (Shelton et al., 2012), there was no effect of order, suggesting that having seen the artist figures first did not make the triangles more or less sensitive to the influence of social skills. Moreover, seeing the triangles followed by artist figures did not make the artist figures more or less sensitive. However, as noted above, these were already different classes of objects. Here we consider what happens to perceived agency when targets share certain features but not others.

To address these issues, we ran a series of experiments that compare the original targets from the previous study, plain triangles and artist figures, to other variations that might convey more or less agency as evidenced by an influence of social skills on VSPT performance. First, we set out to establish whether adding human-like features to an object would increase the sensitivity to social skill influence. In Experiment 1, we compared plain triangles to triangles with eyes affixed to them, making them appear more human-like (or at least Muppet-like) via visual features. In Experiment 2, we compared the artist figures

condition from our previous study, which showed the relationship between social skills and VSPT, to agents with even more human-like qualities, fashion dolls. In both Experiments 1 and 2, we contrasted conditions that vary on known visual features and counterbalanced order to allow us to explore any modulation of the social skill relationships due to experiential context. Lastly, we wanted to ask whether we could increase sensitivity to social skill influence by conceptually manipulating the meaning of a target of perspective taking. In Experiment 3, we return to the plain triangles, but now refer to them as “Aliens” in an effort to convey that these could be creatures with agent-like qualities. This manipulation allowed us to ask whether a conceptual cue to agency can be robust enough to bring about the influence of social skills on VSPT performance. Overall, results from these studies reveal the complexity of the relationship between social skills and VSPT, highlighting the susceptibility of individual differences to contextual and cross-domain influences.

GENERAL MATERIALS AND METHODS

All three of the correlational experiments used the same basic paradigm, varying only the target of the VSPT task.

PARTICIPANTS

All participants were Johns Hopkins University undergraduate students between the ages of 18–22 who participated in return for extra credit in psychology courses. All procedures were approved and conducted in accordance with the Johns Hopkins Homewood Institutional Review Board. For all studies, inclusion in the study was based on the 0° orientation trials (described in more detail in the subsequent section). Because this type of VSPT task is difficult and a wide range of scores is typically observed, we did not want to exclude individuals merely because they fell along the lower end of the distribution. Therefore, we reasoned that if participants could correctly identify their own view, then we could assume that they understood and were engaged in the task. As such, we excluded individuals who made more than one error on these trials. Across all of the experiments, this criterion seemed to successfully separate those who were on task from those who were not. Moreover, for each experiment, an effort was made to obtain approximately equal numbers of males and females. The analyses for each experiment included examining potential differences in performance between the genders. Consistent with previous findings obtained by Shelton et al. (2012), the differences between males and females on all measures and correlations were negligible and will not be discussed further.

MATERIALS, DESIGN, AND PROCEDURES

For each experiment, participants completed a set of measures that included the three buildings task, paper-and-pencil spatial skill tests, and a self-report questionnaire (the AQ described below) in a pseudorandom order. The spatial skill tests administered are part of a standard battery of measures typically included across all experiments in the lab; they were not pertinent to the hypothesis-driven questions being addressed and had little or no relationship to the outcomes presented below.

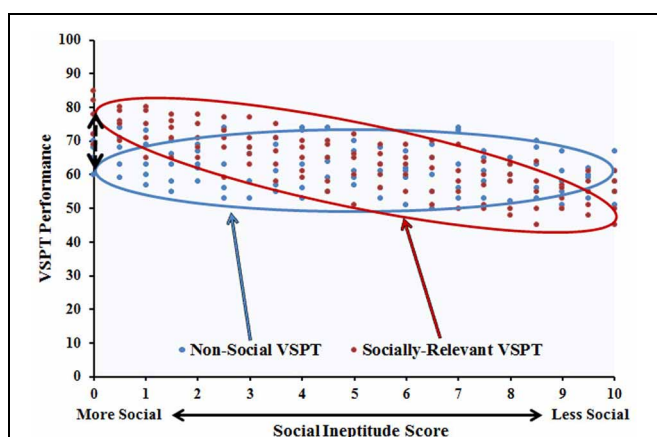


FIGURE 1 | Potential framework depicting the relationship between social skills and VSPT performance for proposed socially-relevant and non-social VSPT. The distinction between these two types of tasks is captured by the degree to which VSPT is incurring a social skill influence indexed by the magnitude of the correlation.

Three buildings (3Bldgs) task

Participants completed the 3Bldgs task, which is an adaptation of Piaget's three mountains perspective-taking test (Piaget and Inhelder, 1967). For this task, participants viewed two different displays. Each display consisted of three unique buildings (6 different buildings total) with each building constructed out of LEGO® building blocks (Lego Group, Billund, Denmark) and placed on 24" diameter plastic disks that were covered in faux grass mats. Each display disk was centered on a 36" diameter wood table and photographed from 8 different orientations separated by 45° increments. Around the building display were seven uniquely colored targets for perspective taking (red, blue, white, black, purple, yellow, and orange). Targets were placed at 45° intervals and corresponded to headings of 45°, 90°, 135°, 180°, 225°, 270°, and 315° with respect to the participants' designated view of 0°. The targets were manipulated across the set of experiments to assess the potential impact on the perceived agency (see **Figure 2**) and are described in greater detail with the corresponding experimental manipulation.

Participants were seated in front of the physical display and viewed images on a laptop computer. Each presented image corresponded to the would-be visual perspective of one of the targets or to the participant's own perspective (0°). Participants were asked to identify the perspective of the image. For each image, irrespective of the task version, the participant was asked, "Which <TARGET> is at this view?"² Participants indicated their response by pressing a key corresponding to the color of the target or the spacebar to indicate that it was his/her own view. Each task version consisted of 40 self-paced trials (5 trials at each

orientation) with a 5-s response deadline. Response latency and accuracy were measured.

The 0° orientation (where the participant was seated) was selected randomly for each participant from one of 4 possible orientations. For each display, the four candidate orientations were selected by randomly choosing one orientation and using the 3 additional orientations that were opposite and orthogonal to that initial orientation. For a given start orientation, targets were placed at the remaining seven orientations. When appropriate, display-condition assignment and order of conditions were counterbalanced, and the order of the target colors around the display was selected randomly for each participant and kept constant for both conditions (when applicable).

Autism quotient (AQ; Baron-Cohen et al., 2001)

The AQ is a self-report questionnaire designed to assess the degree to which individuals vary on five traits typically associated with ASD—social skills, perseveration, attention to detail, communication, and imagination—with higher scores (overall and each subscale separately) reflecting stronger ASD-like traits. For the set of experiments presented here, the critical scales of interest are the social and communication impairment scales, which are designed to capture behaviors on a continuum from socially appropriate to socially inappropriate behaviors. Due to the strong correlation observed in Shelton et al. (2012) between the social and communication impairment subscales, we used the same combined social/communication score as the previous study. For clarity, we term this the social ineptitude score, reflecting the fact that higher scores mean less social. This social ineptitude score is used in all analyses for each experiment.

EXPERIMENT 1: TRIANGLES WITH AND WITHOUT EYES

In this experiment, we set out to assess whether objects could be made sensitive to the influence of social skills by adding features suggestive of agency. Specifically, participants were asked to complete two conditions in which they either took the perspective of a triangle (plain triangles condition) or took the perspective of a triangle that had eyeballs affixed to the top of it (triangles-with-eyes condition). First, we expected no relationship between social skills and performance with plain triangles, replicating our previous work. For the triangles-with-eyes condition, we used the magnitude of the correlation between social skills and performance as an index to determine whether eyeballs were sufficient for inducing potential agency in an object. If any sign of agency "socializes" the task, then we might expect a correlation comparable to that observed for the artist figures in Shelton et al. (2012). However, we may also see a gradation of social skill influence dependent upon the degree of potential agency induced, in which case the triangles-with-eyes would show a weaker correlation than other more agent-like targets. Such a result would necessitate additional comparisons. Finally, we might see that static triangles are objects regardless of whether they have eyes or not, with no correlation observed in either condition.

In addition to the basic comparison of triangles with and without eyes, we also entertained the possibility that the order of the conditions could affect observed correlations. Given that these two conditions use the same basic object, they may provide a

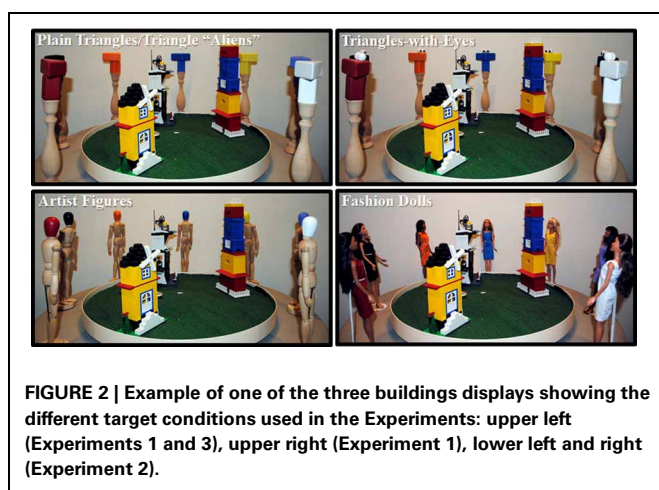


FIGURE 2 | Example of one of the three buildings displays showing the different target conditions used in the Experiments: upper left (Experiments 1 and 3), upper right (Experiment 1), lower left and right (Experiment 2).

²This question was intentionally vague to limit potential bias in the language that might influence how the participant should interpret the task and the target. Using language such as "Who sees this...?" might affect interpretation of agency (i.e., "who" denotes a being; "seeing" is a property of an agent). We also acknowledge that this wording does not explicitly tell the participant to mentally visualize the space. As such, it is possible that some participants engaged more or less visual imagery as opposed to other forms of spatial reasoning. However, we have no reason to suspect that this same variability in strategy does not hold for all VSPT tasks.

more direct contrast of their potential agency than the triangles, cameras, and artist figures used in our previous research study (Shelton et al., 2012). As such, we considered whether seeing the plain triangles first might imbue the subsequent triangles with eyes with more agency than they might convey on their own (when experienced first). This might result in the correlation for triangles-with-eyes being weaker when performed first than when performed after plain triangles. Plain triangles provide an even more interesting case in that they are expected to show no correlation, especially if experienced first. However, it is possible that seeing triangles-with-eyes as potential agents could carryover to subsequent performance with plain triangles, making them more sensitive to the social skill influence.

MATERIALS AND METHODS

Participants

For this experiment, 78 naïve participants were enrolled. Six participants (2 males) failed to meet criterion, leaving 72 participants (34 males) included in all subsequent analyses.

Materials, design, and procedures

Participants completed two versions of the 3Bldgs task using plain triangles and triangles with eyes as targets. Using a set of 14 identical wooden triangular blocks, we created two sets of seven different colored triangles (see above) placed on plain wood pedestals (13" total height). One set served as the plain triangles condition. For the second set, 1" round wooden eyeballs painted white with black circles were affixed to the top of each triangle to create the triangles-with-eyes condition (see Figure 2). For each image, irrespective of the task version (plain triangles/triangles-with-eyes), the participant was asked, "Which Triangle is at this view?" Each participant completed both versions of the task using two different displays. Display-condition assignment and order of conditions were counterbalanced. After applying the exclusion criterion, we had approximately equal numbers in each order (plain triangles first $n = 34$). The order of the target colors around the display was selected randomly for each participant and kept constant for both conditions.

RESULTS

Mean response latency (overall and for correct trials only) and overall accuracy were calculated for both versions of the 3Bldgs task and were separately subjected to a mixed ANOVA with order as a between-subjects variable and target (plain triangles/triangles-with-eyes) as a within-subjects variable (Figure 3). For response latency, there were no significant effects or interactions (all p s > 0.11). For accuracy, we found that the group that performed the plain triangles first were significantly less accurate than the group that performed the triangles-with-eyes condition first, $F_{(1, 70)} = 5.75$, $p = 0.019$, $\eta^2_G = 0.05$. Although this effect was significant, it was a small effect, accounting for only about 5% of the measured variance. Moreover, we also observed that participants were significantly more accurate on the triangles-with-eyes condition than the plain triangles condition, $F_{(1, 70)} = 5.98$, $p = 0.017$, $\eta^2_G = 0.02$. Again, this was a small effect (2% of the measured variance). There was no significant order \times target interaction, $F_{(1, 70)} = 1.97$, $p = 0.165$.

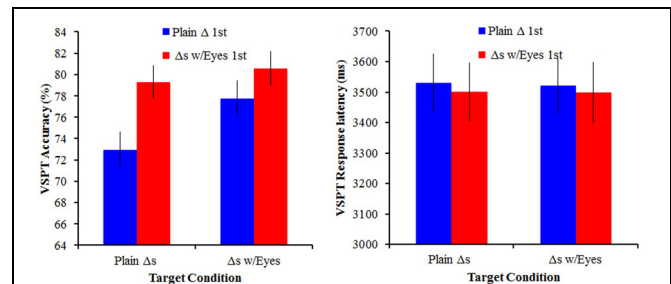


FIGURE 3 | VSPT performance (mean accuracy and response latency) as a function of the target conditions for Experiment 1. Error bars reflect ± 1 standard error of the mean.

Although these effects are fairly small, it suggests that it may be important to consider whether some targets' perspectives are more readily adopted overall, and, more critically, the role order may play when investigating the relationship between social skills and performance on our VSPT conditions.

In addition to assessing conditional differences, we also correlated the performance on the plain triangles and triangles-with-eyes conditions. Accuracy was positively correlated, $r = +0.44$, $p < 0.001$. This relationship is consistent with our previous study (Shelton et al., 2012) and suggests that there is a common spatial component to the VSPT task, irrespective of target condition.

To answer the critical question of social skill influence on VSPT performance we explored the correlations between social skills and accuracy³ on the VSPT task for the two target conditions overall and then for each order separately. All correlations are significant at $\alpha = 0.05$, corrected for the size of the relevant subset of correlations investigated unless otherwise specified. Consistent with Shelton et al. (2012), we observed a significant correlation between the social and communication impairment subscales from the AQ ($r = 0.42$); therefore, due to this observed relationship and to be consistent with previous literature we used a combined scored (social ineptitude score) in all analyses (see Table 1 for separate correlations with subscales). Lower values on the social ineptitude score would be associated with better social skills, whereas higher values on the social ineptitude score would be associated with poorer social skills. Overall, we observed no correlation for the plain triangles condition, $r = -0.18$, $p = 0.12$ and a negative correlation for the triangles-with-eyes condition, $r = -0.46$, indicating that more social individuals had better performance than less social individuals for the triangles-with-eyes condition. A t -test for non-independent r 's was conducted and revealed that these correlations were significantly different, $t_{(69)} = 2.45$, $p = 0.02$, suggesting that adding eyes to the plain triangles allowed objects to become sensitive to social skill influences in VSPT.

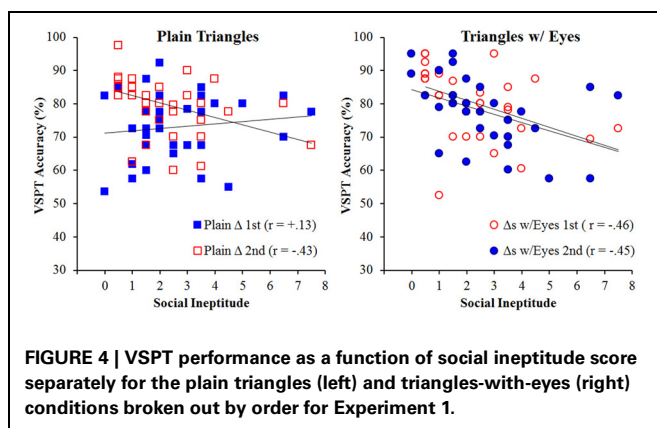
An examination of the effect of order on these correlations paint a more complex picture, as can be observed in Figure 4, which shows the correlations for both conditions separately for each order. For the triangles-with-eyes condition, the correlation

³Consistent with previous data, response latency had no significant correlations (strongest was $r = +0.1$).

Table 1 | Summary of correlations (and *p*-values) between the relevant social skill scores and each condition, overall and by order of conditions where relevant.

	Social subscale	Communication subscale	Social ineptitude score
EXPERIMENT 1			
Plain triangles			
Overall	+0.01 (0.939)	−0.33 (0.005)*	−0.18 (0.123)
1 st	+0.23 (0.202)	−0.04 (0.805)	+0.13 (0.479)
2 nd	−0.20 (0.232)	−0.50 (0.001)*	−0.43 (0.008)*
Triangles-with-eyes			
Overall	−0.26 (0.031)	−0.54 (<0.001)*	−0.46 (<0.001)*
1 st	−0.29 (0.077)	−0.46 (0.003)*	−0.46 (0.004)*
2 nd	−0.22 (0.219)	−0.60 (<0.001)*	−0.45 (0.008)*
EXPERIMENT 2			
Artist figures			
Overall	+0.01 (0.920)	−0.15 (0.195)	−0.08 (0.521)
1 st	−0.25 (0.160)	−0.45 (0.007)*	−0.39 (0.022)
2 nd	+0.20 (0.235)	+0.07 (0.678)	+0.14 (0.391)
Fashion dolls			
Overall	−0.25 (0.032)	−0.48 (<0.001)*	−0.40 (0.001)*
1 st	−0.21 (0.209)	−0.41 (0.011)	−0.33 (0.045)
2 nd	−0.32 (0.067)	−0.61 (<0.001)*	−0.52 (0.002)*
EXPERIMENT 3			
Triangle aliens	−0.27 (0.066)	−0.34 (0.017)*	−0.36 (0.013)*

All *p*-values presented are uncorrected for multiple comparisons; * indicates the correlations that survive the correction for multiple comparisons within related subsets.



was significant regardless of the order in which the conditions were completed, $r = -0.46$ and -0.45 , supporting the claim that placing eyes on the simple triangles was sufficient to bring about the social skill influence. Performance on perspective taking with plain triangles had previously not shown a significant correlation with social skills (Shelton et al., 2012), and that was again the case when this condition was performed as the first condition, $r = +0.13$, $p = 0.48$. However, when the plain triangles followed the triangles-with-eyes, there was a significant negative correlation, $r = -0.43$, that was not significantly different from those observed for the triangles-with-eyes in either order,

$ps > 0.45$. Additionally, this correlation was significantly different from the correlation when the plain triangles were viewed first as evidenced by a *z*-transform for independent correlations, $z = -2.35$, $p = 0.009$. As such, plain triangles incurred as much influence from social skills as triangles-with-eyes when participants had experienced the triangles-with-eyes first.

In cases where we observed a correlation with the social ineptitude score, we also compared their magnitude to the correlation observed for artist figures in the previous study ($r = -0.58$; Shelton et al., 2012) using a *z*-transform for independent correlations. None of the comparisons were significant, all $ps > 0.16$, suggesting that the triangles-with-eyes and the plain triangles (when presented after the triangles-with-eyes condition) were showing correlations in the same range as previously observed for artist figures.

DISCUSSION

The goal of Experiment 1 was to determine whether an object previously shown to be insensitive to social skill influence could be made sensitive by adding an agent-like feature. Adding eyes to triangles had two important impacts on performance. First, we did observe the significant correlation between performance on the triangles-with-eyes condition and social skills, suggesting that features such as eyes can create targets of perspective taking that are sensitive to the influence of social skills. In essence, it appears placing static features on an object may convey a sense of agency similar to what was observed for animating shapes (e.g., Zwickel, 2009). In addition, we observed that our plain triangles could also be imbued with some agent-like attributions by simply following the experience of the task with the triangles-with-eyes.

It is tempting to argue that the act of performing the task with an implied social context might keep participants in a state of “social-ness” rather than actually imparting the agency or sociality on the plain triangles. This seems unlikely given that we failed to find any order effects (or even trends that would suggest order effects) in the previous study (Shelton et al., 2012) and pilot work when the targets were different kinds of objects. Instead, it seems that the shared properties of the triangles with and without eyes may have allowed the agency (or sensitivity to social influence) induced by the eyes to carry over. In Experiment 2, we examine a variation on this carry over effect by contrasting two different representations of human form.

EXPERIMENT 2: ARTIST FIGURES AND FASHION DOLLS

Experiment 1 started with an object and examined whether we could induce sensitivity to social skill influence. In Experiment 2, we started with the artist figures that were first used to demonstrate the correlation between social skills and VSPT in this paradigm and compared them to a target with more human-like features to assess whether the correlation might be sensitive to the degree or extent of implied agency. In addition, we again varied the order of the conditions to examine whether experience with the putatively stronger potential agent might strengthen or weaken the sensitivity of the putatively weaker potential agent.

MATERIALS AND METHODS

Participants

For this experiment, 82 naïve participants were enrolled. Ten participants (5 males) were excluded due to failure to reach criterion, leaving 72 participants (30 males) eligible for all subsequent analyses.

Materials, design, and procedures

Participants completed two versions of the 3Bldgs task. In the artist figures condition, each target was a 13" tall wooden artist figure with its head painted one of seven unique colors. In the fashion dolls condition, we used a set of 7 distinct Barbie™ dolls (Mattel, Inc., El Segundo, CA), with each fashion doll wearing a colored dress corresponding to the colors used in the artist figures condition (see **Figure 2**). For each image, irrespective of the task version (artist figures/fashion dolls), the participant was asked, "Which Doll is at this view?" Each participant completed both versions of the task using two different displays. Display-condition assignment and order of conditions were counterbalanced. After applying the exclusion criterion, we had approximately equal numbers in each order (artist figures first $n = 38$). The order of the target colors around the display was selected randomly for each participant and kept constant for both conditions.

RESULTS

Mean response latency (overall and for correct trials only) and overall accuracy were calculated for both versions of the 3Bldgs task and were separately subjected to a mixed ANOVA with order as a between-subjects variable and target (artist figures/fashion dolls) as a within-subjects variable (see **Figure 5**). For both response latency and accuracy, there were no significant effects or interactions (all p s > 0.23 for response latency and all p s > 0.07 for accuracy). Again, we observed a significant positive correlation between accuracy with fashion dolls and accuracy with artist figures, $r = +0.46$, $p < 0.001$.

Given the observed correlation between the social and communication impairment subscales from the AQ ($r = 0.69$), all subsequent correlations were run between the social ineptitude score and the accuracy on the two agency conditions overall and then for each order separately (see **Table 1** for separate correlations with subscales). All correlations are significant at $\alpha = 0.05$,

corrected for the size of the relevant subset of correlations investigated unless otherwise specified. Overall, we observed a negative correlation for the fashion dolls, $r = -0.40$, indicating better accuracy with better social skills. Surprisingly, no such correlation was observed for the artist figures, $r = -0.08$, $p = 0.52$. This is contrary to our previous studies where we have consistently observed this correlation (Shelton et al., 2012). A t -test for non-independent r 's confirmed that these two correlations were significantly different from each other, $t_{(69)} = 2.82$, $p = 0.006$.

The unexpected result in the artist figures overall made the motivation for examining order effects even stronger. **Figure 6** shows the correlations broken down by order. For the fashion dolls, the correlation between social ineptitude score and performance was weaker when fashion dolls were presented first, $r = -0.33$, $p = 0.04$ uncorrected, but met the criterion for multiple comparisons when fashion dolls were presented second, $r = -0.52$. Although the correlation numerically increased when fashion dolls were presented second, the difference between the two correlations was not significant, $z = 0.97$, $p = 0.17$, suggesting that the correlation was similar irrespective of order. An examination of the artist figures condition revealed a more complicated picture. That is, when artist figures came first, performance showed a correlation with the social ineptitude score similar to what was observed previously (Shelton et al., 2012), $r = -0.39$, $p = 0.02$ uncorrected (comparison to $r = -0.58$, $z = 1.33$, $p = 0.09$), but when these same artist figures followed the experience with fashion dolls, the correlation with the social ineptitude score was weak and in the opposite direction, $r = +0.14$, $p = 0.39$. The correlations for the artist figures in the two different orders were significantly different, $z = 2.26$, $p = 0.001$, suggesting that the social skill sensitivity was modulated by the context in which the particular targets were experienced.

DISCUSSION

Experiment 2 provides a second example of how the degree to which a target of perspective taking appears to be socially relevant can be influenced by the experience of other targets. Artist figures, which were used to establish the initial correlation between social skills and VSPT in this paradigm, were essentially stripped of their sensitivity to social skill influence when they were experienced after performing the task with fashion dolls.

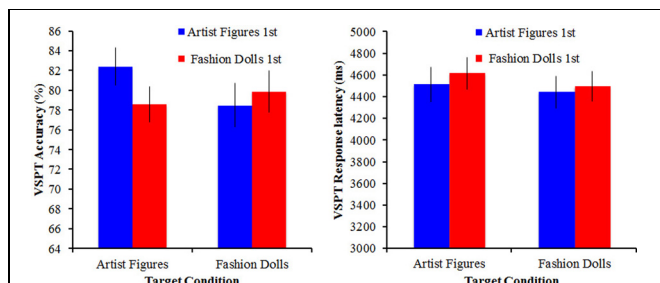


FIGURE 5 | VSPT performance (mean accuracy and response latency) as a function of the target conditions for Experiment 2. Error bars reflect ± 1 standard error of the mean.

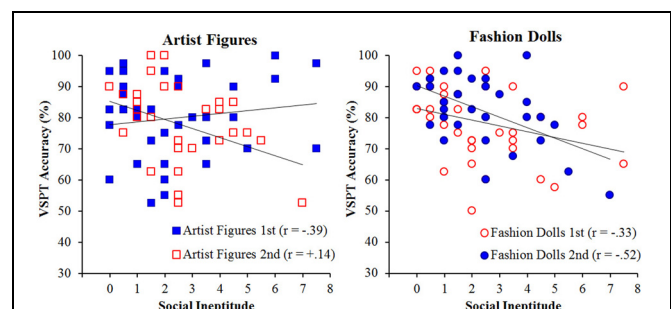


FIGURE 6 | VSPT performance as a function of social ineptitude score separately for the artist figures (left) and fashion dolls (right) conditions broken out by order for Experiment 2.

Although both Experiments 1 and 2 demonstrate how context can affect the sensitivity to social skill influences, the results may seem contradictory. In Experiment 1, having the putatively more agent-like target first increased the sensitivity for the subsequent less agent-like target, whereas Experiment 2 showed the opposite effect. However, the figures used in Experiment 2 were not the same object varying in a feature or two; they were two different representations of human form that varied on a variety of visual features (continuity of form, faces, hair, clothing, etc.). Although the artist figures can clearly convey agency in a way that allows a social skill influence, they are also affected by the context in which they are experienced.

The broader issue of context effects and target influences will be addressed in more detail in the General Discussion, but first we turn our attention to another form of context. All of our manipulations of potential agent-like features so far have been visual features. It is also possible to create conditions in which objects might be viewed as agents via conceptual context. Whether we can induce sensitivity to social skill influence using a conceptual context is the question for Experiment 3.

EXPERIMENT 3: TRIANGLE “ALIENS”

One of the clear conclusions of Experiments 1 and 2 is that objects can be sensitive to social skill influences as a function of having features that suggest potential agency. Whether affixing eyes to simple shapes or using representations of the human form, these features appear to affect how individuals approach the perspective-taking task. We also observed a form of conceptual carryover from the triangles-with-eyes to the plain triangles. As a final test, we asked whether a purely conceptual manipulation could also make an object sensitive to social skill influence. Using the plain triangles again, we offered an alternative interpretation of the triangles as potential agents by calling them “aliens” during the perspective-taking trials. If triangles with eyeballs on top motivate the task to become more social in nature, then perhaps simply suggesting a type of being, be it alien or otherwise, might operate in a similar manner.

MATERIALS AND METHODS

Participants

For this experiment, 53 naïve participants were enrolled. Five participants (3 males) failed to meet criterion, leaving 48 participants (24 males) included in all analyses.

Materials, design, and procedures

Participants completed the 3Bldgs task using the same 7 triangular blocks on pedestals described in Experiment 1. For each image the participant was asked, “Which Alien is at this view?” Across participants, the display type was counterbalanced and the order of the target colors around the display was selected randomly for each participant.

RESULTS AND DISCUSSION

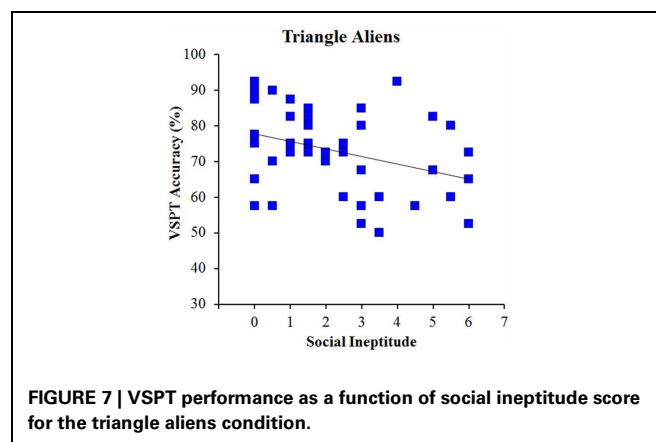
Mean response latency was 3018 and 3054 ms overall and for correct trials only, respectively, and overall accuracy was 72.9%. Again, we observed a significant correlation between the AQ social and communication impairment subscales, $r = +0.46$, so we again used the social ineptitude score (see Table 1

for separate correlations with subscales). The critical correlation between the social ineptitude score and performance on the VSPT task with triangle aliens was significant, $r = -0.36$ (see Figure 7). Moreover, this correlation was not significantly different from the correlation obtained for the triangles-with-eyes condition either overall or separated by order in Experiment 1, $ps > 0.65$. These results suggest that even in the absence of a visual feature, an object can become sensitive to social skill influence on VSPT through conceptual suggestion.

GENERAL DISCUSSION

One of the key motivations for this special issue on developing a framework for integrating the “social” and the “spatial” is the recent acknowledgment of a clear relationship between VSPT and aspects of one’s savvy in social situations (Hamilton et al., 2009; Zwicker et al., 2011; Brunyé et al., 2012; Shelton et al., 2012). By definition, VSPT tasks involve the need to consider/imagine/reason about a target perspective that is different from one’s own. As such, it is tempting to conclude that VSPT is both a spatial and social task. However, these tasks can vary dramatically with respect to who or what is the placeholder for the target perspective. Our previous work has shown that the relationship between VSPT performance and social skills depends on the nature of the target; targets that could be seen as potential agents were sensitive to social skill influence, whereas targets that did not have agent-like features were insensitive (Shelton et al., 2012). These and similar results (Zwicker, 2009; Zwicker et al., 2011; Schilbach et al., 2012; Zwicker and Müller, 2013) suggest that VSPT tasks *acquire* some social relevance when the target has potential agency. Using correlations between measures of social skills and performance on VSPT, the present study offers a more detailed account of what features and conditions can affect the degree to which any given target might convey agency and become sensitive to social skill influence.

One of the first implications of the present work is that simple physical features can induce an object to be sensitive to social skill influence. In the previous study, the physical form of a wooden artist figure appeared to motivate participants to engage the VSPT



task in a more social manner than either plain triangles or cameras (Shelton et al., 2012). Although triangles do not resemble human forms, in the present study, we were able to observe the same social skill influence on VSPT by simply adding eyeballs to the triangles. This small featural change seemed to push participants to engage the triangles as if they were potential agents, like the artist figures. Taken together, this body of work suggests that there are different visual features that can motivate one to perceive an object as a potential agent. The global shape of the artist figures likely conveyed the sense that this could be a person form, whereas the presence of eyes (which the artists figure did not actually have) likely conveyed a similar sense for triangles. These differences raise important questions about what other features or combinations might be more or less effective in engaging the social mechanisms that appear to be brought to bear, but they establish the very basic notion that minimal change can assert a strong influence.

In addition to the observation that physical features can induce agency, we also observed that the context defined by experiences surrounding the introduction of particular targets in perspective taking can also play a role in the perceived potential agency. First, we found that when a target closely resembles another object that has recently been attributed with agency, the similarity of object features may be sufficient to convey carryover agency. Performing VSPT with plain triangles as targets is generally insensitive to social skill influence when performed prior to other conditions or in the context of artist figures or cameras. That is, on their own, they are not potential agents. Despite their fundamental characterization as inanimate objects, these same plain triangles can be engaged as if they were potential agents when they immediately follow experience with identical triangles with eyes affixed to them. In other words, experiencing triangles-with-eyes as agents allowed plain triangles to be viewed as agents.

The artist figures provide a second case of experiential context. When performing the task with artist figures in the context of plain triangles or cameras, the artist figures represent the condition that shows the strongest numerical relationship between social skills and performance (Shelton et al., 2012). Similarly, when participants experienced the task with the artist figures first in Experiment 2, we saw a similar performance-social skill correlation to other “agent” conditions. However, when participants experienced the artist figures after exposure to the fashion dolls, this correlation was diminished completely for the artist figures condition, such that it had the same magnitude correlation as plain triangles alone. In this case, it was as if experiencing the fashion dolls as agents made the artist figures seem not only *less* agent-like but not at all agent-like, akin to a purely inanimate object.

As noted previously, the two types of triangles and two types of dolls represent opposing effects of experiential context (for a more in-depth discussion of experiential social context or historicity, see Schilbach et al., 2013). For the triangles, we had identical objects that differed only in the presence or absence of a single agent-like feature (eyes). Whatever attributes the eyes appeared to bring to the triangles on which they were affixed

(triangles-with-eyes condition) lingered when the triangles were presented again without eyes (plain triangles). One argument might be that the triangles were viewed as the same triangles with and without eyes. That is, once the triangles had been imbued with potential agency, having them “return” without eyes did not immediately strip them of the attributes that were motivating the task to take on social relevance. By contrast, the fashion dolls and artists figures are clearly not the same object but are different representations of human form. Although artist figures alone may be sufficient to engage the social mechanisms that affect performance with agent-like targets, they appear to lose any attributes that convey a sense of agency when one has had experience with the more representative form of the fashion doll. Anecdotally, some participants even remarked that the artist figures seemed “creepy” after seeing the fashion dolls. This type of comment has never been noted before in our previous studies nor in the condition when the artist figures came first, suggesting that the experiential context was asserting a strong influence over the perception of the objects at a very basic level.

Although not significantly different from the cases presented here, it is notable that the artist figures in Shelton et al. (2012) had a numerically larger correlation than any of the conditions in the present study. This may be due to the fact that the artist figures were experienced in the context of two other objects that did not possess agent-like qualities (plain triangles and cameras). In 2/3 of the orders used for the study, the artist figures would have come after one or both of the object conditions, which may have magnified the correlation relative to the conditions used in the present study. The lack of order effects on the correlations in the original study is potentially problematic for this argument. However, given that the artist figures when presented first had a correlation of -0.39 in the present study it is possible that the original study was insufficiently powered to detect changes from such a high baseline correlation. Again, this suggests that the role of experiential context may be a critical factor in the way individuals engage a VSPT task with different targets.

Both the induction of agency and contextual effects noted above are driven by physical features. For example, the carry-over of agency from triangles-with-eyes to plain triangles likely depended on the agency evoked by the physical features of the eyes carrying over to the highly similar plain triangles. Likewise, the many agent-like features of the fashion dolls appeared to convey agency, and the diminished agency for artist figures following experience with fashion dolls likely depended on the contrast of these “rich” agents to the highly dissimilar artist figures. However, the conveyance of agency is not limited to physical features. In Experiment 3, the performance on the VSPT task with plain triangles was again correlated with social skills when the triangles were referred to as aliens rather than triangles. Implying a being, albeit an alien being, appears to be similar to adding physical features such as eyeballs, motivating participants to engage in the task in a way that allows social skills to assert influence. This suggests that when interpreting the targets of a VSPT task, people may have a very low threshold for allowing a target to be “social.”

A running theme through this work is the introductory framework in which the target and possibly other features of a VSPT task can affect whether the task itself is socially relevant or not. Using this framework, we have demonstrated that the degree to which a task appears to be social or not involves the complex interaction of target features and experiential context, which includes the presence of other targets and the language used to identify the targets. One might be tempted to conclude that this work is largely about methodology, and in some sense, this is the case. Although we have not exhaustively tested all types of targets or target combinations, our findings clearly offer some suggestions for how to craft a VSPT task that is or is not sensitive to social skill influence. For example, if one's goal is to understand VSPT in isolation, what we have termed non-social VSPT, then one can design a task that limits the potential agency as much as possible. However, this work also speaks to and raises many deeper theoretical issues about how the human brain processes spatial information in order to reason about the world.

A first critical point is that the sensitivity of VSPT performance to targets and experiential context is consistent with the notion that VSPT in real-world settings is not a task that happens in isolation. For example, imagine sitting in the stands at a ball game waiting for a friend. Your friend is lost, but you can see him under the scoreboard. By phone, you might give him directions based on what you know he can currently see. The particular directions you give might be influenced by a wide variety of concerns—your friends known ability to mix up left and right, the urgency with which you want to get him to the seat before first pitch, whether you want him to pass the concession stand to grab refreshments, etc. In this example, the ability to relate to the friend's situation involves both the understanding of visual-spatial perspective (what the friend can see) but also the socially relevant situational factors (the friend's state, abilities, goals). Therefore, one's performance should be dependent on the interaction of spatial and social skills such that this socially relevant VSPT situation will benefit if social skills are strong but might be hindered if they are weaker.

This still leaves the broader question of why the specific judgments participants were asked to do in our VSPT task should be influenced by social skills when targets appear to convey agency. That is, one could imagine that the ballgame example could be accomplished by having the spatial reasoning done by a "purely" spatial computational process with social skills only entering at the point of deciding how to communicate that information to an agent. Our task only requires the judgment and not a tailored communication of the outcome, suggesting that the social influence may be operating throughout the process of perspective taking. Although our results are ambivalent with respect to how the presence of an agent-like target might be altering the underlying computations, we offer some speculation about how this interaction might come about. One possibility is that VSPT involves first essentially embodying the target of the potential perspective one is attempting to assess (i.e., one is attempting to assume the target's position in order to see its viewpoint). As may be the case with spontaneous perspective taking

(Tversky and Hard, 2009), a potential agent as a target may automatically induce one to consider the target's social/personality attributes. One's comfort level in understanding or appreciating these attributes may then serve to gate how readily the perspective can be assumed. For example, an individual who is more socially savvy might be able to more readily recognize the utility of a potential agent through efficient assessment of relevant attributes (e.g., the eyes on the fashion doll means she might have the ability to see) and dismissal of irrelevant and unknown attributes (e.g., she is happy and has good fashion sense), whereas an individual with less social savvy might experience inhibition in trying to take the perspective of a potential agent because he/she cannot apprehend the attributes as readily and choose those that would facilitate embodiment. By contrast, when the perspective taking involves an object as the target, there are no obvious social attributes, so one's social savvy will neither hurt nor help, as it is irrelevant.

In this working model, we do not propose different mechanisms or processes for VSPT with agents vs. objects. Instead, we are suggesting that there is a common spatial component irrespective of the target, which is consistent with the observed correlations between the different versions of the VSPT task (targets with and without agency). However, when there is an agent, social skills may act as a gateway for the initial step of embodying the target. This proposed role suggests that we will still see individual differences in the spatial aspects of the tasks, regardless of target type, but we will have additional variability due to social skills when the target is a potential agent. Whether the proposed model above or an alternative framework ultimately captures the interaction of social and spatial skills, this work motivates a deeper question: what are the advantages of having a system that is generally sensitive to the presence of agents for a seemingly spatial task given that this sensitivity can benefit some individuals and hinder others relative to non-social conditions? Is this driven by the folk wisdom that humans are simply social beings? These are open questions and ones not readily addressed empirically, but they provide fodder for thinking critically about the interactive nature of human cognition.

The overarching goal of this project was to deepen the exploration of factors that determine whether and when VSPT might be sensitive to the influence of social skills. Taken together, the results suggest that the social influence on VSPT is mediated by a complex relationship that includes the task, the target, and the context in which the target is perceived. Future studies may continue to elaborate on the various boundary conditions that evoke agency or take it away, but the broader message from our work and similar studies is the importance of thinking beyond the bounds of a single domain to explain the complexity of human behavior.

ACKNOWLEDGMENTS

We would like to thank Ben Nelligan for helpful comments on previous versions of this manuscript. Portions of this work were supported in part by a Woodrow Wilson undergraduate research fellowship that was awarded to Alexandra J. Murray.

REFERENCES

- Ames, D. L., Jenkins, A. C., Banaji, M. R., and Mitchell, J. P. (2008). Taking another person's perspective increases self-referential neural processing. *Psychol. Sci.* 19, 642–644.
- Baron-Cohen, S. (1992). The theory of mind hypothesis of autism: history and prospects of the idea. *Psychologist* 5, 9–12.
- Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., and Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J. Autism Dev. Disord.* 31, 5–17. doi: 10.1023/A:1005653411471
- Best, C., Moffat, V., Power, M., Owens, D., and Johnstone, E. (2008). The boundaries of the cognitive phenotype of autism: theory of mind, central coherence and ambiguous figure perception in young people with autistic traits. *J. Autism Dev. Disord.* 38, 840–847. doi: 10.1007/s10803-007-0451-8
- Brunyé, T. T., Ditman, T., Giles, G. E., Mahoney, C. R., Kessler, K., and Taylor, H. A. (2012). Gender and autistic personality traits predict perspective-taking ability in typical adults. *Pers. Individ. Dif.* 52, 84–88. doi: 10.1016/j.paid.2011.09.004
- David, N., Aumann, C., Santos, N. S., Bewernick, B. H., Eickhoff, S. B., Newen, A., et al. (2008). Differential involvement of the posterior temporal cortex in mentalizing but not perspective taking. *Soc. Cogn. Affect. Neurosci.* 3, 279–289. doi: 10.1093/scan/nsn023
- David, N., Bewernick, B. H., Cohen, M. X., Newen, A., Lux, S., Fink, G. R., et al. (2006). Neural representations of self versus other: visual-spatial perspective taking and agency in a virtual ball-tossing game. *J. Cogn. Neurosci.* 18, 898–910. doi: 10.1162/jocn.2006.18.6.898
- Eppel, M. L., Gatz, M., and Frank, S. J. (1983). The relationship between children's development of abilities in social and spatial perspective-taking. *J. Genet. Psychol.* 143, 143–144.
- Gould, E., Tarbox, J., O'Hara, D., Noone, S., and Bergstrom, R. (2011). Teaching children with autism a basic component skill of perspective-taking. *Behav. Intervent.* 26, 50–66. doi: 10.1002/bin.320
- Hamilton, A. F., Brindley, R., and Frith, U. (2009). Visual perspective taking impairment in children with autistic spectrum disorder. *Cognition* 113, 37–44. doi: 10.1016/j.cognition.2009.07.007
- Hobson, R. P. (1984). Early childhood autism and the question of egocentrism. *J. Autism Dev. Disord.* 14, 85–104.
- Kessler, K., and Rutherford, H. (2010). The two forms of visuo-spatial perspective taking are differently embodied and subserved different spatial prepositions. *Front. Psychol.* 1:213. doi: 10.3389/fpsyg.2010.00213
- Kessler, K., and Thomson, L. A. (2010). The embodied nature of spatial perspective taking: embodied transformation versus sensorimotor interference. *Cognition* 114, 72–88. doi: 10.1016/j.cognition.2009.08.015
- Kessler, K., and Wang, H. (2012). Spatial perspective taking is an embodied process, but not for everyone in the same way: differences predicted by sex and social skills score. *Spat. Cogn. Comput. Interdiscip. J.* 12, 133–158. doi: 10.1080/13875868.2011.634533
- Piaget, J., and Inhelder, B. (1967). *The Child's Conception of Space*. New York, NY: Norton.
- Ruby, P., and Decety, J. (2001). Effect of subjective perspective taking during simulation of action: a PET investigation of agency. *Nat. Neurosci.* 4, 546–550.
- Schilbach, L., Eickhoff, S. B., Cieslik, E. D., Kuzmanovic, B., and Vogeley, K. (2012). Shall we do this together? Social gaze influences action control in a comparison group, but not in individuals with high-functioning autism. *Autism* 16, 151–162. doi: 10.1177/1362361311409258
- Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., et al. (2013). Toward a second-person neuroscience. *Behav. Brain Sci.* 36, 393–414. doi: 10.1017/S0140525X12000660
- Schober, M. F. (1998). "How addressees affect spatial perspective choice in dialogue," in *Representation and Processing of Spatial Expressions*, eds P. Olivier and K.-P. Gapp (Mahwah, NJ: Lawrence Erlbaum Associates), 231–245.
- Shelton, A. L., Clements-Stephens, A. M., Lam, W. Y., Pak, D. M., and Murray, A. J. (2012). Should social savvy equal good spatial skills? The interaction of social skills with spatial perspective taking. *J. Exp. Psychol. Gen.* 141, 199–205. doi: 10.1037/a0024617
- Tversky, B., and Hard, B. M. (2009). Embodied and disembodied cognition: spatial perspective-taking. *Cognition* 110, 124–129. doi: 10.1016/j.cognition.2008.10.008
- Zwicker, J. (2009). Agency attribution and visuospatial perspective taking. *Psychon. Bull. Rev.* 16, 1089–1093. doi: 10.3758/PBR.16.6.1089
- Zwicker, J., and Müller, H. J. (2013). On the relation between spontaneous perspective taking and other visuospatial processes. *Mem. Cogn.* 41, 558–570. doi: 10.3758/s13421-012-0283-4
- Zwicker, J., White, S. J., Coniston, D., Senju, A., and Frith, U. (2011). Exploring the building blocks of social cognition: spontaneous agency perception and visual perspective taking in autism. *Soc. Cogn. Affect. Neurosci.* 6, 564–571. doi: 10.1093/scan/nsq088

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 02 July 2013; accepted: 05 August 2013; published online: 05 September 2013.

Citation: Clements-Stephens AM, Vasiljevic K, Murray AJ and Shelton AL (2013) The role of potential agents in making spatial perspective taking social. *Front. Hum. Neurosci.* 7:497. doi: 10.3389/fnhum.2013.00497

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2013 Clements-Stephens, Vasiljevic, Murray and Shelton. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The influence of visual perspective on the somatosensory steady-state response during pain observation

Dora L. Canizales^{1,2,3}, Julien I. A. Voisin^{3,4}, Pierre-Emmanuel Michon³, Marc-André Roy² and Philip L. Jackson^{1,2,3*}

¹ École de Psychologie, Université Laval, Québec, QC, Canada

² Centre de Recherche de l'Institut Universitaire en Santé Mentale de Québec, Québec, QC, Canada

³ Centre Interdisciplinaire de Recherche en Réadaptation et Intégration Sociale, Québec, QC, Canada

⁴ Département de Réadaptation, Université Laval, Québec, QC, Canada

Edited by:

Antonia Hamilton, University of Nottingham, UK

Reviewed by:

Cosimo Urgesi, University of Udine, Italy

Leonie Koban, University of Colorado Boulder, USA

*Correspondence:

Philip L. Jackson, École de Psychologie, Université Laval, Pavillon Félix-Antoine-Savard, 2325, rue des Bibliothèques, Québec, QC G1V 0A6, Canada
e-mail: philip.jackson@psy.ulaval.ca

The observation and evaluation of other's pain activate part of the neuronal network involved in the actual experience of pain, including those regions subserving the sensori-discriminative dimension of pain. This was largely interpreted as evidence showing that part of the painful experience can be shared vicariously. Here, we investigated the effect of the visual perspective from which other people's pain is seen on the cortical response to continuous 25 Hz non-painful somatosensory stimulation (somatosensory steady-state response: SSSR). Based on the shared representation framework, we expected first-person visual perspective (1PP) to yield more changes in cortical activity than third-person visual perspective (3PP) during pain observation. Twenty healthy adults were instructed to rate a series of pseudo-dynamic pictures depicting hands in either painful or non-painful scenarios, presented either in 1PP (0–45° angle) or 3PP (180° angle), while changes in brain activity was measured with a 128-electrode EEG system. The ratings demonstrated that the same scenarios were rated on average as more painful when observed from the 1PP than from the 3PP. As expected from previous works, the SSSR response was decreased after stimulus onset over the left caudal part of the parieto-central cortex, contralateral to the stimulation side. Moreover, the difference between the SSSR was of greater amplitude when the painful situations were presented from the 1PP compared to the 3PP. Together, these results suggest that a visuospatial congruence between the viewer and the observed scenarios is associated with both a higher subjective evaluation of pain and an increased modulation in the somatosensory representation of observed pain. These findings are discussed with regards to the potential role of visual perspective in pain communication and empathy.

Keywords: pain observation, visual perspective, empathy, electroencephalography, somatosensory steady-state

INTRODUCTION

Seeing pain in other people is an experience susceptible to trigger responses akin those felt when we hurt ourselves. Indeed, several neuroimaging studies found a partial overlap between cerebral circuits involved during actual experience of pain (known as the pain matrix) and during the observation of other's pain (for a recent review see Fan et al., 2011; Lamm et al., 2011). More specifically, the observation of other's pain engages in a similar way some of the neuronal systems subserving the sensory (e.g., somatosensory cortices) and the affective components (e.g., insula, anterior cingulate cortex) of self pain perception (Fitzgibbon et al., 2010; Corradi-Dell'Acqua et al., 2011). The activation of these regions suggests, indirectly, that observing pain produces a fine-grained multidimensional mental representation of the other's pain even in the absence of somatosensory input. Even though this mental representation of pain enables an observer to partially share the subjective experience of other's pain, the extent of this sharing

mechanism can vary according to different factors (Coll et al., 2011). Note that the specificity of the nociceptive cerebral representation itself (signature; Wager et al., 2013) is currently being debated and the extent to which this representation also codes for the pain of others (Krishnan et al., 2013) and social pain (Iannetti et al., 2013) appears to be more limited than initially thought.

Nevertheless, recent neuroimaging studies have provided strong evidence that observation of pain can involve the somatosensory cortex, known to contribute to sensory processing of noxious stimuli (Bufalari et al., 2007; Lamm et al., 2007b; Cheng et al., 2008; Betti et al., 2009; Han et al., 2009; Yang et al., 2009; Voisin et al., 2011a). Moreover, the recruitment of sensory processes of pain is distinctly demonstrated by a decrease of the somatosensory response during pain observation (Cheng et al., 2008; Voisin et al., 2011a; Marcoux et al., 2013). Although the involvement of the sensory cerebral circuits during pain observation has been repetitively demonstrated (see Keysers et al., 2010;

Lamm et al., 2011, for reviews), the variables that modulate these sensory processes remain unclear.

A growing body of evidence supports the idea that somatosensory activity is influenced by *perspective taking* (PT; Ruby and Decety, 2001, 2003, 2004; Jackson et al., 2006a,b). PT can be defined as the ability to adopt someone else's point of view in order to understand their situation (Decety et al., 2006). This ability represents an essential component of empathy, which refers to the faculty to understand and to share other's emotions and feelings and to respond appropriately (Decety et al., 2006). Studies generally distinguish cognitive PT, which requires the individual to imagine being the other person (e.g., Ruby and Decety, 2003; Jackson et al., 2006b; Dosch et al., 2010) from visual PT, which involves seeing a scene or a situation from different angles (e.g., Jackson et al., 2006a; Kessler and Thomson, 2010). Regarding cognitive PT, several studies demonstrated that thinking about oneself in a specific situation generates different behavioral and cerebral responses than imagining another person in the same situation (Ruby and Decety, 2004; Jackson et al., 2005, 2006b; Lamm et al., 2007a, 2008; Li and Han, 2010). While self-perspective requires fast and automatic processes, which are more related to agency (i.e., ability to attribute the origin of an action), adopting the perspective of others appears to engage more deliberate and regulatory mechanisms (van der Heiden et al., 2013).

Visual PT generally results from a mental rotation of one's own perspective toward the other's perspective in order to consider the spatial information from the other's viewpoint that may be different from the subject's one (Kozhevnikov and Hegarty, 2001; Kessler and Thomson, 2010). Visual PT provides crucial spatial information that enables a person to appropriately conduct social interaction and understand other's mental states (Langdon and Coltheart, 2001; Kaiser et al., 2008; Lambrey et al., 2008; Kessler and Thomson, 2010). The manipulation of the visual perspective is broadly used in cinematography (e.g., subjective/objective camera), particularly in horror movies and video games (e.g., first/third-person games) in order to generate the feeling in the spectator of sharing the point of view of the character.

Behavioral studies generally report faster reaction time and increased accuracy performance when an object or an action is seen in a *first-person* visual perspective (1PP) (i.e., seeing a situation from the onlooker's viewpoint) compared to a *third-person* visual perspective (3PP) (i.e., seeing a situation presented in someone else's viewpoint) (Jackson et al., 2006a; Kaiser et al., 2008). Jackson et al. (2006a) demonstrated that seeing or imitating actions performed in the 1PP yielded stronger sensorimotor activation in comparison to the 3PP. This supports the assumption that adopting 1PP generates more robust sensorimotor representation of the action in the onlooker's brain that may be close to actual execution of the action. 3PP also seems to involve specific neuronal processes associated with spatial transformations (Jackson et al., 2006a; Kaiser et al., 2008; Callan et al., 2012), visual motion perception (Bundo et al., 2000; de Lussanet et al., 2008), and executive functions such as inhibition and attention (Hampshire et al., 2010; Dodds et al., 2011). Altogether, these findings show that adopting 1PP and 3PP requires distinct neuronal processes: the former may be more associated with

automatic embodiment (resonance) and the latter with cognitive functions such as visuospatial processing and inhibition.

The main objective of this study was to determine if the point of view of the observer (visual perspective) can specifically modulate the behavioral and cerebral responses to painful visual stimuli. To do so, we compared the modulation of the somatosensory steady-state response (SSSR) during the observation of painful visual stimuli depicted in a 1PP and 3PP. Firstly, we hypothesized that participants would attribute higher pain ratings to painful pictures depicted in the 1PP. Secondly, we suggested that seeing the pictures would produce an automatic decrease of the SSSR amplitude (i.e., *initial gating effect*), which will occur mainly over the left parietal cortex as this region was previously found to be more responsive to steady-state somatosensory stimulation (Voisin et al., 2011a,b; Marcoux et al., 2013). Thirdly, we predicted that this SSSR modulation would be *a priori* greater when painful situations were presented in a first-person compared to a 3PP (i.e., *visual perspective effect*). Finally, we have also examined the association between the SSSR response and self-reported measures of different components of empathy.

METHODS

PARTICIPANTS

The sample was composed of 20 healthy right-handed Caucasian volunteers (nine men, mean age = 25 ± 5 years). Participants had no history of neurological, psychiatric or pain related disorders, and visual acuity was normal or corrected. This participant had not completed the whole experiment and quit due to discomfort during the task. The study was approved by the Research Ethic Committee of the Institut de Réadaptation en Déficience Physique de Québec. Participants gave written informed consent and received a small monetary compensation for their participation.

MATERIAL

VISUAL STIMULI AND EXPERIMENTAL PROCEDURE

Pseudo-dynamic visual stimuli presented the right hand of adult Caucasian (half male, half female) displayed in 12 different everyday life scenarios (e.g., cutting food with a knife). These scenarios were shown either in a first (1PP: arm at 0–45° angle) or third person visual perspective (3PP: arm at ~180° angle). The scenarios ended in a painful or nonpainful situation (Pain vs. NoPain condition). There were 12 different scenarios displaying two types of visual perspective (1PP and 3PP), two pain levels (Pain and NoPain), and two models' sex (Male and Female), giving a total of 96 different visual stimuli. Visual stimuli were perceived as dynamic because they were composed of a sequence of three pictures, respectively displayed for 750, 250 and 1500 ms for a total length of 2500 ms. The participants could see the type of visual perspective from the first picture on, but the painful vs. nonpainful outcome appeared only in the third picture. This was done to equate pain anticipation across conditions. Note that the motor and sensory components in the stimuli varied (hand moving away from a situation, danger approaching hand) but these were distributed across conditions so as to avoid bias. This relative heterogeneity improves ecological validity and reduces repetitiveness, which could lead to habituation effects.

The experimental task, scripted in E-Prime (Version 2.0, Psychology Software Tools, Inc.), contained eight blocks of 24 trials in which the four conditions (two Pain levels [pain, no pain] \times two Perspectives [first, third]) were presented six times each in random order. Each trial comprised a fixation cross (2500 ms), the dynamic visual stimulus (2500 ms) and a verbal numerical rating scale ranging from 0 (No Pain) to 10 (Worst Pain) (3000 ms) (see **Figure 1**). The total length of a block was 4 min. The gender of the person on the visual stimuli was equally and randomly distributed in each block. The participants were instructed to verbally rate the intensity of pain observed after each picture once the numeric scale appeared on the screen. To make sure that the instructions were well understood, the participants also completed a short practice session (12 trials) before the experimental session. A trial sample is shown in **Figure 1**.

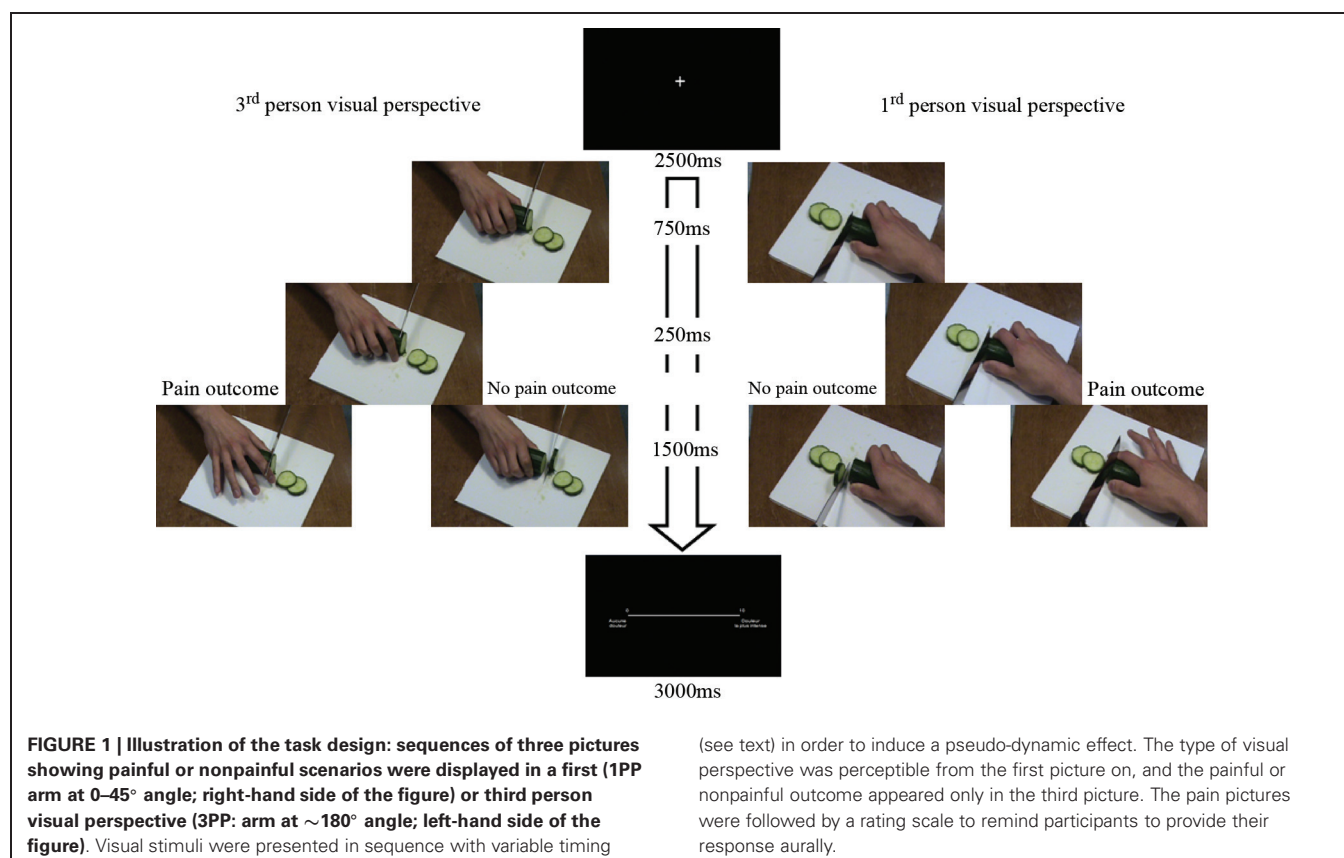
Throughout each trial block, a continuous and nonpainful mechanical stimulation (25 Hz) provided by a custom-made cylinder-shaped vibrotactile stimulator (10 cm long, 3 cm diameter) held in the participants' right hand. The right hand rested on an armrest and electromyographic activity was recorded (MP150 system, Biopac Inc.) with Ag-AgCl surface electrodes positioned in bipolar configuration over the first dorsal interosseus muscle (FDI). Participants were told to not contract the stimulator with their hand during the experimental session. EMG activity was visually examined to monitor that participants did not change the grip significantly on the stimulator.

INTERPERSONAL REACTIVITY INDEX QUESTIONNAIRE

A French translation of the Interpersonal Reactivity Index (IRI; Davis, 1980) self-report questionnaire was administered to the participants. The IRI is a measure of dispositional empathy in which participants had to determine the level of agreement or disagreement about thoughts and feelings in a variety of situations using a 5-point Likert-type scale. The IRI contains four 7-item subscales: the Empathic Concern (EC) scale measures the tendency to experience feelings of sympathy and compassion for others; the Personal Distress (PD) scale evaluates the inclination to feel discomfort and helplessness in response to other's people distress; the PT scale assesses the propensity to adopt other person's point of view; and the Fantasy (F) scale measures the tendency to imagine oneself into fictional situations. The score on each subscale is used as independent measure of four abilities related to empathy.

EEG

During the EEG data acquisition, the participants were comfortably seated on a chair with armrest in a quiet dark room. An EEG helmet with 124 + 4 Ag/AgCl electrodes contacting scalp surface by way of saline-soaked sponges (Electrical Geodesic Inc., OR, USA) was used to record the cerebral activity. The sampling rate was set at 500 Hz and the electrodes impedances were kept below 50 k Ω .



ANALYSES

BEHAVIORAL ANALYSES

First, trials where participants gave an incorrect response were discarded from subsequent analyses. An incorrect response occurred when a participant rated a painful picture as nonpainful (rating of zero) and a nonpainful picture as painful (ratings of 1 to 10). This procedure was applied to make sure that the participants have correctly categorized the visual stimuli and were paying attention to the task. The analyses were conducted on correct painful trials only. A repeated measure ANOVA (Pain, Perspective and Block conditions) was computed to confirm that evaluation of painful visual stimuli were not influenced by habituation across blocks. Ratings for the nonpainful pictures were not kept for the subsequent analyses. A paired *t*-test was calculated for the difference between mean rating of painful pictures presented in the 1PP and 3PP. All statistical analyses were computed with the SPSS v.13 software (SPSS Inc., Chicago, IL, USA).

EEG PRE-PROCESSING AND ANALYSES

The EEG analyses were ran using the locally-developed software ELAB plus the ELAN-Pack (Aguera et al., 2011) and MATLAB software (version 6.5; The Math-Works inc., Natick, MA). Note that trials that were removed from the behavioral data (see above) were also rejected from the EEG data. Visual inspection of very high level of noisy data led to the rejection of all data from a second participant who, furthermore, demonstrated important signs of anxiety and agitation relatively to the EEG apparatus. The EEG signal was cleaned from blinks, muscle activity, fast baseline shift, and high inter-electrodes impedance. More specifically, any 100 ms-long sample was rejected if it included one of these events: (i) in the same electrode channel, the scalp potential exhibited variation over 50 μV within a 10 ms time window; (ii) in the same electrode channel, the energy content was more than 500 μV^2 in the 60–100 Hz band; or (iii) 800 μV^2 in the 23–27 Hz band; and (iv) in any electrode channel, the scalp potential exhibited a variation larger than 150 μV within a 200 ms time window. A total of 20.18 % (SD = 10.5%) of the samples was rejected according to these criteria, without distinction for the type of stimuli (Pain-1PP: 20.76%, Pain-3PP: 19.57%; NoPain-1PP: 21.46%; NoPain-3PP: 18.94%). Moreover, a participant was rejected if each block contained at least 50% of noise. With respect to this criterion, a third participant was removed from the subsequent analyses. Next, a spherical spline interpolation process (Tikhonov regularization) was applied to the remaining data samples. The extraction of the 25 Hz energy band frequency was then performed on EEG data by applying complex Gaussian Morlet's wavelets in order to produce time-frequency maps of the SSSR corresponding to the 25 Hz vibrotactile stimulation in a time interval. Notice that the combination of these two steps requires to reject either the whole sample or the whole electrode each time a faulty sample is found in one electrode, which leads to an increase in the number of rejected samples (Voisin et al., 2011b,c). As a final control of quality, samples with oscillatory activity over > 600 μV^2 and any trial those reconstructed from less than 70% of the original raw data were rejected from subsequent analyses. At the end of this pre-processing, 29.8% (SD = 17.6%) of data were rejected.

Determination of the a priori region of interest (ROI) was similar to Voisin et al. (2011a,b,c) and Marcoux et al. (2013). The grand mean of the signal (i.e., blind to the actual experimental conditions) of five combinations of three surrounding electrodes over the parietal cortex were examined using paired *t*-tests during the last 200 ms pre-stimulus (i.e., during the fixation cross) versus the 200 ms time bins during the first picture presentation (i.e., before the subject could identify the condition). This approach, similar in spirit to a localisationer run in fMRI, was conducted specifically to select which group of three electrodes showed the highest gating response to the vibrotactile stimulation. As the experimental condition is not used as a criterion, the procedure has no impact on the statistical tests for the visual perspective. SSSR during the overall time course was divided in 200 ms wide time bins sampled every 100 ms that were used for the subsequent analyses to detect differences in the modulation with higher accuracy.

To assess the impact of the experimental conditions on SSSR modulation, a statistical analysis similar to that described in Decety et al. (2010) and Li and Han (2010) was used. Namely, the statistical analyses were conducted separately on two time windows for which the experimental conditions differed. In first initial gating window, no indication of pain was present, so the analysis focused on the presence of a SSSR modulation after picture onset. This variation consisted in the difference between the last 200 ms pre-stimulus (i.e., during the fixation cross), and two 200 ms long time bins immediately following the first picture presentation (i.e., 200 to 500 ms, with 100 ms overlap). The mean energy of this initial gating window was computed with paired *t* tests.

In the second time window, which refers to the specific gating, the raw SSSR was normalized to its corresponding baseline (the last 200 ms portion of the second picture, i.e., before the painful picture outcome apparition), using the following equation: (SSSR-baseline)/baseline (see Marcoux et al., 2013). Then, mean of SSSR ratios of all 200 ms time bin (total of 9 consecutive time bins) within the specific gating window (1200–2200 ms) were computed in order to systematically assess the visual perspective effect (1PP*3PP) during pain observation using paired *t*-tests. *P*-values are reported with the Bonferroni corrected alpha values. A Pearson *r* correlation analysis (two-tailed, statistical thresholds: $p < 0.05$) was performed to measure the relation between SSSR (significant time bins only) and pain ratings of painful visual stimuli in 1PP and 3PP condition. Correlation analyses were also conducted on IRI subscales and SSSR when participants were watching painful situations presented in 1PP and 3PP.

RESULTS

BEHAVIORAL RESULTS

The percentage of correct responses was very high (mean 96.6%, SD = 3.3). Lower percentage of incorrect responses tends to be found in 1PP (mean = 2.8%, SD = .18) comparatively to 3PP (mean = 3.6%, SD = .21), but this difference was not significant ($t(16) = -1.82, p = 0.09$). A repeated measures ANOVA performed on mean pain ratings for each Perspective (1PP vs. 3PP) and Pain (Pain vs. NoPain) levels revealed no significant difference across

the blocks (8) (Interaction : $F(7, 9) = 1.77, p = .21$), indicating that the pain ratings did not differ over time between these conditions (e.g., no significant habituation effect). Paired t -test performed on means of ratings of pain intensity in painful pictures showed a significant difference in Perspective condition ($t(16) = 2.25, p = .02$). Participants rated painful pictures in 1PP significantly higher (mean = 5.39, SD = 2.04) than those in 3PP (mean = 5.31, SD = 2.06).

EEG RESULTS

The initial gating window (200 to 500 ms): The map of the cortical amplitudes in the 25 Hz band confirmed that electrodes 66, 67 and 71 (128-HydroCell Geodesic Net, Electrical Geodesic) showed the highest gating response to the vibrotactile stimulation, which corresponded to the posterior parieto-central region contralateral to the stimulation (comparable to electrodes P3-P1-PZ of the 10–20 coordinate system). The paired t -tests conducted on this ROI revealed a significant decrease of the SSSR amplitude values (all $ps < .001, \alpha = .03$) during the display of the first picture (200 ms just after first image onset), and this suppression remained for all 200 ms time bins during the first picture presentation.

The visual PT effect (specific gating window, i.e., 1200 to 2200 ms): A significant effect of the Perspective condition on the SSSR was found for the 1900 to 2100 ms period (i.e., 900 and 1100 ms after the onset of the third picture) ($t(16) = -2.89, p = .005, \alpha = .006$). The mean of SSSR ratios showed a larger decrease for painful pictures depicted in 1PP relative to those in 3PP. No significant effect of visual perspective was found in the other SSSR 200 ms time bins during the specific gating window (1200–1400 ms: $t(16) = -.98, p = .17$, 1300–1500 ms: $t(16) = -1.24, p = .12$, 1400–1600 ms: $t(16) = -1.41, p = .09$, 1500–1700 ms: $t(16) = -.06, p = .48$, 1600–1800 ms: $t(16) = .15, p = .44$, 1700–1900 ms: $t(16) = -.19, p = .43$, 1800–2000 ms: $t(16) = -1.01, p = .17$, 2000–2200 ms: $t(16) = -2.07, p = .03$, all $\alpha = .006$), although some tendencies were found which did not survive the correction for multiple tests. Notice that, although known too severe, Bonferroni correction here reaches the same conclusion as more powerful correction such as Holm-Bonferroni method. SSSR modulation differences between conditions are presented in **Figure 2**.

CORRELATION ANALYSES

No significant correlations were found between SSSR initial gating and IRI subscales (EC: $r = .47, p = .06$, PD: $r = -.3, p = .34$, PT: $r = -.33, p = .2$, F: $r = .03, p = .92$). Ratios of individual SSSR decrease during observation of painful pictures in 1PP and 3PP (i.e., during the 3rd picture presentation) were not significantly correlated neither with corresponding pain ratings (1PP: $r = -.11, p = .63$; 3PP: $r = -.12, p = .64$), nor with other IRI subscales (1PP: EC: $r = -.03, p = .91$, PD: $r = .09, p = .73$, PT: $r = -.17, p = .52$, F: $r = -.01, p = .96$; 3PP: EC: $r = -.2, p = .94$, PD: $r = .2, p = .43$, PT: $r = .21, p = .43$, F: $r = .15, p = .59$).

DISCUSSION

The present study demonstrated, for the first time, that the visual perspective from which pain is observed could influence both

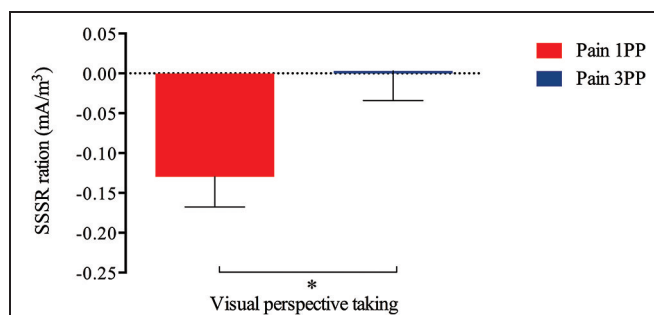


FIGURE 2 | A significant difference ($t(16) = -2.89, p = .005, \alpha = .006$) was detected according to visual perspective after the painful outcome onset (i.e., while third picture presentation; 1900 to 2100 ms). More specifically, painful stimuli seen in 1PP produced a higher SSSR decrease ($M = -.13, SD = .15$) than those observed in 3PP ($M = .003, SD = .15$).

the modulation of a somatosensory response and subjective pain evaluation. Using steady-state EEG, this study revealed differences in SSSR according to the visual perspective through which pain was observed in others. The results also confirmed the hypothesis that viewing pseudo-dynamic pictures in 1PP produced higher ratings of pain intensity relatively to those in 3PP. However, the absence of a significant relation between SSSR and subjective ratings of painful visual stimuli could suggest that these responses to pain observation may be underpinned by different empathic constructs.

AN OVERVIEW ON THE SOMATOSENSORY CEREBRAL MODULATION

As hypothesized, the results revealed a strong general gating appearing early at the visual stimuli onset before being specific to pain observation. Given that the visual stimuli were the same for all conditions before the pain outcome onset, this general gating, found in overall somatosensory activation, is not necessarily specific to pain, but rather to observed hands in action. This result is consistent with the hypothesis that somatosensory gating reflects an attention filtering process (Cromwell et al., 2008). This “gating” effect might represent an attention filtering process that rejects incoming irrelevant somatosensory information to focus on those that are motivationally relevant (Montoya and Sitges, 2006).

THE EFFECT OF VISUAL PERSPECTIVE DURING PAIN OBSERVATION

Pain intensity ratings may vary according to the visual perspective from which painful pictures are presented. When pain is seen with self-proximity, as in 1PP, it is perceived as more intense than when it is watched in another’s viewpoint. Moreover, a higher tendency to make incorrect evaluation of pain intensity came out when people viewed pictures in 3PP. So, both cerebral and behavioral findings support the hypothesis that watching pain in one’s own viewpoint enhanced neurophysiological activity and pain intensity judgments. Interestingly, participants declared to have noticed different perspectives in the visual stimuli although they were not directly required to take self or other’s perspective. In other words, this point suggests that participants did not ignore the orientation of visual stimuli while judging pain intensity.

The current study also demonstrated that observing pain from a 1PP or 3PP influences the somatosensory neuronal activity. As mentioned previously, to understand another person's visual perspective, one has to transpose the other's spatial image onto the self perspective (Kozhevnikov and Hegarty, 2001; Kessler and Thomson, 2010). Thus, in either 1PP or 3PP visual perspectives, people have to mentally simulate an egocentric visual representation of the context seen. Similarly, a specific somatosensory modulation was found when the participants were rating pain intensity presented in painful pictures in both visual perspectives. However, a stronger SSSR decrease was found in 1PP relative to 3PP when the participants were evaluating the pain intensity seen in painful scenarios.

These results suggest that painful situations observed in a visual perspective consistent with one's own engage to a greater extent the sensory processes of pain perception comparatively to situations seen from another's person point of view. These findings are consistent with a previous study that has showed that changing the context from which one imagines pain (pain in self compared to pain in others) influences the level of activity in the secondary somatosensory cortex (Jackson et al., 2006b). The sensory-discriminative dimension of pain encodes the main properties of an actual painful sensation such as stimulus localization, intensity and quality discrimination (Treede et al., 1999). Thus, higher somatosensory gating effect may suggest that looking at painful situation from our point of view induces a greater encoding of the stimulus properties. In line with this result, an early review of brain imaging paper on pain perception has suggested that the pattern of activity within different regions is closer to what is found during nociception when the pain is referenced to the self as opposed to another person (Jackson et al., 2006c). This pattern of neural response may be more closely linked to the actual pain experience (Derbyshire, 2000; Jackson et al., 2006c). As mentioned previously, the general somatosensory gating is related to observation of action displaying hands before seeing the painful outcome. Therefore, an alternative hypothesis is that these results might suggest an advantage of 1PP for action understanding that consequently lead to enhanced pain perception.

One interpretation for the finding that less SSSR decrease was found in painful pictures observed in another's person of view is that 3PP involved different cognitive processes (e.g., complex spatial transformations) to mentally rotate the stimuli to an egocentric perspective (van der Heiden et al., 2013). In accordance with this suggestion, Li and Han (2010) reported a decrease of the event-related brain potentials amplitude when changing cognitive perspective during pain observation in the late top-down controlled component but not the early automatic component. Their results indicated that pain observation initially modulates the ERP response whether the participants had to imagine that they were in a painful situation or that an unfamiliar person was in the same painful situation. Cognitive perspective processes later reduces this neural response to observed pain (Li and Han, 2010). The timing of the SSSR variation could be an interesting variable to assess precisely in future research. Altogether, these findings demonstrate that the general process of visual PT is associated with a common somatosensory neuronal response pattern. However, distinct processes are engaged when

one has to evaluate pain situations observed according to the visual perspective i.e., in first- or third-person visual perspective.

Taken together, the present findings support the hypothesis that visual PT yields higher cognitive processes, and modulates the somatosensory neural activity in pain observation. However, no significant results support the relationship between pain intensity ratings and EEG data. Some studies also failed to detect significant statistical correspondence between behavioral and cerebral measures (Danziger et al., 2009; Voisin et al., 2011a). Thus, it is reasonable to suggest that seeing and evaluating pain might engage distinct constructs. Further, this leads to the idea that other behavioral measures, such as response latency, could probably be more related to time-frequency neurophysiological data and should be considered in future work.

Some limitations need to be addressed in this study. First, our sample size was relatively small, reducing the quantity of EEG and behavioral data. To overcome this inconvenience, we used a specific EEG pre-processing that keeps an optimal amount of EEG data for analyses. Second, we did not include neuropsychological tests, so a comparison could not be made for possible interaction between visual perspective abilities and specific cognitive functions (e.g., inhibition). Third, the type of strategy that participants used to evaluate pain intensity was not controlled in the experiment. Someone who evaluates pain intensity based of the visual perspective could give different ratings between 1PP and 3PP, while another person could refer to his personal experience, regardless of the orientation of stimuli. These current limitations should be considered as possible avenues for future research on visual PT.

CONCLUSION

The neuronal and behavioral mechanisms of visual PT were examined in a pain observation paradigm, a widely recognized methodology for the study of different components of empathy (Decety et al., 2006; Fitzgibbon et al., 2010; Lamm et al., 2011). The present results demonstrated that seeing pain from self- or other- visual perspectives produce partly similar neuronal responses, which enable a person to share and to understand another's person pain experience even if it is different from his or her point of view. This study further illustrated that the characteristics of the somatosensory cerebral modulation could differ between self and other's visual perspective. The current study lays the basis for further studies on pain communication where the consideration of different points of view can be influenced by the visual perspective from which a situation is perceived. We also emphasize the relevance for further investigations using a similar experimental paradigm with psychiatric populations, such as schizophrenia, in which general PT deficits are observed (Langdon et al., 2006; Montag et al., 2007; Derntl et al., 2009).

ACKNOWLEDGMENTS

We would like to thank Pierre-Olivier Lauzon and Michel-Pierre Coll for their technical contribution. This study was supported by a Discovery Grant from NSERC, a New Investigator Award from the Brain and Behavior Research Foundation, as well as a Leaders Opportunity Fund from the Canadian Foundation for

Innovation, to Philip L. Jackson. The study was also made possible thanks to scholarships from CRCN awarded to Dora L. Canizales, and salary grants from the FRQS and CIHR to Philip L. Jackson.

REFERENCES

- Aguera, P. E., Jerbi, K., Caclin, A., and Bertrand, O. (2011). ELAN: a software package for analysis and visualization of MEG, EEG and LFP signals. *Comput. Intell. Neurosci.* 2011:158970. doi: 10.1155/2011/158970
- Betti, V., Zappasodi, F., Rossini, P. M., Aglioti, S. M., and Tecchio, F. (2009). Synchronous with your feelings: sensorimotor gamma band and empathy for pain. *J. Neurosci.* 29, 12384–12392. doi: 10.1523/jneurosci.2759-09.2009
- Bufalari, I., Aprile, T., Avenanti, A., Di Russo, F., and Aglioti, S. M. (2007). Empathy for pain and touch in the human somatosensory cortex. *Cereb. Cortex* 17, 2553–2561. doi: 10.1093/cercor/bhl161
- Bundo, M., Kaneoke, Y., Inao, S., Yoshida, J., Nakamura, A., and Kakigi, R. (2000). Human visual motion areas determined individually by magnetoencephalography and 3D magnetic resonance imaging. *Hum. Brain Mapp.* 11, 33–45. doi: 10.1002/1097-0193(200009)11:1<33::aid-hbm30>3.0.co;2-c
- Callan, D. E., Gamez, M., Cassel, D. B., Terzibas, C., Callan, A., Kawato, M., et al. (2012). Dynamic visuomotor transformation involved with remote flying of a plane utilizes the 'Mirror Neuron' system. *PLoS One* 7:e33873. doi: 10.1371/journal.pone.0033873
- Cheng, Y., Yang, C. Y., Lin, C. P., Lee, P. L., and Decety, J. (2008). The perception of pain in others suppresses somatosensory oscillations: a magnetoencephalography study. *Neuroimage* 40, 1833–1840. doi: 10.1016/j.neuroimage.2008.01.064
- Coll, M.-P., Grégoire, M., Latimer, M., Eugène, F., and Jackson, P. L. (2011). Perception of pain in others: implication for caregivers. *Pain Manag.* 1, 257–265. doi: 10.2217/pmt.11.21
- Corradi-Dell'Acqua, C., Hofstetter, C., and Vuilleumier, P. (2011). Felt and seen pain evoke the same local patterns of cortical activity in insular and cingulate cortex. *J. Neurosci.* 31, 17996–18006. doi: 10.1523/jneurosci.2686-11.2011
- Cromwell, H. C., Mears, R. P., Wan, L., and Boutros, N. N. (2008). Sensory gating: a translational effort from basic to clinical science. *Clin. EEG Neurosci.* 39, 69–72. doi: 10.1177/155005940803900209
- Danziger, N., Faillenot, I., and Peyron, R. (2009). Can we share a pain we never felt? Neural correlates of empathy in patients with congenital insensitivity to pain. *Neuron* 61, 203–212. doi: 10.1016/j.neuron.2008.11.023
- de Lussanet, M. H., Fadiga, L., Michels, L., Seitz, R. J., Kleiser, R., and Lappe, M. (2008). Interaction of visual hemifield and body view in biological motion perception. *Eur. J. Neurosci.* 27, 514–522. doi: 10.1111/j.1460-9568.2007.06009.x
- Decety, J., Jackson, P. L., and Brunet, E. (2006). "The cognitive neuropsychology of empathy," in *Empathy in Mental Illness and Health*, eds T. F. Farrow and P. W. Woodruff (New York: Cambridge University Press), 239–260.
- Decety, J., Yang, C. Y., and Cheng, Y. (2010). Physicians down-regulate their pain empathy response: an event-related brain potential study. *Neuroimage* 50, 1676–1682. doi: 10.1016/j.neuroimage.2010.01.025
- Derbyshire, S. W. (2000). Exploring the pain "neuromatrix". *Curr. Rev. Pain* 4, 467–477. doi: 10.1007/s11916-000-0071-x
- Derntl, B., Finkelmeyer, A., Toygar, T. K., Hulsman, A., Schneider, F., Falkenberg, D. I., et al. (2009). Generalized deficit in all core components of empathy in schizophrenia. *Schizophr. Res.* 108, 197–206. doi: 10.1016/j.schres.2008.11.009
- Dodds, C. M., Morein-Zamir, S., and Robbins, T. W. (2011). Dissociating inhibition, attention and response control in the frontoparietal network using functional magnetic resonance imaging. *Cereb. Cortex* 21, 1155–1165. doi: 10.1093/cercor/bhq187
- Dosch, M., Loenneker, T., Bucher, K., Martin, E., and Klaver, P. (2010). Learning to appreciate others: neural development of cognitive perspective taking. *Neuroimage* 50, 837–846. doi: 10.1016/j.neuroimage.2009.12.013
- Fan, Y., Duncan, N. W., de Greck, M., and Northoff, G. (2011). Is there a core neural network in empathy? An fMRI based quantitative meta-analysis. *Neurosci. Biobehav. Rev.* 35, 903–911. doi: 10.1016/j.neubiorev.2010.10.009
- Fitzgibbon, B. M., Giummarra, M. J., Georgiou-Karistianis, N., Enticott, P. G., and Bradshaw, J. L. (2010). Shared pain: from empathy to synaesthesia. *Neurosci. Biobehav. Rev.* 34, 500–512. doi: 10.1016/j.neubiorev.2009.10.007
- Hampshire, A., Chamberlain, S. R., Monti, M. M., Duncan, J., and Owen, A. M. (2010). The role of the right inferior frontal gyrus: inhibition and attentional control. *Neuroimage* 50, 1313–1319. doi: 10.1016/j.neuroimage.2009.12.109
- Han, S., Fan, Y., Xu, X., Qin, J., Wu, B., Wang, X., et al. (2009). Empathic neural responses to others' pain are modulated by emotional contexts. *Hum. Brain Mapp.* 30, 3227–3237. doi: 10.1002/hbm.20742
- Iannetti, G. D., Salomons, T. V., Moayed, M., Mouraux, A., and Davis, K. D. (2013). Broken hearts and broken bones: contrasting mechanisms of social and physical pain. *Trends Cogn. Sci.* 17, 371–378. doi: 10.1016/j.tics.2013.06.002
- Jackson, P. L., Brunet, E., Meltzoff, A. N., and Decety, J. (2006b). Empathy examined through the neural mechanisms involved in imagining how I feel versus how you feel pain. *Neuropsychologia* 44, 752–761. doi: 10.1016/j.neuropsychologia.2005.07.015
- Jackson, P. L., Meltzoff, A. N., and Decety, J. (2005). How do we perceive the pain of others? A window into the neural processes involved in empathy. *Neuroimage* 24, 771–779. doi: 10.1016/j.neuroimage.2004.09.006
- Jackson, P. L., Meltzoff, A. N., and Decety, J. (2006a). Neural circuits involved in imitation and perspective-taking. *Neuroimage* 31, 429–439. doi: 10.1016/j.neuroimage.2005.11.026
- Jackson, P. L., Rainville, P., and Decety, J. (2006c). To what extent do we share the pain of others? Insight from the neural bases of pain empathy. *Pain* 125, 5–9. doi: 10.1016/j.pain.2006.09.013
- Kaiser, S., Walther, S., Nennig, E., Kronmüller, K., Mundt, C., Weisbrod, M., et al. (2008). Gender-specific strategy use and neural correlates in a spatial perspective taking task. *Neuropsychologia* 46, 2524–2531. doi: 10.1016/j.neuropsychologia.2008.04.013
- Kessler, K., and Thomson, L. A. (2010). The embodied nature of spatial perspective taking: embodied transformation versus sensorimotor interference. *Cognition* 114, 72–88. doi: 10.1016/j.cognition.2009.08.015
- Keysers, C., Kaas, J. H., and Gazzola, V. (2010). Somatosensation in social perception. *Nat. Rev. Neurosci.* 11, 417–428. doi: 10.1038/nrn2833
- Kozhevnikov, M., and Hegarty, M. (2001). A dissociation between object manipulation spatial ability and spatial orientation ability. *Mem. Cognit.* 29, 745–756. doi: 10.3758/bf03200477
- Krishnan, A., Woo, C.-W., Ruzic, L., Fan, J., Gu, X., Chang, L., et al. (2013). *Physical and Observed Pain are Represented by Distinct Neural Patterns*. Paper presented at the Annual Meeting of the Organization for Human Brain Mapping, Seattle, WA.
- Lambrey, S., Amorim, M. A., Samson, S., Noulhiane, M., Hasboun, D., Dupont, S., et al. (2008). Distinct visual perspective-taking strategies involve the left and right medial temporal lobe structures differently. *Brain* 131, 523–534. doi: 10.1093/brain/awn317
- Lamm, C., Batson, C. D., and Decety, J. (2007a). The neural substrate of human empathy: effects of perspective-taking and cognitive appraisal. *J. Cogn. Neurosci.* 19, 42–58. doi: 10.1162/jocn.2007.19.1.42
- Lamm, C., Decety, J., and Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *Neuroimage* 54, 2492–2502. doi: 10.1016/j.neuroimage.2010.10.014
- Lamm, C., Nusbaum, H. C., Meltzoff, A. N., and Decety, J. (2007b). What are you feeling? Using functional magnetic resonance imaging to assess the modulation of sensory and affective responses during empathy for pain. *PLoS One* 2:e1292. doi: 10.1371/journal.pone.0001292
- Lamm, C., Porges, E. C., Cacioppo, J. T., and Decety, J. (2008). Perspective taking is associated with specific facial responses during empathy for pain. *Brain Res.* 1227, 153–161. doi: 10.1016/j.brainres.2008.06.066
- Langdon, R., and Coltheart, M. (2001). Visual perspective-taking and schizotypy: evidence for a simulation-based account of mentalizing in normal adults. *Cognition* 82, 1–26. doi: 10.1016/s0010-0277(01)00139-1
- Langdon, R., Coltheart, M., and Ward, P. B. (2006). Empathetic perspective-taking is impaired in schizophrenia: evidence from a study of emotion attribution and theory of mind. *Cogn. Neuropsychiatry* 11, 133–155. doi: 10.1080/1354680044000218
- Li, W., and Han, S. (2010). Perspective taking modulates event-related potentials to perceived pain. *Neurosci. Lett.* 469, 328–332. doi: 10.1016/j.neulet.2009.12.021
- Marcoux, L.-A., Michon, P.-E., Voisin, J. I. A., Lemelin, S., Vachon-Pressseau, E., and Jackson, P. L. (2013). The modulation of somatosensory resonance by psychopathic traits and empathy. *Front. Hum. Neurosci.* 7:724. doi: 10.3389/fnhum.2013.00274
- Montag, C., Heinz, A., Kunz, D., and Gallinat, J. (2007). Self-reported empathic abilities in schizophrenia. *Schizophr. Res.* 92, 85–89. doi: 10.1016/j.schres.2007.01.024

- Montoya, P., and Sitges, C. (2006). Affective modulation of somatosensory-evoked potential elicited by tactile stimulation. *Brain Res.* 1068, 205–212. doi: 10.1016/j.brainres.2005.11.019
- Ruby, P., and Decety, J. (2001). Effect of subjective perspective taking during simulation of action: a PET investigation of agency. *Nat. Neurosci.* 4, 546–550. doi: 10.1038/87510
- Ruby, P., and Decety, J. (2003). What you believe versus what you think they believe: a neuroimaging study of conceptual perspective-taking. *Eur. J. Neurosci.* 17, 2475–2480. doi: 10.1046/j.1460-9568.2003.02673.x
- Ruby, P., and Decety, J. (2004). How would you feel versus how do you think she would feel? A neuroimaging study of perspective-taking with social emotions. *J. Cogn. Neurosci.* 16, 988–999. doi: 10.1162/0898929041502661
- Treede, R. D., Kenshalo, D. R., Gracely, R. H., and Jones, A. K. (1999). The cortical representation of pain. *Pain* 79, 105–111. doi: 10.1016/s0304-3959(98)00184-5
- van der Heiden, L., Scherpiet, S., Konicar, L., Birbaumer, N., and Veit, R. (2013). Inter-individual differences in successful perspective taking during pain perception mediates emotional responsiveness in self and others: an fMRI study. *Neuroimage* 65, 387–394. doi: 10.1016/j.neuroimage.2012.10.003
- Voisin, J. I., Marcoux, L. A., Canizales, D. L., Mercier, C., and Jackson, P. L. (2011a). I am touched by your pain: limb-specific modulation of the cortical response to a tactile stimulation during pain observation. *J. Pain* 12, 1182–1189. doi: 10.1016/j.jpain.2011.06.005
- Voisin, J. I., Mercier, C., Jackson, P. L., Richards, C. L., and Malouin, F. (2011c). Is somatosensory excitability more affected by the perspective or modality content of motor imagery? *Neurosci. Lett.* 493, 33–37. doi: 10.1016/j.neulet.2011.02.015
- Voisin, J. I., Rodrigues, E. C., Hetu, S., Jackson, P. L., Vargas, C. D., Malouin, F., et al. (2011b). Modulation of the response to a somatosensory stimulation of the hand during the observation of manual actions. *Exp. Brain Res.* 208, 11–19. doi: 10.1007/s00221-010-2448-3
- Wager, T. D., Atlas, L. Y., Lindquist, M. A., Roy, M., Woo, C. W., and Kross, E. (2013). An fMRI-based neurologic signature of physical pain. *N. Engl. J. Med.* 368, 1388–1397. doi: 10.1056/NEJMoa1204471
- Yang, C. Y., Decety, J., Lee, S., Chen, C., and Cheng, Y. (2009). Gender differences in the mu rhythm during empathy for pain: an electroencephalographic study. *Brain Res.* 1251, 176–184. doi: 10.1016/j.brainres.2008.11.062

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 13 July 2013; accepted: 21 November 2013; published online: 09 December 2013.

Citation: Canizales DL, Voisin JIA, Michon P-E, Roy M-A and Jackson PL (2013) The influence of visual perspective on the somatosensory steady-state response during pain observation. Front. Hum. Neurosci. 7:849. doi: 10.3389/fnhum.2013.00849

This article was submitted to the journal Frontiers in Human Neuroscience.

Copyright © 2013 Canizales, Voisin, Michon, Roy and Jackson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Perspective taking in language: integrating the spatial and action domains

Madeleine E. L. Beveridge* and Martin J. Pickering

Department of Psychology, University of Edinburgh, Edinburgh, UK

Edited by:

Klaus Kessler, University of Glasgow, UK

Reviewed by:

Anna M. Borghi, University of Bologna and Institute of Cognitive Sciences and Technologies, Italy
Giovanni Pezzulo, National Research Council of Italy, Italy

*Correspondence:

Madeleine E. L. Beveridge,
Department of Psychology, University of Edinburgh, 7 George Square,
Edinburgh, EH8 9JZ, UK
e-mail: m.e.l.beveridge@sms.ed.ac.uk

Language is an inherently social behavior. In this paper, we bring together two research areas that typically occupy distinct sections of the literature: perspective taking in spatial language (whether people represent a scene from their own or a different spatial perspective), and perspective taking in action language (the extent to which they simulate an action as though they were performing that action). First, we note that vocabulary is used inconsistently across the spatial and action domains, and propose a more transparent vocabulary that will allow researchers to integrate action- and spatial-perspective taking. Second, we note that embodied theories of language comprehension often make the narrow assumption that understanding action descriptions involves adopting the perspective of an agent carrying out that action. We argue that comprehenders can adopt embodied action-perspectives other than that of the agent, including those of the patient or an observer. Third, we review evidence showing that perspective taking in spatial language is a flexible process. We argue that the flexibility of spatial-perspective taking provides a means for conversation partners engaged in dialogue to maximize similarity between their situation models. These situation models can then be used as the basis for action language simulations, in which language users adopt a particular action-perspective.

Keywords: embodied cognition, spatial perspective, agency, action perspective, situation models

INTRODUCTION

Over the past decade, research into language comprehension has increasingly been framed in terms of a link between perceptual and motor systems, and higher level cognitive tasks. A central assumption of such *Embodied Cognition* frameworks is that people's understanding of language is grounded in their physical interactions with the world (e.g., Barsalou, 1999, 2008; Pulvermüller, 2005; Fischer and Zwaan, 2008; Glenberg et al., 2008a; Glenberg and Gallese, 2012). In strong versions of Embodied Cognition, language comprehension is achieved through mental representations that correspond, in perceptual or motor qualities, to the object or action being described. Such accounts draw on evidence that comprehenders are faster to correctly match sentences to images that correspond to the perceptual characteristics implied by the sentence context, such as orientation (Stanfield and Zwaan, 2001), shape (Zwaan et al., 2002; Pecher et al., 2009), and implied movement (Kaschak et al., 2005, 2006).

In addition, Action-Sentence Compatibility Effects (ACE; Glenberg and Kaschak, 2002) demonstrate that language comprehension is linked to action execution. Participants are faster to respond to sentences that imply moving the hand away from or towards one's body (e.g., "Close/Open the drawer"), when the direction of response required (away from or towards their body) matches the direction of movement implied in the sentence. Aravena et al. (2010) recently provided evidence of a neural signature for ACE effects by recording event-related (brain) potentials.

In this study, participants listened to sentences implying an open or closed hand shape, and indicated their understanding by responding with either an open or closed hand shape. Incongruent trials, where the hand-shape implied by the sentence did not match the hand-shape required by the response, resulted in an N400 effect (associated with difficulty integrating stimuli into a given semantic context; Kutas and Federmeier, 2000). Such evidence is consistent with the viewpoint that action language comprehension involves representing an action as though you were performing it yourself—that is, from an agent's perspective.

In this paper, we explore research into action-perspective taking (from whose perspective do language users simulate a described action?), and spatial-perspective taking (from whose perspective do language users conceive spatial relations?). We propose that these two forms of perspective taking are fundamentally linked: in order for language users to perform an action simulation, they must first establish a spatial context for that action, by locating it within a situation model. In dialogue, spatial-perspective taking can be used by interlocutors to negotiate or align on situation models that specify similar spatial relations between entities, to ensure a mutually understood spatial context for actions. Actions are performed in space, and, therefore, we might expect considerable cross-over between the literatures on action- and spatial-perspective taking, but this does not appear to be the case. We argue that one reason for this situation is the use of inconsistent and conflicting terminology across the two fields.

Our goal in this paper is to unite action- and spatial-perspective taking in an account of action language comprehension. First, we propose a vocabulary for discussing action-perspective taking that will allow action- and spatial-perspective taking to be integrated. Next, we explore evidence from the Embodied Cognition literature, investigating which action-perspective comprehenders typically adopt. We argue that, contrary to some Embodied Cognition accounts where action-perspective taking is typically assumed to be fixed on the agent, several other perspectives are in fact available. We then review research into which spatial-perspective people tend to adopt in language use, and how such perspective taking is negotiated in dialogue. Finally, we propose the Spatial Grounding Hypothesis, which states that action simulations are grounded in spatial context. We discuss the evidence in favor of this hypothesis, and explore the role of situation models in providing this context.

REPRESENTING OTHER PEOPLE'S ACTIONS

At the same time as theories of action-language processing have stressed the primacy of motor representations, theories of action understanding have argued that the same mental representations are involved in both performing and perceiving actions (e.g., Grèzes and Decety, 2001; Prinz and Hommel, 2002). For example, Common Coding theory (Prinz, 1997; Hommel et al., 2001) proposes that codes for planned actions and perceived actions share a common representational domain. In support of this account, behavioral research suggests first, that participants are less able to perceive a static stimulus (left or right pointing arrow) when performing a congruent action (left or right button press; Müsseler and Hommel, 1997), and second, that perceiving an action while planning an incompatible action affects action execution (Brass et al., 2000; Kilner et al., 2003). In other words, the link between perception and action affects our ability both to *perceive* stimuli, and to *perform* actions. Such findings are echoed by recent neurological research showing evidence of “mirror matching”, where regions of the motor system that are activated when performing an action are also activated when passively perceiving an action (e.g., Buccino et al., 2001; Grèzes et al., 2003; for a review see Rizzolatti and Craighero, 2004).

Much research has argued that the perceiver of an action mentally simulates executing that action herself (Decety, 2002). This *simulation theory* has counterparts in simulation theories of mind that propose that understanding another person involves simulating their mental activity (e.g., Gallese and Goldman, 1998). Indeed, it could be argued that a successful theory of mind is one that allows us to predict and understand our own and other peoples' actions, and that this is achieved through simulation (Ruby and Decety, 2001). The close link between self and other then begs the question: how do we distinguish our own actions or mental activities from those of other people? The ability to distinguish ourselves from other people is critical to successful social interaction, but in a system in which our own actions share representations with the actions of other people, action attribution becomes a key computational problem (Decety and Sommerville, 2003; de Vignemont and Haggard, 2008).

The mechanism by which the separation of self and other is maintained is beyond the scope of this paper (see, for example,

Ruby and Decety, 2001, 2004; Decety and Sommerville, 2003). But however it is achieved, the self-other distinction is tightly connected with perspective taking. First, self must be successfully distinguished from other in order for there to be the possibility of different perspectives (Jeannerod, 2006). Second, the ability to represent other people's actions in a similar way to their own allows people to take an agent's perspective on an action, even when they are describing or hearing about an action performed by somebody else.

A TAXONOMY OF PERSPECTIVE

As highlighted above, a large body of research now suggests a link between language processing and sensorimotor activation (see Kiefer and Pulvermüller, 2012; Meteyard et al., 2012 for recent reviews). This link can best be captured by Embodied Cognition accounts of language processing.¹ Embodied Cognition seeks to distinguish itself from “traditional” psycholinguistic accounts by insisting that language representations are modal rather than amodal (e.g., Zwaan and Taylor, 2006; Barsalou, 2008). What is often not made explicit in Embodied Cognition accounts is that modal representations are inherently perspective-based. For a representation to be modal, it must assume a given perspective. In other words, the perspective is necessary to ground the representation. However, discussion of perspective taking in action language is often opaque, and this is particularly problematic if we wish to relate action-perspective taking and spatial-perspective taking.

In visual cognition, researchers distinguish between two types of spatial-perspective taking. Level 1 perspective involves understanding what falls within another individual's line of sight—for example, is a particular object occluded by another object as that person looks at it? Level 2 perspective involves understanding how the world appears from another person's perspective—for example, is a particular object to the left or the right of another object as that person looks at it? (Flavell et al., 1981; Michelon and Zacks, 2006). In the present paper, we limit our review of spatial-perspective to this second level, focusing on spatial relations, rather than visibility. Kessler and Rutherford (2010) argued that Level 2, but not Level 1 spatial-perspective taking, appears to involve some form of covert mental rotation or simulation. As such, Level 2 spatial-perspective entails a level of embodiment that Level 1 does not, and is therefore closer to the perspective-bound simulations proposed by Embodied Cognition accounts of action-language understanding.

With respect to Level 2 spatial-perspective taking, we can contrast intrinsic, absolute, and relative reference frames (see Levinson, 1996, 2003). In an intrinsic reference frame, the position of an object is described relative to a reference object (e.g., “The window is above the door”). In an absolute reference frame, the position of an object is described in terms of stable environmental features, such as points of the compass, as in “The

¹Note our use of the term “embodied” refers specifically to Embodied (or “grounded”) Cognition accounts of language, not to “embodied” versus “disembodied” perspectives, as sometimes discussed in the spatial perspective literature (e.g., Tversky and Hard, 2009).

ship is south of the island". Neither of these reference frames locates an object relative to an observer. A relative reference frame, on the other hand, does just that: for example, "The car is to my left". Within a relative reference frame, one can adopt an *egocentric* or *allocentric* perspective. An *egocentric* perspective entails representing objects in a scene from your own viewpoint, and an *allocentric* perspective entails representing objects from the viewpoint of someone other than yourself (see Levinson, 2003 for a fuller treatment of spatial reference frames). The terms *egocentric* and *allocentric* therefore have specific and well-established meanings in the spatial literature: *egocentric* means conceptualizing space from your own point of view, and *allocentric* means conceptualizing space from another's point of view. In the literature on Embodied Cognition, however, researchers often use *egocentric* to refer to putting oneself in someone else's shoes (for example, interpreting a sentence such as "John kicked Mary" as though the comprehender herself were performing the act of kicking; e.g., Willems et al., 2010). This use of the term is opposite that in spatial-perspective taking and is therefore confusing. In addition, using the term *egocentric* perspective in action language, or *allocentric* perspective in spatial language, does not specify *whose* shoes the comprehender is putting herself into. In spatial language, this underspecification is typically not problematic, since the perspective adopted in a sentence such as "John is looking at the picture on the left" can be explicitly clarified. The comprehender can legitimately ask "on whose left?", and the speaker can reply "on *my* left", "on *your* left", "on *his* left", etc. However, in action language, perspective-taking is implicit, rather than explicit, and no such clarification is possible. For example, a comprehender who responded to the sentence "John is looking at the picture on the left", with the query "who is looking?" would receive the reply "John", and remain no clearer about whose perspective the speaker was adopting. Therefore, unlike spatial language, when discussing action language it is necessary for embodied accounts to specify whose perspective is being adopted for a particular action: the term *egocentric* perspective tells us that comprehenders are putting themselves in somebody else's shoes, but crucially not whose shoes. Similarly, researchers often speak of "situated simulations" (Marino et al., 2012), or "sensorimotor experience" (Pecher et al., 2009) without specifying from whose perspective this simulation or resonance occurs. We suggest that this lack of specification derives from a widely held assumption in embodied cognition accounts that the agent's perspective is adopted. However, we also suggest that this assumption is unwarranted.

There are in fact different Embodied Cognition accounts of language processing, and researchers in this field place varying importance on the role of sensorimotor processing in semantics (see Meteyard et al., 2012 for a recent review of positions advocating different degrees of embodiment). However, a prevailing view conceives language comprehension as an internal simulation of the described action, as if the comprehender were performing that action herself (e.g., Barsalou, 1999; Zwaan and Taylor, 2006; Borghi and Scorolli, 2009; Bergen and Wheeler, 2010). If it is true that action-perspective taking is fixed on the agent's perspective, then the underspecification of *egocentric*, outlined above, is not a problem; the perspective adopted would

always coincide with the agent of the described action. However, as we shall see, it is not clear that an agent's perspective is always adopted. Researchers in action language therefore need to make clear exactly whose perspective they assume is being adopted.

For example, in understanding "John kicked Mary", there are at least two embodied perspectives that could be adopted for the action of kicking: that of John (the embodied agent); and that of Mary (the embodied patient). If the comprehender has reason to believe that other people are witnesses to the event (i.e., if she has reason to include bystanders in her situation model), then she can also adopt the perspective of a bystander watching the kicking event unfold (the embodied observer). For example, if a previous sentence implied the existence of a crowd gathering around Mary and John, the comprehender can adopt the perspective of a member of this crowd, observing John kicking Mary. In each case, the comprehender represents the action from the perspective of a person present in the comprehender's model of that event. In taking the embodied agent's perspective, the comprehender represents the action of kicking as though she herself were the agent of that action, by activating the same systems involved in executing a kicking action. In taking the embodied patient's perspective, the comprehender represents the action of kicking as though she herself were the patient of that action (presumably activating some form of empathic response to the pain, such as wincing). In taking the embodied observer's perspective, the comprehender represents that action as though she were watching it unfold, by activating the same systems that would be recruited when observing such an action. In addition to these embodied perspectives, there is another perspective that the comprehender could take: that of the non-embodied observer. Unlike an embodied participant or observer, the non-embodied observer represents the action without running a simulation from any particular point of view. We propose that action-perspective taking is grounded in spatial context (see section Situation Models: Linking Spatial- and Action-Perspectives); comprehenders will run an action simulation wherever possible, but if there is insufficient spatial context to simulate the action from a particular perspective, comprehenders will adopt the non-embodied observer's perspective instead.

The sentence "John kicked Mary" refers to a transitive event with two participants. There are of course, more complex sentences in which further embodied perspectives exist. This is the case for sentences describing ditransitive events (e.g., "John passed the child to his wife"), or sentences where a thematic role is occupied by more than one entity (e.g., "John kicked Mary and Sam"). The number of potential embodied perspectives available for a given sentence is therefore the number of participants in that event plus that any embodied observers licensed by the comprehender's situation model. We propose that these perspectives (e.g., embodied agent, embodied patient, embodied recipient, plus embodied observer and non-embodied observer) provide a transparent basis for discussing action perspective taking. Using these terms, researchers can not only distinguish between embodied and non-embodied representations, but within the embodied representations, it is possible to distinguish whose perspective is adopted.

DO LANGUAGE USERS CONSISTENTLY ADOPT THE AGENT'S PERSPECTIVE?

We noted above that many embodied accounts of language assume that if a perspective is adopted for action language, it is the agent's perspective (e.g., Glenberg and Kaschak, 2002; Zwaan and Taylor, 2006; Wu and Barsalou, 2009). Such an assumption is consistent with results from studies using isolated action verbs, for example, showing somatotopic activation for specific body parts. Research using functional magnetic resonance imaging (fMRI) has found that passive listening to an arm-word ("pick") leads to increased activation in areas of the premotor and primary motor cortex associated with arm movements; passive listening to a face-word ("lick") leads to increased activation in areas associated with the face; and passive listening to a foot-word ("kick") lead to increased activation in areas associated with the feet (Hauk et al., 2004; see also Aziz-Zadeh et al., 2006). In other words, the activation appears to be associated with particular acts from the perspective of the agent of the act (e.g., the kicker) rather than (for example) the patient (e.g., the person or thing that is kicked). Further work using magnetoencephalography (MEG) has demonstrated that such somatotopic activation occurs extremely quickly, within 200 ms of word presentation, and even when participants are concentrating on an unrelated, non-language based task (Pulvermüller et al., 2005). These findings suggest that adopting an embodied agent's perspective may occur automatically in the early stages of semantic processing, at least in isolated words.²

More evidence that people adopt the embodied agent's perspective (as though the comprehender herself were carrying out an action) comes from evidence for "body-specific" representations of manual action verbs (e.g., *throw*) in a Dutch lexical decision task (Willems et al., 2010). Left-handed participants showed activation in the right pre-motor hand area, but right-handed participants showed activation in the left pre-motor hand area, despite there being no manual responses on critical trials. These results echo findings of "body-specific" activation for motor imagery, where left- and right-handed participants imagined performing actions described by manual action verbs (Willems et al., 2009). It therefore appears that people tend to adopt the embodied agent's perspective for isolated verbs, representing the verb according to how they personally would perform those actions with their particular bodies (i.e., right-handed for right-handed participants; left-handed for left-handed participants).

However, verbs are usually processed not in isolation, but in the context of sentences featuring noun phrases that refer to particular entities. Do language users also adopt an embodied agent's perspective in action sentences, as well as isolated verbs?

The evidence that they do is mixed. Participants undergoing fMRI were presented with mouth-, leg-, or hand-related action sentences featuring the pronoun ("I") in the agent's role (e.g., "Mordo la mela" [I bite the apple]; "Afferro il coltello" [I grasp the knife]; "Calcio il pallone" [I kick the ball]; Tettamanti et al., 2005). The results showed evidence of somatotopic activation similar to that observed in isolated verb processing (e.g., Hauk et al., 2004), implying that participants were simulating the described actions from the agent's perspective. However, in this study, the agent's perspective coincided with the perspective of the potentially self-referential pronoun "I": participants may have adopted a perspective in line with the thematic role assigned to the pronoun "I", rather than the perspective of the agent *per se*. A better indication of whether participants routinely adopt the embodied agent's perspective comes from studies investigating ACE effects (Glenberg and Kaschak, 2002; Glenberg et al., 2008b). When sentences were given in the form of an imperative (e.g., "Close the drawer"), participants were faster to respond when the direction of the response was congruent with the movement implied by the agent in the sentence than when it was incongruent. In other words, they appeared to adopt the perspective of an agent closing a drawer. However, in sentences featuring two arguments, one of whom could refer to the participant, participants were faster to respond when the direction of the response was congruent with the movement relative to the pronoun "you". For example, participants were faster to respond with away movements to sentences such as "You delivered the pizza to Andy", but faster to respond with towards movements to sentences such as "Andy delivered the pizza to you". Therefore, this suggests that when a sentence involves a potentially self-referential pronoun ("you", "I"), comprehenders tend to adopt the perspective of the thematic role assigned to that pronoun, whether or not this coincides with the thematic agent of the action. In a dialogue context, where sentences such as "You are / I am cutting the tomato" are uttered and understood by each participant in turn, the situation is more complex. Participants appear to prioritize adopting opposing perspectives for "you" and "I", over maintaining a consistent perspective (e.g., embodied agent, embodied observer) for either of the pronouns (Pickering et al., 2012).

Several studies have addressed whether people adopt the agent's perspective when the agent of a described action is not self-referential, in the absence of a second self-referential argument. In Embodied Cognition accounts that conceive action language as an extension of mirror-matching, where representations of other people's actions are inherently similar to representations of one's own actions (e.g., Rizzolatti and Arbib, 1998; Pulvermüller, 2005), descriptions of actions performed by third-person agents should elicit similar effects to descriptions of actions performed by first- or second-person agents. In line with this prediction, Buccino et al. (2005) used transcranial magnetic stimulation (TMS) to stimulate the left-hemispheric hand or foot motor areas, as participants listened to third person hand- or foot-related action sentences (e.g., "Cuciva la gonna" [He sewed the shirt]; "Marciaiva sul posto" [He marched on the spot]), compared with control abstract sentences (e.g., "Amava la moglie" [He loved his wife]). Motor evoked potentials (MEPs) from the hand and foot muscles were recorded. Hand MEPs were

²A general note of caution is needed when interpreting studies that show similar activation in action execution and action language comprehension. These studies are typically cited as evidence that during language comprehension, participants simulate performing the action (in our terminology, they adopt an embodied agent's perspective). However, research into mirror-matching suggests that observing and executing an action also activate similar neural substrates (e.g., Grèzes et al., 2003; for reviews, see Decety and Sommerville, 2003; Rizzolatti and Craighero, 2004). Therefore, it is possible that activation in motor areas during language comprehension in fact reflects the participant mentally "observing", rather than "executing", the described action.

modulated specifically when listening to hand-related action sentences, and foot MEPs were modulated specifically when listening to foot-related sentences. These results suggest at least some tendency to adopt an embodied agent's perspective for third-person sentences.

However, without a direct comparison between first- and third-person sentences, we cannot know whether action perspective-taking in third-person sentences matches action perspective-taking in first-person sentences. Behavioral evidence suggests that comprehenders reading self-referential and non-self-referential sentences adopt different action-perspectives. Brunyé et al. (2009) used a sentence-picture matching task with first-, second-, and third-person action sentences, and "internal" or "external" action images. In the "internal" images, the position of the hands meant they could plausibly be interpreted as those of the participant. In the "external" images, the position of the hands meant they could not plausibly be interpreted as those of the participant. Instead, they could most plausibly be interpreted as those of an agent who the participant was observing perform the action. Selecting an internal image would imply adopting the embodied agent's perspective. Selecting an external image would imply adopting the perspective of an embodied observer. Brunyé et al. (2009) found that participants were faster to correctly match first- and second-person sentences to internal rather than external images, and to correctly match third-person sentences to external rather than internal images. In other words, participants adopted the embodied agent's perspective when the agent of the sentence could be attributed to the comprehender, but not otherwise (see also Ditman et al., 2010; Sato and Bergen, 2013). In an fMRI study, Tomasino et al. (2007) found no difference in primary motor cortex activation between silent reading of German action phrases presented in the first-person (e.g., "Ich hämmere" [I hammer]) versus third-person (e.g., "Er hämmert" [He hammers]). However, Papeo et al. (2011) had participants silently read action or non-action Italian verbs conjugated in the first- or third-person (e.g., "Scrivo" [I write]; "Scrive" [he writes]; "Medito" [I wonder]; "Medita" [he wonders]). They found that TMS-induced MEPs in the relevant motor area (e.g., hand) increased for the first-person action verbs, but that the third-person action verbs behaved like the non-action verbs, and showed no increase in MEPs. Embodied Cognition accounts need not predict total parity between first- and third-person action representations. However, the posited involvement of the motor system in action language comprehension (e.g., Fischer and Zwaan, 2008) should imply at least some difference between third-person action and non-action verbs. The fact that a difference between action and non-action verbs was found only in first-person sentences led Papeo et al. (2011) to conclude that motor simulation of an action sentence occurs only when the self is identified as the agent of the action.

What could be behind the conflicting results of Tomasino et al. (2007), and Papeo et al. (2011)? One important difference may be in the task. Participants in Tomasino et al.'s study were asked to decide whether a described event took place inside or outside a building, and thus could complete the task without paying attention to whether the verb was presented in the first- or third-person. On the other hand, Papeo et al. instructed participants

to determine the syntactic subject of a phrase, thus focussing attention on the contrast between first- and third-person agents. Researchers are becoming increasingly aware of the role of task demands and context in studies of Embodied Cognition. The conflicting results here add to evidence suggesting that motor representations of action language may not be activated automatically, but depend on aspects of the task, including depth of processing (Sato et al., 2008), sentence tense (Bergen and Wheeler, 2010), and relevance to task goals (Hoedemaker and Gordon, 2013). Indeed, it is possible to view the emphasis on the agent's perspective in action-language research, as a result of task demands. The link between action and language has typically been investigated by studying congruency effects when participants execute actions during sentence processing (Zwaan and Taylor, 2006; Taylor and Zwaan, 2008), after sentence processing (Glenberg and Kaschak, 2002; Glenberg et al., 2008b), or before sentence processing (Glenberg et al., 2008a). When the emphasis of the task is to execute an action, it is perhaps not surprising that results seem to indicate that participants adopt the agent perspective. Other paradigms in embodied approaches to language follow sentence processing with image presentation rather than action execution. For example, participants are typically faster and more accurate to recognize an image of an object when it is presented in the same orientation (vertical/horizontal) as implied by the preceding sentence (Stanfield and Zwaan, 2001; see also Zwaan et al., 2002; Pecher et al., 2009). The authors interpret these findings as evidence that comprehenders run visual simulations of an event (i.e., they adopt an embodied observer's perspective). The perspective adopted by comprehender may therefore depend on the task used to investigate it. It may even be possible to use the task to prime participants to adopt a given action-perspective, although we know of no study that has investigated this possibility.

In summary, some Embodied Cognition accounts of action language assume that people adopt an embodied agent's perspective when comprehending action language, based on an internal simulation of performing that action (Zwaan and Taylor, 2006; Barsalou, 2009). Moreover, strong Embodied Cognition accounts assume that the agent's perspective is automatically activated, regardless of contextual factors such as the reference of the sentence, as determined, for example, by the subject pronoun (Pulvermüller, 2005; Pulvermüller et al., 2005). The evidence outlined above suggests that people do adopt the embodied agent's perspective for isolated verbs, and for sentences in which a potentially self-referential pronoun ("you", "I") is specified as the agent (Hauk et al., 2004; Pulvermüller et al., 2005; Willems et al., 2010). However, when a self-referential pronoun occupies a thematic role other than agent, comprehenders appear to adopt the perspective of the thematic role assigned to that pronoun, and not the perspective of the agent (Glenberg and Kaschak, 2002). When a third party is specified as the agent of an action, and no self-referential pronoun is present, some evidence suggests that comprehenders adopt the embodied agent's perspective (Buccino et al., 2005; Tomasino et al., 2007), whereas other evidence suggests that people adopt an embodied observer's perspective (Brunyé et al., 2009; Papeo et al., 2011). Although more data are clearly needed in order to draw firm conclusions about which perspective comprehenders adopt under which circumstances,

current data demonstrate that adopting an agent's perspective is not the only possibility during action language comprehension. As a consequence, the underspecified terms *egocentric* or *internal* perspective should be avoided when discussing action-perspective taking. Instead, researchers in Embodied Cognition should seek to employ more transparent terms that specify *in whose shoes* the comprehender is placing herself (e.g., *embodied agent*, *embodied patient*, *embodied observer*).

SPATIAL-PERSPECTIVE TAKING

So far, we have reviewed evidence examining whose action-perspective language users tend to adopt when processing action language sentences. However, language users can also adopt a range of spatial-perspectives during language production or comprehension. Of particular interest is whether people adopt an egocentric spatial-perspective (conceiving spatial relations from their own point of view), or an allocentric spatial-perspective (conceiving spatial relations from another's point of view).

Schober (1993) asked participants to describe the location of objects, either alone, to an imaginary addressee, or when in the same room as a conversational partner. Participants were more likely to describe the location from the addressee's point of view, using terms such "on your left", than from their own point of view. Schober (1995) also found that speakers tended to adopt the addressee's perspective in task requiring the speaker to identify particular objects to an addressee. Interestingly, participants in Schober (1993) who described objects to an imaginary addressee were *more* likely to use the addressee's perspective than participants whose conversation partners were present. With an addressee absent and unable to provide feedback, it may be safer for the speaker to assume the addressee's perspective as often as possible. Duran et al. (2011), using a virtual reality paradigm, also found that participants were more likely to adopt an allocentric spatial perspective when told that they were interacting with a virtual, rather than real partner. It appears that believing that their partner was real allowed participants to shift more of the burden of mutual comprehension to their partner. The tendency to shift responsibility for effective communication to a conversation partner may be stronger when, as in Duran et al.'s (2011) study, that partner is making a request rather than providing information. Yoon et al. (2012) found that speakers in a modified referential communication task were more likely to use allocentric perspective when requesting something from their partner compared with giving information to their partner. Since it is in speakers' interests to ensure that their requests are successfully understood, it is sensible for listeners to assume that speakers will adopt an allocentric perspective when making that request.

The above results show that spatial-perspective taking, like action-perspective taking, is a flexible process. By changing the perspective they adopt, speakers or listeners can shift more or less of the burden of mutual comprehension on to their partner. Further research suggests that during dialogue, people may attempt to minimize not only their own effort, but the collective effort of both conversation partners, by obeying what Clark and Wilkes-Gibbs (1986) term the principle of *least collaborative effort*. Speakers and listeners often appear to adopt spatial perspectives in a way that maximizes the resources available. The principle

of least collaborative effort appears to be adopted especially in cases where one partner is judged less able to complete the communication task (Schober and Brennan, 2003). For example, Mainwaring et al. (2009) found that speakers were more likely to use an (allocentric) addressee's perspective when the addressee was under increased cognitive load. Schober (2009) studied what happens when, unbeknownst to the participants, one partner in a conversation has better spatial ability than another, as determined by mental rotation test results. Participants were paired into a director and a matcher, with no knowledge of their own or their partner's results on the mental rotation tests. The matcher selected a target circle from an array, based on the director's spatial descriptions. Low-ability directors were more likely to take their own (egocentric) perspective, while high-ability directors were more likely to take their partner's (allocentric) perspective. Over the course of the experiment, high-ability directors who were paired with low-ability matchers increased their use of allocentric perspective, whereas low-ability directors who were paired with high-ability matchers decreased their use of allocentric perspective. Note that these opposite patterns of behavior between high- and low-ability directors is in itself reason to be cautious of basing our understanding of spatial perspective-taking in language on university students of (presumably) high cognitive ability.

We argue that this online adaptation to a partner's ability to engage in the communicative task is compatible with conversation as conceived as a joint action (Clark, 1996; Sebanz et al., 2006; Gambi and Pickering, 2011). In the case of spatial perspective-taking, the perspective that people adopt appears to depend at least partly on the ability of their partner to engage in the task. In the next section, we argue that maximising the collective resources in this way allows conversation partners to establish coherent situation models in both partners. Once these situation models have been established, language users are in a position to adopt a particular action-perspective when performing mental simulations of actions. However, interlocutors do not adapt only their use of spatial-perspective within a relative reference frame; they also appear to adapt their choice of reference frame itself. Evidence that conversation partners align on their use of reference frame comes from studies using a confederate-priming paradigm. Watson et al. (2004) studied participants' use of an intrinsic versus a relative reference frame. Participants were more likely to use an intrinsic reference frame after the confederate had used an intrinsic frame than after the confederate had used a relative reference frame. Importantly, Watson et al. found participants regularly switched between reference frames. Spatial-perspective taking in dialogue is therefore highly flexible in order to allow for maximal alignment and hence maximal similarity in situation models. Whether such alignment on situation models occurs as a result of automatic priming (e.g., Pickering and Garrod, 2004, 2006), or of negotiating common ground (e.g., Clark, 1996) is beyond the scope of this paper, but we assume both possibilities remain open.

SITUATION MODELS: LINKING SPATIAL- AND ACTION-PERSPECTIVES

Much research on Embodied Cognition can be traced back to studies of situation models in language processing (e.g., Johnson-Laird, 1983; Van Dijk and Kintsch, 1983). According to recent

accounts, situation models are representations of specific situations described in language, where events are connected along five dimensions: space, time, protagonist, causality, and intentionality (Zwaan et al., 1995; for a review of situation models in language, see Zwaan and Radvansky, 1998). Evidence suggests it is the content of these models, rather than linguistic form of the language itself, which is typically retained in memory and integrated into updated models as comprehension continues (Sachs, 1967; Johnson-Laird and Stevenson, 1970). For example, Bransford et al. (1972) demonstrated that participants who read the sentence “Three turtles rested on a floating log, and a fish swam beneath them” frequently selected the linguistically different but situationally equivalent sentence “Three turtles rested on a floating log, and a fish swam beneath it” in a recognition test (see also Barclay, 1973; Honeck, 1973). Many modern studies in the Embodied Cognition literature have found similar effects when the focus is shifted to online rather than memory processes. For example, Borghi et al. (2004) found that participants were faster to verify items typically found inside a given object (e.g., “steering wheel”) following a preamble placing them inside that same object (e.g., “You are driving a car”) versus outside it (e.g., “You are refuelling a car”). They proposed that participants used a mental simulation grounded in modal representations (e.g., of being inside or outside a car), which then guides property verification (see also Kosslyn et al., 1978).

Such mental simulations are a defining feature of embodied theories of language, and differ from the situation models discussed in text or discourse processing in that they appear to capture online processing during language comprehension. Whereas situation models represent the integration of knowledge about events and situations into a coherent, existing framework, mental simulations are concerned with the online action-perspective taking about a particular act (see also Zwaan, 2008 for discussion of the differences). We propose that this “nesting” of action simulations within situation models is what links spatial- and action-perspective taking in language. In order for a comprehender to adopt an embodied perspective on an action, that action must be grounded in a spatial context. This spatial context is provided by the comprehender’s situation model. Situation models are conceived from a particular spatial perspective; in dialogue, conversation partners maximize their resources and align on spatial-perspective and reference frames, in order to ensure suitably similar situation models, for example by making use of the principle of least collaborative effort (Clark, 1996). Recall that situation models can specify events across a number of dimensions (space, time, causality, etc.; Zwaan et al., 1995). For our purposes, “suitably similar” situation models means that the situation models of both interlocutors specify the same protagonists in roughly the same spatial relations to one another.

The spatial relations between objects and people are a fundamental part of situation models (Tversky, 1991), and might be specified at various levels of granularity, from coarse grained, specifying only overall direction, to fine grained, specifying exact distances. We propose that the minimum information required in a situation model in order to run an action simulation is the participants in that action and some (coarse-grained) information about the spatial relations in which they stand. This

allows comprehenders to establish the direction and perhaps rough distance in which an action occurs, and thus to simulate it, adopting a particular action-perspective. When a sentence is interpreted self-referentially (because it involves pronouns such as “you” or “I”—and perhaps also, although we know of no study demonstrating this—when it refers to the comprehender by name), the comprehender creates a situation model grounded in his or her own body; other participants in the action are by default conceived as located in front of the comprehender. For example, in Glenberg and Kaschak (2002), sentences such as “You delivered the pizza to Andy” elicited ACE effects because the direction of an action could be established (away from the comprehender’s body), and an action-perspective could be adopted in line with the thematic role assigned to the self-referential pronoun (embodied agent). We refer to the idea that spatial context grounds action-perspective taking as the Spatial Grounding Hypothesis.

The Spatial Grounding Hypothesis can explain the diverging results we discussed earlier regarding first-person and third-person language. Recall that Papeo et al. (2011) found that comprehenders appeared to adopt an embodied agent’s perspective for first-person language, but no embodied perspective for third-person language; whereas the results of Tomasino et al. (2007) suggested that first- and third-person language elicited similar action perspectives. The Spatial Grounding Hypothesis explains these results as follows. In Papeo’s study, the first-person sentences ground the situation model in the comprehender’s own body, allowing an action simulation to occur; in the third-person sentences, the situation model contains insufficient spatial information for action simulation. In Tomasino et al.’s (2007) study, the task was to decide whether the described action took place inside or outside, thus encouraging the construction of situation models in which to situate first- and third-person actions. Task demands may therefore play an important role in action language understanding, in the extent to which they provide, or encourage participants to create, spatial context for the described actions.

For example, third-person sentences in which the direction of the described action (e.g., turning a knob clockwise or anti-clockwise) is apparent from the sentence context (e.g., raising or lowering the volume) also elicit ACE-type effects where the comprehender adopts an embodied agent’s perspective (Zwaan and Taylor, 2006). Further work suggests that these effects only occur once the direction of movement (clockwise or anti-clockwise) has been specified (Taylor et al., 2008). On the other hand, some evidence suggests that where a described action lacks suitable spatial grounding—for example, when it is described in the third-person, and the spatial relations between participants are not specified—action-perspective taking does not occur. Gianelli et al. (2011) replicated the ACE effects in sentences featuring second-person agents (e.g., “You gave a pizza to Louis”), but not third-person agents (e.g., “Lea gave a pizza to Louis”). When avatars provided spatial locations for the third-person agents, the ACE effect reappeared. In other words, participants only adopted an embodied agent’s action-perspective when their situation model afforded adequate spatial context.

We have suggested that spatial context grounds action-perspective taking, such that a comprehender can only simulate an action from a particular perspective if her situation model

specifies the participants in that action, and their spatial relations (thus giving her access to the direction in which an action would occur). We have argued that this proposal, the Spatial Grounding Hypothesis, can incorporate apparently conflicting results about action-perspective taking into a coherent framework. But there are other factors that support the Spatial-Grounding Hypothesis. First, it predicts that conversation partners will align on spatial-perspective and choice of reference frame, in order to establish similar situation models in both partners. We saw in the previous section that this is indeed the case. Second, it can explain why the presence of a potential agent other than the speaker affects how likely the speaker is to shift her spatial perspective. Tversky and Hard (2009) investigated the influence of a potential agent on how likely people were to adopt an allocentric perspective. Participants viewed photographs of scenes in which an actor was reaching for objects (and thus, in a position to act on that object), scenes with no actor, and scenes with an actor who was not reaching. Participants were more likely to adopt an allocentric spatial perspective (that of the actor in the photograph) when the actor was reaching versus not reaching for an object. Similarly, Zwickel (2009) investigated what spatial-perspective participants adopted when watching clips of animated triangles that they perceived as more or less agentive (Abell et al., 2000). Zwickel provided some evidence that participants only adopt an allocentric perspective when they view the other entity as an agent with specific states of mind, rather than a non-agentive entity moving at random. Mazzarella et al. (2012) recently extended Tversky and Hard's (2009) study by manipulating the extent to which the actor was in a position to act on the object (grasping versus gazing). Images in which the actor was in a better position to act on the object (grasping) triggered more use of allocentric spatial perspective in participants compared with images in which the actor was in a less good position to act on the object (gazing). All of this suggests that participants are more likely to adopt an allocentric spatial-perspective in the presence of someone they perceive as a potential agent.

On the other hand, research suggests that the ability to extract information useful for object interaction (e.g., size) is diminished when participants adopt an allocentric, rather than egocentric spatial-perspective (Campanella et al., 2011). In addition, participants are faster to execute a reach-to-grasp movement when the object also falls within the peripersonal, rather than extrapersonal, space of a second person, implying that people tend to be faster to interact with objects in the presence of another potential agent (Gianelli et al., 2013). Given that participants want to interact with objects more quickly in the presence of another potential agent, and given that adopting an allocentric perspective may impede their ability to do so, why, then, would participants be more likely to adopt an allocentric perspective in the presence of another potential agent? Tversky and Hard (2009) suggested that their participants, in order to make sense of the scene, tried to understand the possibility that the other person can interact with the objects. We propose that people find it easier to understand another person's potential actions when they understand the spatial relations in the other person's situation model; that is, when they conceive space from that person's perspective. Spatial-perspective taking can therefore augment a situation model by

increasing awareness of an agent's *potential* actions, even when no action is described.

One argument against the Spatial Grounding Hypothesis is that that situation models are often underspecified, and do not provide comprehenders with the necessary spatial context in which to simulate action simulations. In particular, isolated verbs provide no explicit spatial context, and yet evidence suggests that comprehenders do adopt an embodied agent's perspective on the actions that the verbs describe (e.g., Hauk et al., 2004; Willems et al., 2010). We suggest that participants typically interpret these isolated verbs as self-referential (even when they are not presented in the imperative). Thus, like explicitly self-referential language, the comprehender's own body grounds her situation model in this case. In other cases, where the comprehender's situation model does not allow her to establish at least the coarsely-coded spatial relations involved in an action, she cannot adopt an embodied action-perspective, because the action simulation cannot be run. However, this does not mean that the sentence describing an action cannot be understood. Rather, the comprehender can adopt the perspective of a non-embodied observer. This perspective is not an embodied perspective, in the sense that it does not involve a simulation of the action from the perspective of any of the participants. However, it is sufficient to allow the comprehender to understand the sentence, even if that understanding is somewhat less fully specified than the situation in which an embodied action-perspective can be adopted. Researchers have found that non-ice hockey players respond more slowly and show less pre-motor activation than expert ice hockey players do when reading sentences about ice hockey (Beilock et al., 2008), but this does not mean that they fail to understand the sentences. Their understanding may be impoverished relative to that of the expert players, but comprehension is not an all or nothing process (Taylor and Zwaan, 2013). Just as non-expert players may supplement their understanding of ice hockey using information and inferences about similar experiences (e.g., playing field hockey), comprehenders with inadequate situation models may supplement their models by adopting a non-embodied observer's perspective based on memories or inferences about similar situations.

CONCLUSIONS

In this paper, we have attempted to reconcile two largely distinct literatures concerned with spatial-perspective taking and action-perspective taking. We have proposed a transparent vocabulary for action-perspective taking, which we hope will facilitate research between these two domains. At the heart of our proposal is the suggestion that researchers working in Embodied Cognition must specify *from whose perspective* a given action is being simulated. Although an agent's perspective seems in many cases the most natural candidate, other perspectives are possible, and are often adopted when self-referential pronouns are assigned a thematic role other than agent.

We have argued that comprehenders can only adopt an action-perspective if they have a spatial context for that action (the Spatial Grounding Hypothesis). In the case of isolated verbs and self-referential pronouns, people typically take their spatial grounding from their own bodies. But in the absence

of self-referential language, action-perspective taking can only occur when the spatial relations between participants in the action have been established within the comprehender's situation

REFERENCES

- Abell, F., Happé, F., and Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cogn. Dev.* 15, 1–16. doi: 10.1016/s0885-2014(00)00014-9
- Aravena, P., Hurtado, E., Riveros, R., Cardona, J. F., Manes, F., and Ibáñez, A. (2010). Applauding with closed hands: neural signature of action-sentence compatibility effects. *PLoS One* 5:e11751. doi: 10.1371/journal.pone.0011751
- Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., and Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Curr. Biol.* 16, 1818–1823. doi: 10.1016/j.cub.2006.07.060
- Barclay, J. R. (1973). The role of comprehension in remembering. *Cogn. Psychol.* 254, 229–254. doi: 10.1016/0010-0285(73)90013-3
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–609. doi: 10.1017/s0140525x99002149
- Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645. doi: 10.1146/annurev.psych.59.103006.093639
- Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1281–1289. doi: 10.1098/rstb.2008.0319
- Beilock, S. L., Lyons, I. M., Mattarella-Micke, A., Nusbaum, H. C., and Small, S. L. (2008). Sports experience changes the neural processing of action language. *Proc. Natl. Acad. Sci. USA* 105, 13269–13273. doi: 10.1073/pnas.0803424105
- Bergen, B., and Wheeler, K. (2010). Grammatical aspect and mental simulation. *Brain Lang.* 112, 150–158. doi: 10.1016/j.bandl.2009.07.002
- Borghi, A. M., Glenberg, A. M., and Kaschak, M. P. (2004). Putting words in perspective. *Mem. Cognit.* 32, 863–873. doi: 10.3758/bf03196865
- Borghi, A. M., and Scorolli, C. (2009). Language comprehension and dominant hand motion simulation. *Hum. Mov. Sci.* 28, 12–27. doi: 10.1016/j.humov.2008.07.002
- Bransford, J. D., Barclay, J. R., and Franks, J. (1972). Sentence memory: a constructive versus interpretive approach. *Cogn. Psychol.* 3, 193–209. doi: 10.1016/0010-0285(72)90003-5
- Brass, M., Bekkering, H., Wohlschläger, A., and Prinz, W. (2000). Compatibility between observed and executed finger movements: comparing symbolic, spatial, and imitative cues. *Brain Cogn.* 44, 124–143. doi: 10.1006/brcg.2000.1225
- Brunyé, T. T., Ditman, T., Mahoney, C. R., Augustyn, J. S., and Taylor, H. A. (2009). When you and I share perspectives: pronouns modulate perspective taking during narrative comprehension. *Psychol. Sci.* 20, 27–32. doi: 10.1111/j.1467-9280.2008.02249.x
- Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., et al. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *Eur. J. Neurosci.* 13, 400–404. doi: 10.1111/j.1460-9568.2001.01385.x
- Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., and Rizzolatti, G. (2005). Listening to action-related sentences modulates the activity of the motor system: a combined TMS and behavioral study. *Brain Res. Cogn. Brain Res.* 24, 355–363. doi: 10.1016/j.cogbrainres.2005.02.020
- Campanella, F., Sandini, G., and Morone, M. C. (2011). Visual information gleaned by observing grasping movement in allocentric and egocentric perspectives. *Proc. Biol. Sci.* 278, 2142–2149. doi: 10.1098/rspb.2010.2270
- Clark, H. (1996). *Using Language*. Cambridge: Cambridge University Press.
- Clark, H. H., and Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition* 22, 1–39. doi: 10.1016/0010-0277(86)90010-7
- Decety, J. (2002). “Neurophysiological evidence for simulation of action,” in *Simulation and Knowledge of Action*, eds J. Dokic and J. Proust (Amsterdam: John Benjamins), 53–72.
- Decety, J., and Sommerville, J. A. (2003). Shared representations between self and other: a social cognitive neuroscience view. *Trends Cogn. Sci.* 7, 527–533. doi: 10.1016/j.tics.2003.10.004
- de Vignemont, F., and Haggard, P. (2008). Action observation and execution: what is shared? *Soc. Neurosci.* 3, 421–433. doi: 10.1080/17470910802045109
- Ditman, T., Brunyé, T. T., Mahoney, C. R., and Taylor, H. (2010). Simulating an enactment effect: pronouns guide action simulation during narrative comprehension. *Cognition* 115, 172–178. doi: 10.1016/j.cognition.2009.10.014
- Duran, N. D., Dale, R., and Kreuz, R. J. (2011). Listeners invest in an assumed other's perspective despite cognitive cost. *Cognition* 121, 22–40. doi: 10.1016/j.cognition.2011.06.009
- Fischer, M. H., and Zwaan, R. (2008). Embodied language: a review of the role of the motor system in language comprehension. *Q. J. Exp. Psychol. (Hove)* 61, 825–850. doi: 10.1080/17470210701623605
- Flavell, J. H., Everett, B. A., Croft, K., and Flavell, E. R. (1981). Young children's knowledge about visual perception: further evidence for the Level 1-Level 2 distinction. *Dev. Psychol.* 17, 99–103. doi: 10.1037/0012-1649.17.1.99
- Gallese, V., and Goldman, A. (1998). Mirror neurons and the mind-reading. *Trends Cogn. Sci.* 2, 493–501. doi: 10.1016/s1364-6613(98)01262-5
- Gambi, C., and Pickering, M. J. (2011). A cognitive architecture for the coordination of utterances. *Front. Psychol.* 2:275. doi: 10.3389/fpsyg.2011.00275
- Gianelli, C., Farnè, A., Salemme, R., Jeannerod, M., and Roy, A. C. (2011). The agent is right: when motor embodied cognition is space-dependent. *PLoS One* 6:e25036. doi: 10.1371/journal.pone.0025036
- Gianelli, C., Scorolli, C., and Borghi, A. M. (2013). Acting in perspective: the role of body and language as social tools. *Psychol. Res.* 77, 40–52. doi: 10.1007/s00426-011-0401-0
- Glenberg, A. M., and Gallese, V. (2012). Action-based language: a theory of language acquisition, comprehension, and production. *Cortex* 48, 905–922. doi: 10.1016/j.cortex.2011.04.010
- Glenberg, A. M., and Kaschak, M. P. (2002). Grounding language in action. *Psychon. Bull. Rev.* 9, 558–565. doi: 10.3758/bf03196313
- Glenberg, A. M., Sato, M., and Cattaneo, L. (2008a). Use-induced motor plasticity affects the processing of abstract and concrete language. *Curr. Biol.* 18, R290–R291. doi: 10.1016/j.cub.2008.02.036
- Glenberg, A. M., Sato, M., Cattaneo, L., Riggio, L., Palumbo, D., and Buccino, G. (2008b). Processing abstract language modulates motor system activity. *Q. J. Exp. Psychol. (Hove)* 61, 905–919. doi: 10.1080/17470210701625550
- Grèzes, J., Armony, J., Rowe, J., and Passingham, R. (2003). Activations related to “mirror” and “canonical” neurones in the human brain: an fMRI study. *Neuroimage* 18, 928–937. doi: 10.1016/s1053-8119(03)00042-9
- Grèzes, J., and Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: a meta-analysis. *Hum. Brain Mapp.* 12, 1–19. doi: 10.1002/1097-0193(200101)12:1<1::aid-hbm10>3.0.co;2-v
- Hauk, O., Johnsrude, I., and Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron* 41, 301–307. doi: 10.1016/s0896-6273(03)00838-9
- Hoedemaker, R. S., and Gordon, P. C. (2013). Embodied language comprehension: encoding-based and goal-driven processes. *J. Exp. Psychol. Gen.* [Epub ahead of print].
- Hommel, B., Müsseler, J., Aschersleben, G., and Prinz, W. (2001). The Theory of Event Coding (TEC): a framework for perception and action planning. *Behav. Brain Sci.* 24, 849–878. doi: 10.1017/s0140525x01000103
- Honeck, R. P. (1973). Interpretive versus structural effects on semantic memory. *J. Verbal Learning Verbal Behav.* 12, 448–455. doi: 10.1016/s0022-5371(73)80025-8
- Jeannerod, M. (2006). *Motor Cognition: What Actions Tell the Self*. Oxford: Oxford University Press.
- Johnson-Laird, P. N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference and Consciousness*. Cambridge: Cambridge University Press.
- Johnson-Laird, P. N., and Stevenson, R. (1970). Memory for syntax. *Nature* 227, 412.
- Kaschak, M. P., Madden, C. J., Theriault, D. J., Yaxley, R. H., Aveyard, M., Blanchard, A., et al. (2005). Perception of motion affects language processing. *Cognition* 94, B79–B89. doi: 10.1016/j.cognition.2004.06.005

- Kaschak, M. P., Zwaan, R., Aveyard, M., and Yaxley, R. H. (2006). Perception of auditory motion affects language processing. *Cogn. Sci.* 30, 733–744. doi: 10.1207/s15516709cog0000_54
- Kessler, K., and Rutherford, H. (2010). The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions. *Front. Psychol.* 1:213. doi: 10.3389/fpsy.2010.00213
- Kiefer, M., and Pulvermüller, F. (2012). Conceptual representations in mind and brain: theoretical developments, current evidence and future directions. *Cortex* 48, 805–825. doi: 10.1016/j.cortex.2011.04.006
- Kilner, J. M., Paulignan, Y., and Blake-more, S.-J. (2003). An interference effect of observed biological movement on action. *Curr. Biol.* 13, 522–525. doi: 10.1016/s0960-9822(03)00165-9
- Kosslyn, S. M., Ball, T. M., and Reiser, B. J. (1978). Visual images preserve metric spatial information: evidence from studies of image scanning. *J. Exp. Psychol. Hum. Percept. Perform.* 4, 47–60. doi: 10.1037/0096-1523.4.1.47
- Kutas, M., and Federmeier, K. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends Cogn. Sci.* 4, 463–470. doi: 10.1016/s1364-6613(00)01560-6
- Levinson, S. C. (1996). “Frames of reference and Molyneux’s question: cross-linguistic evidence,” in *Language and Space*, eds P. Bloom, M. Peterson, L. Nadel and M. Garrett (Cambridge, MA: MIT Press), 109–169.
- Levinson, S. C. (2003). *Space and Language in Cognition: Explorations in Cognitive Diversity*. Cambridge: Cambridge University Press.
- Mainwaring, S. D., Tversky, B., Ohgishi, M., and Schiano, D. J. (2009). Descriptions of simple spatial scenes in english and japanese. *Spat. Cogn. Comput.* 3, 3–42. doi: 10.1207/s15427633scc0301_2
- Marino, B. F. M., Gallese, V., Buccino, G., and Riggio, L. (2012). Language sensorimotor specificity modulates the motor system. *Cortex* 48, 849–856. doi: 10.1016/j.cortex.2010.12.003
- Mazzarella, E., Hamilton, A., Trojano, L., Mastroianni, B., and Conson, M. (2012). Observation of another’s action but not eye gaze triggers allocentric visual perspective. *Q. J. Exp. Psychol. (Hove)* 65, 2447–2460. doi: 10.1080/17470218.2012.697905
- Meteyard, L., Cuadrado, S. R., Bahrami, B., and Vigliocco, G. (2012). Coming of age: a review of embodiment and the neuroscience of semantics. *Cortex* 48, 788–804. doi: 10.1016/j.cortex.2010.11.002
- Michelon, P., and Zacks, J. M. (2006). Two kinds of visual perspective taking. *Percept. Psychophys.* 68, 327–337. doi: 10.3758/bf03193680
- Müsseler, J., and Hommel, B. (1997). Blindness to response-compatible stimuli. *J. Exp. Psychol. Hum. Percept. Perform.* 23, 861–872.
- Papeo, L., Corradi-Dell’Acqua, C., and Rumiati, R. I. (2011). “She” is not like “I”: the tie between language and action is in our imagination. *J. Cogn. Neurosci.* 23, 3939–3948. doi: 10.1162/jocn_a_00075
- Pecher, D., van Dantzig, S., Zwaan, R. A., and Zeelenberg, R. (2009). Language comprehenders retain implied shape and orientation of objects. *Q. J. Exp. Psychol. (Hove)* 62, 1108–1114. doi: 10.1080/17470210802633255
- Pickering, M. J., and Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behav. Brain Sci.* 27, 169–190. doi: 10.1017/s0140525x04000056
- Pickering, M. J., and Garrod, S. (2006). Alignment as the basis for successful communication. *Res. Lang. Comput.* 4, 203–228. doi: 10.1007/s11168-006-9004-0
- Pickering, M. J., McLean, J. F., and Gambi, C. (2012). Do addressees adopt the perspective of the speaker? *Acta Psychol. (Amst)* 141, 261–269. doi: 10.1016/j.actpsy.2012.06.001
- Prinz, W. (1997). Perception and action planning. *Eur. J. Cogn. Psychol.* 9, 129–154. doi: 10.1080/713752551
- Prinz, W., and Hommel, B. (Eds.). (2002). *Common Mechanisms in Perception and Action: Attention and Performance XIV*. Oxford: Oxford University Press.
- Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nat. Rev. Neurosci.* 6, 576–582. doi: 10.1038/nrn1706
- Pulvermüller, F., Shtyrov, Y., and Ilmoniemi, R. (2005). Brain signatures of meaning access in action word recognition. *J. Cogn. Neurosci.* 17, 884–892. doi: 10.1162/0898929054021111
- Rizzolatti, G., and Arbib, M. A. (1998). Language within our grasp. *Trends Neurosci.* 21, 188–194. doi: 10.1016/s0166-2236(98)01260-0
- Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192. doi: 10.1146/annurev.neuro.27.070203.144230
- Ruby, P., and Decety, J. (2001). Effect of subjective perspective taking during simulation of action: a PET investigation of agency. *Nat. Neurosci.* 4, 546–550. doi: 10.1038/87510
- Ruby, P., and Decety, J. (2004). How would you feel versus how do you think she would feel? A neuroimaging study of perspective-taking with social emotions. *J. Cogn. Neurosci.* 16, 988–999. doi: 10.1162/0898929041502661
- Sachs, J. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Percept. Psychophys.* 2, 437–442. doi: 10.3758/bf03208784
- Sato, M., and Bergen, B. K. (2013). The case of the missing pronouns: does mentally simulated perspective play a functional role in the comprehension of person? *Cognition* 127, 361–374. doi: 10.1016/j.cognition.2013.02.004
- Sato, M., Mengarelli, M., Riggio, L., Gallese, V., and Buccino, G. (2008). Task related modulation of the motor system during language processing. *Brain Lang.* 105, 83–90. doi: 10.1016/j.bandl.2007.10.001
- Schober, M. (1993). Spatial perspective-taking in conversation. *Cognition* 47, 1–24. doi: 10.1016/0010-0277(93)90060-9
- Schober, M. (1995). Speakers, addressees, and frames of reference: whose effort is minimized in conversations about locations? *Discourse Process.* 20, 219–247. doi: 10.1080/01638539509544939
- Schober, M. (2009). “Spatial dialogue between partners with mismatched abilities,” in *Spatial Language and Dialogue*, eds K. Coventry, T. A. Tenbrink and J. Bateman (Oxford: Oxford University Press), 23–39.
- Schober, M., and Brennan, S. E. (2003). “Processes of interactive spoken discourse: the role of the partner,” in *Handbook of Discourse Processes*, eds A. C. Graesser, M. A. Gernsbacher and S. R. Goldman (Mahwah, NJ: Lawrence Erlbaum), 123–164.
- Sebanz, N., Bekkering, H., and Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends Cogn. Sci.* 10, 70–76. doi: 10.1016/j.tics.2005.12.009
- Stanfield, R., and Zwaan, R. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychol. Sci.* 12, 153–156. doi: 10.1111/1467-9280.00326
- Taylor, L. J., Lev-Ari, S., and Zwaan, R. A. (2008). Inferences about action engage action systems. *Brain Lang.* 107, 62–67. doi: 10.1016/j.bandl.2007.08.004
- Taylor, L. J., and Zwaan, R. A. (2008). Motor resonance and linguistic focus. *Q. J. Exp. Psychol. (Hove)* 61, 896–904. doi: 10.1080/17470210701625519
- Taylor, L. J., and Zwaan, R. A. (2013). “Fault-tolerant comprehension,” in *Language and Action in Cognitive Neuroscience*, eds Y. Coello and A. Bartolo (Hove: Psychology Press), 145–158.
- Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., et al. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *J. Cogn. Neurosci.* 17, 273–281. doi: 10.1162/0898929053124965
- Tomasino, B., Werner, C. J., Weiss, P. H., and Fink, G. R. (2007). Stimulus properties matter more than perspective: an fMRI study of mental imagery and silent reading of action phrases. *Neuroimage* 36(Suppl. 2), T128–T141. doi: 10.1016/j.neuroimage.2007.03.035
- Tversky, B. (1991). Spatial mental models. *Psychol. Learn. Motiv. Adv. Res. Theory* 27, 109–145.
- Tversky, B., and Hard, B. M. (2009). Embodied and disembodied cognition: spatial perspective-taking. *Cognition* 110, 124–129. doi: 10.1016/j.cognition.2008.10.008
- Van Dijk, T. A., and Kintsch, W. (1983). *Strategies of Discourse Comprehension*. New York: Academic Press.
- Watson, M. E., Pickering, M. J., and Branigan, H. P. (2004). “Alignment of reference frames in dialogue,” in *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (Mahwah, NJ: Lawrence Erlbaum).
- Willems, R. M., Hagoort, P., and Casasanto, D. (2010). Body-specific representations of action verbs: neural evidence from right- and left-handers. *Psychol. Sci.* 21, 67–74. doi: 10.1177/0956797609354072
- Willems, R. M., Toni, I., Hagoort, P., and Casasanto, D. (2009). Body-specific motor imagery of hand actions: neural evidence from right- and left-handers. *Front. Hum. Neurosci.* 3:39. doi: 10.3389/fpsy.2009.02.002
- Wu, L., and Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: evidence from property generation. *Acta Psychol. (Amst)* 132, 173–189. doi: 10.1016/j.actpsy.2009.02.002
- Yoon, S. O., Koh, S., and Brown-Schmidt, S. (2012). Influence of

- perspective and goals on reference production in conversation. *Psychon. Bull. Rev.* 19, 699–707. doi: 10.3758/s13423-012-0262-6
- Zwaan, R. A. (2008). Time in language, situation models, and mental simulations. *Lang. Learn.* 58, 13–26. doi: 10.1111/j.1467-9922.2008.00458.x
- Zwaan, R. A., Langston, M. C., and Graesser, A. C. (1995). The construction of situation models in narrative comprehension: an event-indexing model. *Psychol. Sci.* 6, 292–297. doi: 10.1111/j.1467-9280.1995.tb00513.x
- Zwaan, R. A., and Radvansky, G. (1998). Situation models in language comprehension and memory. *Psychol. Bull.* 123, 162–185. doi: 10.1037/0033-2909.123.2.162
- Zwaan, R. A., Stanfield, R., and Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychol. Sci.* 13, 168–171. doi: 10.1111/1467-9280.00430
- Zwaan, R. A., and Taylor, L. J. (2006). Seeing, acting, understanding: motor resonance in language comprehension. *J. Exp. Psychol. Gen.* 135, 1–11. doi: 10.1037/0096-3445.135.1.1
- Zwicker, J. (2009). Agency attribution and visuospatial perspective taking. *Psychon. Bull. Rev.* 16, 1089–1093. doi: 10.3758/pbr.16.6.1089
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Received: 20 June 2013; accepted: 28 August 2013; published online: 17 September 2013.
- Citation: Beveridge MEL and Pickering MJ (2013) Perspective taking in language: integrating the spatial and action domains. *Front. Hum. Neurosci.* 7:577. doi: 10.3389/fnhum.2013.00577
- This article was submitted to the journal *Frontiers in Human Neuroscience*. Copyright © 2013 Beveridge and Pickering. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Reference frame selection in dialog: priming or preference?

Katrin Johannsen* and Jan P. De Ruiter

Faculty of Linguistics and Literary Studies, Bielefeld University, Bielefeld, Germany

Edited by:

Sarah H. Creem-Regehr, University of Utah, USA

Reviewed by:

Holly A. Taylor, Tufts University, USA
Marios Avraamides, University of Cyprus, Cyprus

*Correspondence:

Katrin Johannsen, Faculty of Linguistics and Literary Studies, Universitätsstraße 25, D-33615 Bielefeld, Germany
e-mail: katrin.johannsen@uni-bielefeld.de

We investigate effects of priming and preference on frame of reference (FOR) selection in dialog. In a first study, we determine FOR preferences for specific object configurations to establish a baseline. In a second study, we focus on the selection of the relative or the intrinsic FOR in dialog using the same stimuli and addressing the questions whether (a) interlocutors prime each other to use the same FOR consistently or (b) the preference for the intrinsic FOR predominates priming effects. Our results show effects of priming (more use of the relative FOR) and a decreased preference for the intrinsic FOR. However, as FOR selection did not have an effect on target trial accuracy, neither effect alone represents the key to successful communication in this domain. Rather, we found that successful communication depended on the adaptation of strategies between interlocutors: the more the interlocutors adapted to each other's strategies, the more successful they were.

Keywords: spatial perspective, priming, spatial frames of reference, psycholinguistic, dialog

INTRODUCTION

Localizing an object with reference to another object is common in natural language. For instance, consider the sentence “The book is to the left of the chair.” It is ambiguous whether the book is at the chair's left or whether it is to the left of the chair as viewed from the speaker's perspective. In order to refer to these different perspectives, frames of reference (FOR) are used. FOR are a set of axes that parse space (Carlson, 1999) and can be considered as coordinate systems that impose an orientation on the environment, people, or objects. These coordinate systems have an origin constituted by the point of intersection (Miller and Johnson-Laird, 1976), a direction and an orientation (Logan and Sadler, 1996). Following Levinson (1996, 2003), three different FORs can be distinguished which differ with regard to their origin and the spatial relationship they establish. The *relative* FOR establishes viewpoint-dependent ternary spatial relationships. A description of the object configuration in **Figure 1** according to the relative FOR would be “The plant is in front of the chair”; the spatial relationship comprises the plant, the chair and the viewer. In the present study, the origin of the relative FOR always lies in the viewer; the coordinate system is thus oriented egocentrically and the spatial relationship comprises the speaker's viewpoint and two objects. In the case of the *intrinsic* FOR, the relationships are binary and viewpoint-independent. The intrinsic FOR is used when the origin lies in the object itself and the direction of the FOR is oriented according to the inherent axes of the object (“The plant is next to the chair” in **Figure 1**). The *absolute* FOR is based on environmental features such as gravity or the cardinal directions and will not be considered in this study.

Situations in which the relative and the intrinsic FOR can be used interchangeably exhibit a high potential for ambiguities. If both FOR are equally likely to be used and speakers do not indicate which FOR they are using, the probability that an interlocutor interprets the FOR correctly is at chance level. Attempts to define preferences for specific FORs have led to ambiguous results. The relative FOR, being perceptually available

and avoiding the extra computational effort needed for mental rotation, has been considered predominant by some authors (Linde and Labov, 1975; Levelt, 1982, 1989) whereas other authors have claimed that the intrinsic FOR predominates (Miller and Johnson-Laird, 1976) or is at least preferred (Carlson-Radvansky and Irwin, 1993; Carlson-Radvansky and Radvansky, 1996; Taylor et al., 1999). This disagreement and the potential for ambiguities has led to an extensive body of psycholinguistic investigations of which factors contribute to the selection and processing of spatial FOR, mostly using monolog studies. The factors identified range from functional relations between objects (Carlson-Radvansky and Radvansky, 1996) to motion characteristics (Levelt, 1984), gravity (Friederici and Levelt, 1990), priming effects (Watson et al., 2004; Carlson and Van Deman, 2008; Johannsen and de Ruiter, 2013), scene type (Johannsen and de Ruiter, 2013), and properties of the object configuration such as the rotation of the reference object and the position of the located object (Ziegler et al., 2012).

However, monolog studies do not allow us to investigate how interlocutors deal with FOR ambiguities in dialog. Dialog differs from monolog in that dialog “language use is really a form of joint action” (Clark, 1996, p. 3) which suggests that FOR interpretation in dialog requires that interlocutors coordinate FOR selection in order to communicate successfully. Even though there have been attempts to investigate spatial perspective taking involving an imagined interlocutor (Herrmann, 1988; Duran et al., 2011), studies using real dialogs are rare. Perspective taking between interlocutors who have different physical vantage points and thus different perspectives on the same scene is not considered in the present study, but interesting results can be found elsewhere (e.g., Bürkle et al., 1986; Schober, 1993; Galati et al., 2013). Watson et al. (2004) showed that dialog partners tended to align on FORs, revealing a tendency to use the same FOR that their interlocutor had previously used. This was interpreted as a case of alignment resulting from priming effects. Pickering and Garrod (2004, p. 173) claim that alignment is a key factor for successful

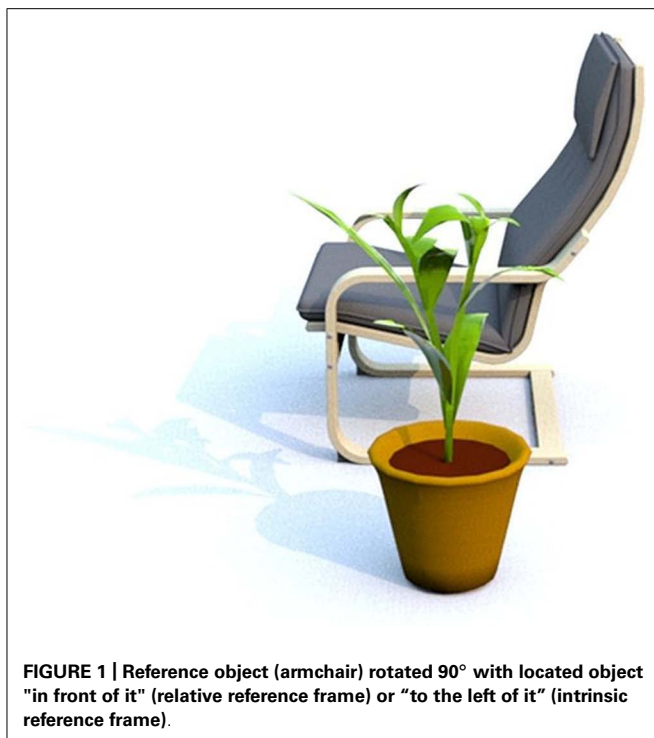


FIGURE 1 | Reference object (armchair) rotated 90° with located object "in front of it" (relative reference frame) or "to the left of it" (intrinsic reference frame).

communication and results from priming which is "essentially resource-free and automatic." However, Watson et al. (2004) used a confederate as one of their participants, and the confederate's utterances were scripted. Assuming that people do not merely adopt the interlocutor's strategies but rather mutually influence one another in dialog, a confederate may not represent a natural dialog counterpart. Thus, FOR selection in a real dialog with two naïve participants may reveal different effects.

However, following the attempts to specify FOR preferences (as described above), it has been shown that there is a general preference for the intrinsic FOR in specific object configurations (Ziegler et al., 2012). This study also demonstrated an effect of the located object's position with regard to the reference object's FOR. If the located object was positioned on the front/back axis of a FOR, this made the selection of the respective FOR more probable. Thus, these axis-dependent preferences may reduce variability in FOR selection independent from priming effects. Furthermore, the general preference for the intrinsic FOR may lead interlocutors to establish a conceptual pact (comparable to conceptual pacts in lexical choices as discussed by Brennan and Clark, 1996) and use it consistently. However, in such cases, conflicts may arise from the opposing impacts of priming and preference of FOR.

The interaction between preference and priming effects has not, to our knowledge, previously been investigated. We expect that if automatic priming is a prevailing effect in conversation that leads to FOR alignment (cf. Pickering and Garrod, 2004; Watson et al., 2004), this should override FOR preferences. However, if FOR preferences lead to conceptual pacts with regard to FOR selection, this may override priming effects.

To investigate effects of preference and priming in dialog, we developed a priming study in which pairs of naïve participants

described pictures of object configurations to each other. In each round, one of the participants was the director, who described a spatial configuration displayed on a monitor while the other participant (matcher) had to choose between two displayed pictures. While half of the stimuli only allowed the use of the relative FOR (prime trials), the other half consisted of stimuli allowing the use of the intrinsic FOR (target trials). After every two trials, the roles changed and the matcher became director. Thus, hearing the interlocutor A use the relative FOR in the prime trial should, according to the priming account, prime interlocutor B to select the relative FOR in the target trial.

STUDY I: FOR PREFERENCES

In order to be able to separate preference from priming effects, we conducted a study in which we determined FOR preferences for specific object configurations as a baseline for comparisons.

MATERIALS AND METHODS

As described above, FOR preferences are highly dependent on the context and object features. For this reason, we focused on objects from a single category (furniture). Spatial verbal descriptions were elicited in an online experiment in which participants were shown pictures of object configurations and were instructed to define the spatial relations by inserting spatial terms in gapped sentences. Their FOR preferences served as a baseline in the dialog study.

Participants

244 participants were recruited by email invitation. Data from 34 participants had to be excluded (due to a cease of participation or different native language), thus, data from 210 participants (168 women, 42 men) with a mean age of 24.1 years (ranging in age from 7 to 72 years) were used for analysis.

Stimuli and design

Stimuli were pictures of object configurations and German gapped sentences of the form "<located Object> steht _____<reference object>." ("<located Object> stands _____<reference object>."). Thus, participants had to insert a spatial preposition and an article to fill the gap. Pictures were created using Sweet Home 3D, an architectural design software. 66 pictures were created, each consisting of a reference object and a located object. Different orientations of the reference object resulted from rotating it clockwise at angles of 90°, starting at 0° (reference object faces the observer). The located object (a plant or a stool) was placed in four different positions: relatively in front of, to the left of, behind and to the right of the reference object. This led to potential ambiguities in the descriptions of the located object, as a reference object rotated by 90° and a located object placed relatively in front could also be described as "next to" using the intrinsic FOR (see **Figure 1**). Following Graf and Herrmann (1989), we distinguished between vehicle (e.g., chair) and opposite (e.g., shelf) objects that reveal differences in the assignment of the intrinsic left/right axis according to their predominant use. Of the 66 pictures, 36 consisted of vehicle objects (chair, armchair, sofa) in four different orientations (excluding object configurations in which the intrinsic and relative FORs were aligned) and 30 showed opposite objects (wardrobe, bookshelf, chest of drawers). For the opposite objects, only the rotations 0°,

90°, and 270° were used, as these objects are characteristically used with their back to a wall. We distinguished between these object categories in order to control for potential differences in FOR selection.

The randomization procedure took reference objects and their rotation as well as located objects and their positions into account.

PROCEDURE

Participants were recruited by email invitation, in which they received a link to the online study. First, instructions and three examples were given using objects distinct from those in the study. Afterward, the participants were shown the stimuli and asked to fill in the gaps of the sentences. The whole study comprised 66 trials and lasted about 20 min. Participants could then enter a prize draw for one of 10 prizes of 10 Euros.

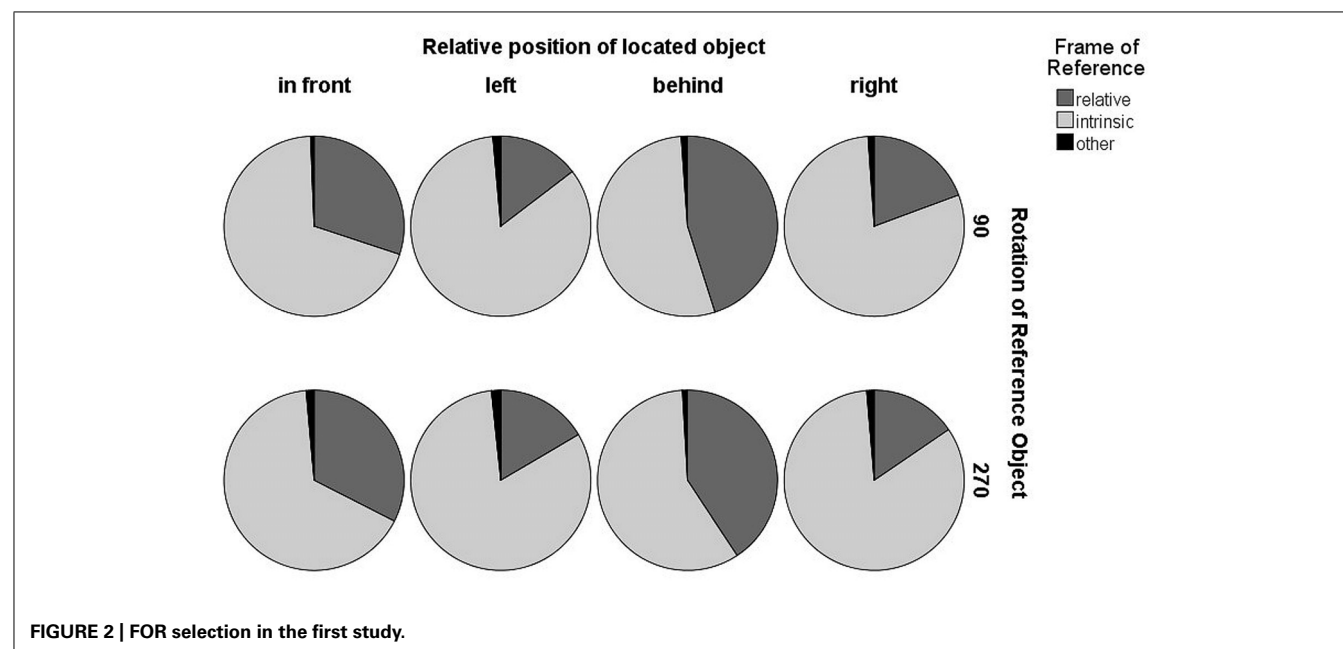
RESULTS¹

Assuming axis-dependent regularities in FOR selection, we investigated the effect of object rotation and position of the located object on FOR selection. Thus, the descriptions of the participants were coded as using “relative FOR,” “intrinsic FOR,” or “other” (for cases on which no FOR was used). Two rotations (90°, 270°) were used for analysis to ensure a constant dissociation of FORs.

Statistical analyses were performed with R (R Core Team, 2013) using the “lme4” package (Bates et al., 2011). Mixed-effects models of logistic regression for binomially distributed outcomes (generalized linear mixed models, GLMM) were used for the analysis of FOR selection. Mixed-effects models are efficient for the analysis of psycholinguistic data as they allow to include random effects of subjects and items “effectively solving the ‘language as a fixed effect fallacy’” (Quené and van den Bergh, 2008, p. 413).

Descriptions that did not use either FOR were excluded (1.37% of the data). As we only used two rotations of the reference object, the position of the located object was either on the relative or on the intrinsic front/back axis of the reference object. In order to investigate FOR preferences resulting from the position of the located object, we fit a logistic mixed-effects model with position of the located object as fixed effect, full random slopes, and intercepts for subjects and items and FOR selection as dependent variable. Positing the relative front position as intercept, we found significant differences to all other positions (relative left, i.e., intrinsic front/back position: $\beta = 2.61$, $SE = 0.3$, $z = 8.66$; relative behind: $\beta = -1.37$, $SE = 0.19$, $z = -7.05$; relative right, i.e., intrinsic front/back position: $\beta = 2.02$, $SE = 0.36$, $z = 5.68$, all $p < 0.001$). These differences in FOR selection resulting from the position of the located object are illustrated in **Figure 2**. Please note that the relative positions “left” and “right” coincide with the intrinsic front–back axis. Regularities of FOR selection suggest an axis-dependent effect, potentially comparable to the distinction between two forms of visuospatial perspective taking (Flavell, 1986). Front/behind judgments are easier to process than left/right relations as they do not require a simulated rotation movement (Kessler and Rutherford, 2010). Furthermore, this axis-effect stands in line with previous research which has shown that the front/back axis is easier to access than the left/right axis due to body asymmetries (Franklin and Tversky, 1990). Additionally, we speculate that the differences between relative “front” and “behind” (i.e., more relative FOR selection when the located object is positioned behind the reference object) might result from the occlusion of the located object. We assume that this occlusion might give more salience to the relative FOR.

We controlled object category in the design in order to eliminate effects of object category on FOR selection. However, when we additionally posited object category as fixed effect in the same



model, model comparison revealed no statistically significant effect of object category. This indicates that FOR selection did not differ between vehicle and opposite objects.

CONCLUSION

Our results reveal a general preference for the intrinsic FOR but also significant effects of the position of the located object. Accordingly, we are now able to differentiate between preferred choice for the object configurations (i.e., the intrinsic FOR) and priming effects.

STUDY II: PRIMING vs. PREFERENCE

MATERIALS AND METHODS

Participants

Fifty four participants were paid volunteers in the experiment. Due to experimenter error, two groups (four participants) had to be excluded, thus data from 50 participants (11 male, 39 female) ranging in age from 19 to 61 ($M = 24.3$, $SD = 6.2$) was used.

Stimuli and design

Using a priming paradigm, we constructed prime and target trials in three conditions which only differed with regard to the prime trials. We thus controlled the target trials within items. Altogether, the experiment consisted of 144 prime-target pairs resulting from three priming conditions for each of the 48 target pictures (3 priming conditions \times 6 reference objects \times 2 rotations \times 4 positions of LO).

The stimuli were pictures created with indoor planning software (Sweet Home 3D). The pictures showed object configurations, consisting of a reference object and a located object. For the prime trials, three types of pictures were created (33 pictures for each type): *neutral*, *same position*, and *different position*. In the neutral pictures, both FOR were available and aligned as the located object was positioned along the vertical axis of a triaxial reference object. The other two types of pictures (*same* and *different position*) comprised a biaxial reference object and a located object which was positioned on one of the horizontal axes. Accordingly, the intrinsic FOR was unavailable and the

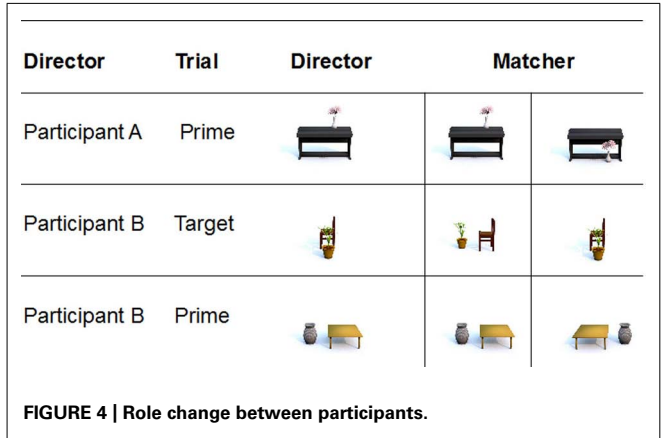
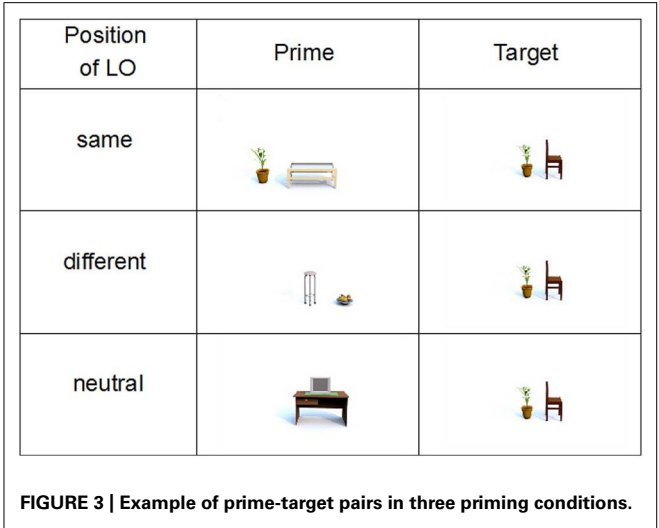
participants had to use the relative FOR. In the *same position* condition, the located object was at the same position in prime and target trials (e.g., to the left of the reference object within the relative FOR, see **Figure 3**). In the *different position* condition, the located object was placed at the opposite side of the reference object than in the target trial (e.g., to the right of the reference object in the prime trial and to the left of the reference object in the target trial, both within the relative FOR). The *same* and *different position* conditions were used to test whether priming effects are stronger when the located object has the same position in prime and target trial which would be a plausible consequence of lexical priming, given that the same prepositions would be used.

The pictures described in the first study (in 90° and 270° rotation) were used as target stimuli. See **Figure 3** for an example of the three priming conditions using the same target trial.

Randomization took into account priming condition, the reference object and its rotation, and the position of the located object. To counterbalance the sequential order, the experiment was conducted in two versions by switching the order for half the participants.

There were two roles for the participants that changed after every two trials: the director and the matcher. The director was shown a single picture of an object configuration and described it to the matcher while the matcher was shown two pictures and had to decide which of the two fitted the director’s description. The matcher’s two pictures always showed the same reference object at the same rotation as on the director’s screen. However, the position of the located object differed so that the director’s descriptions became potentially ambiguous with regard to FOR in the target trials. Thus, if participant B (director) described the target configuration in **Figure 4** as “The plant is in front of the chair,” either picture could plausibly be correct depending on the matcher’s FOR interpretation. Interpreted within a relative FOR, the picture on the left is correct; interpreted within the intrinsic FOR, the picture on the right is correct. However, as only one of the two pictures corresponded to the director’s picture, its choice revealed whether participants successfully solved the problem of ambiguity.

After every two trials, the roles changed so that the director became the matcher and vice versa. Therefore, the description of



the previous director was used as the prime for the description of the subsequent director. Thus, participants took it in turns to prime each other. See **Figure 4** for an example of a prime-target sequence (in the subsequent target trial, participant A would have been the director).

PROCEDURE

Two naïve participants participated together as interlocutors in a dialog task. Each participant sat in front of a computer screen on which the stimuli were displayed. Participants were separated by a movable wall so that they were able to hear each other but could not see each other nor the other's computer screen. At the beginning of the experiment, written instructions were presented on the monitor, informing the participants about the procedure of the experiment. Before the start of the study, participants completed five test trials with stimuli distinct from those used in the study. After that, they were asked if the task was clear to them and if so, the study started. The director was shown a single picture whereas the matcher saw two pictures. The director immediately started describing the spatial configuration. The matchers' task was to determine which of the two pictures matched the director's description and respond by pressing predefined keys on a button box (left key for left picture, right key for right picture). Accuracy ratings were measured using E-Prime (Psychology Tools Software). The matcher was also allowed to give feedback (e.g., ask the matcher for more information, indicate ambiguities). Both participants' pictures remained on the screen until a response was given. The whole study comprised 288 trials (144 prime-target pairs) and lasted about 15 min.

The participants were unaware of the objective of the experiment and of the type of trials they were completing. No feedback was given during the experiment.

RESULTS

Data from 50 participants were used for analysis. Statistical analysis was carried out in PASW Statistics 18 and in "R" (R Core Team, 2013) using the lme4 package (Bates et al., 2011). Mixed-effects models of logistic regression (generalized linear mixed models for binomially distributed outcomes, GLMM) were used for the analysis of FOR selection and accuracy.

Our statistical analysis considered FOR selection in the director's descriptions of target trial stimuli and the matcher's accuracy. Furthermore, we conducted a qualitative analysis of the linguistic behavior in terms of the strategies used to disambiguate descriptions. Rationales for each analysis are given in each section.

Priming of FOR in dialog

In order to investigate effects of priming or preference on FOR selection, we analyzed the FOR selection in the director's descriptions. While a prevailing use of the relative FOR in target trials would indicate priming effects, predominant use of the intrinsic FOR would suggest effects resulting from FOR preference.

The director's descriptions were transcribed and categorized according to FOR use. For categorization, we used the first uninterrupted utterance of the speaker (cf. de Ruiter et al., 2012). In some cases, participants used both FOR at the same time. These

descriptions were categorized as "ambiguous" (10.3% of the data). Descriptions that did not use a specific FOR but rather referred to the location of the object on the screen or the proximity of the objects to the director were categorized as "other" (10.6% of the data). Data that revealed participant's error, for instance when the matchers erroneously described what they saw, were also excluded (0.17%). The rates of FOR selection in prime and target trials are summarized in **Table 1**.

As **Table 2** shows, the relative FOR was chosen more often than the intrinsic FOR. This result was in contrast to the preference for the intrinsic FOR in our first study. Thus, we compared FOR choice between the two studies in order to analyze whether the differences in FOR selection were statistically significant. We fit a logistic mixed-effects model with *study type* (baseline vs. dialog) as a fixed effect and random slopes and intercepts for subjects and items. The results showed a significant main effect of *study type* ($\beta = -3.21$, $SE = 0.88$, $z = -3.67$, $p < 0.001$) confirming the difference in FOR selection between the two studies and revealing more use of the relative FOR in the dialog study.

In the next step, we investigated whether FOR selection in the dialog study had an effect on target trial accuracy. We assumed that if priming or preference was a prevailing mechanism in order to disambiguate FOR, there should be an effect of FOR selection on target trial accuracy (as a measure for communicative success). Thus, we fit a logistic mixed-effects model with FOR selection as fixed effect and random slopes and intercepts for groups and items. There was no significant effect of FOR choice on target trial accuracy ($\beta = -0.79$, $SE = 0.57$, $z = -1.39$, $p = 0.17$).

Differences between priming conditions

As described above, we used three priming conditions (*neutral, same position, different position*) in order to investigate whether

Table 1 | FOR selection and target trial accuracy.

FOR	Prime trial (%) (N = 3600)	Target trial (%) (N = 3600)	Target trial accuracy (% within FOR)
Relative	93.8	50	80.8
Intrinsic	0	29	71.8
Ambiguous	0.2	10.3	82.1
Other	4.7	10.6	97.4
NA	1.3	0.2	–

Table 2 | Differences in FOR index within and between groups.

	Difference of index scores		
	Mean (SD)	Min	Max
Same group	0.18 (0.33)	0	0.97
Subsequent group	0.37 (0.4)	0	1

priming effects are stronger when the located object has the same position in prime and target trial. This would be a plausible consequence of lexical priming, given that the same prepositions would be used. Accordingly, we analyzed whether the three priming conditions differed with regard to FOR selection in the target trials. We excluded “other” and “ambiguous” answers and error cases from analysis.

Again, we fit a mixed-effects model of logistic regression positing priming condition as a fixed effect and using random slopes and intercepts for groups and items. Using the *neutral* condition as intercept, there was no significant effect of priming condition (*same position*: $\beta = -0.16$, $SE = 0.2$, $z = -0.82$, $p = 0.41$; *different position*: $\beta = -0.39$, $SE = 0.23$, $z = -1.74$, $p = 0.08$). This reflects that the three conditions did not differ with regard to FOR selection (relative or intrinsic) in the target trials. However, results revealed a marginal difference between the *neutral* condition and the *different position* condition.

Effects resulting from the position of the located object

As the first study had shown that the position of the located object had a significant effect on FOR selection indicating axis-dependent preferences, we analyzed whether this result was replicated in the second study by fitting a mixed-effects model of logistic regression (using only relative and intrinsic FOR descriptions in the target trials). We posited position of the located object as a fixed effect and used full random slopes and intercepts for groups and items. There was a significant main effect of position of the located object. Using the relative front position as intercept, there were significant differences compared to position of the intrinsic front and back (i.e., the relative left: $\beta = 1.5$, $SE = 0.29$, $z = 5.19$ and relative right: $\beta = 1.25$, $SE = 0.32$, $z = 3.88$, both $p < 0.001$) but not compared to the relative behind position. This indicates that there was a higher amount of relative FOR use when the located object was positioned along the relative front/back axis. Thus, we recoded the positions of the located object in terms of axes so that we were able to distinguish between the relative and intrinsic front/back axes. We then fit a logistic mixed-effects model using *axis* as fixed effect and random slopes and intercepts for groups and items. The results showed a significant effect of the axis ($\beta = 1.26$, $SE = 0.28$, $z = 4.51$, $p < 0.001$, see **Figure 5**) revealing a greater use of the intrinsic FOR

when the located object was positioned on the intrinsic front/back axis.

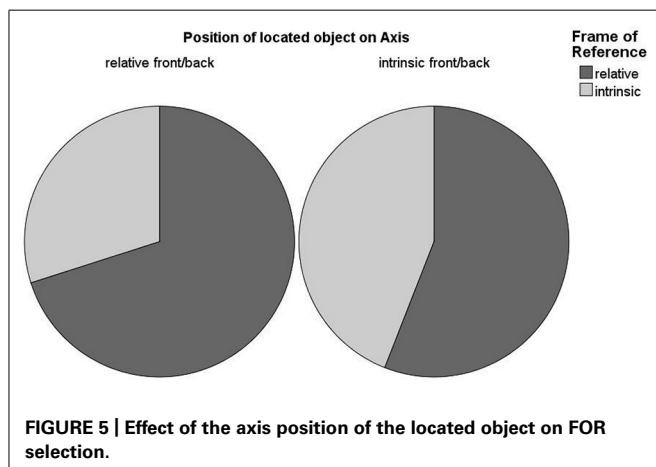
Variability in FOR selection within groups

We expected that participants within groups would adapt the same FOR in order to facilitate mutual comprehension. Thus, we analyzed the variability in FOR selection within and between groups by computing FOR indices (cf. Watson et al., 2006). These indices reflect how similar the descriptions of the participants were with regard to FOR selection within groups, i.e., the more the interlocutors used the same FOR, the lower the FOR index. The FOR indices were computed for each participant by dividing the amount of relative FOR descriptions by the sum of relative and intrinsic descriptions (thus, the analysis excluded the categories ambiguous and other). We subtracted the index from participant B from the index from participant A and squared the result to avoid negative numbers.

To test these within-group indices against indices that would arise between random interlocutors, we subtracted the indices between participants of subsequent groups. These indices from participants that did not engage in a conversation and could not influence each other reflect overall patterns that may arise by chance (see **Table 2**). As the data were not normally distributed (Shapiro–Wilk Test: both $df = 25$; index within groups S–W = 0.6; index between groups: S–W = 0.78, both $p < 0.001$) we compared the two values using Wilcoxon signed-ranks test and found a significant difference ($Z = -2.09$, $p < 0.05$). The scores between random interlocutors were significantly higher than the scores within groups; this reflects that participants adapted to each other and tended to use the same strategies, independent whatever the specific choice of strategy involved.

In order to analyze whether the time course of the dialog led to an increased adaptation of strategies between interlocutors, we calculated sliding averages of the FOR indices within groups. Considering a window of four target trials at a time, again, we proceeded as described above for the computation of FOR indices and, again, subtracted the FOR index of participant B from the FOR index from participant A. By shifting the four-trial-window one trial further at a time, we obtained averages that represented how participants adapted their FOR over time. For illustration, we chose three groups as examples (**Figure 6**) revealing different degrees of adaptation (high, medium, and low) between interlocutors (the respective target trial accuracy for the three groups is depicted in **Figure 7**).

Next, we assumed that a mutual adaptation of FOR reduced misunderstandings thus leading to a more efficient communication (measured here as target trial accuracy). We analyzed whether sliding average FOR indices within groups had an effect on target accuracy. We fit a mixed-effects model of logistic regression with target trial accuracy as dependent measure, FOR indices as fixed effects and full random slopes and intercepts for groups and items. We found a significant main effect of index scores ($\beta = -1.38$, $SE = 0.3$, $z = -4.68$, $p < 0.001$). For an illustration of the relationship between overall target trial accuracy and FOR index scores within groups, see **Figure 7** (case labels include group number and overall target accuracy for the three exemplary groups depicted in **Figure 6**).



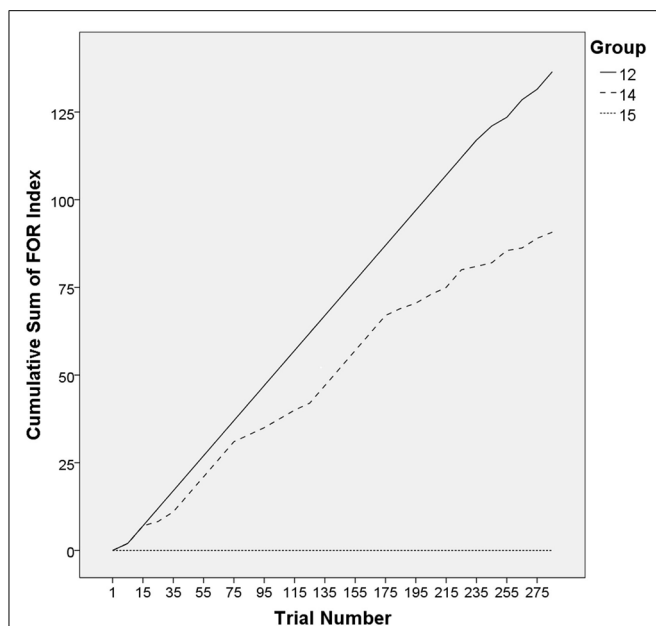


FIGURE 6 | Exemplary time course of difference scores for three groups.

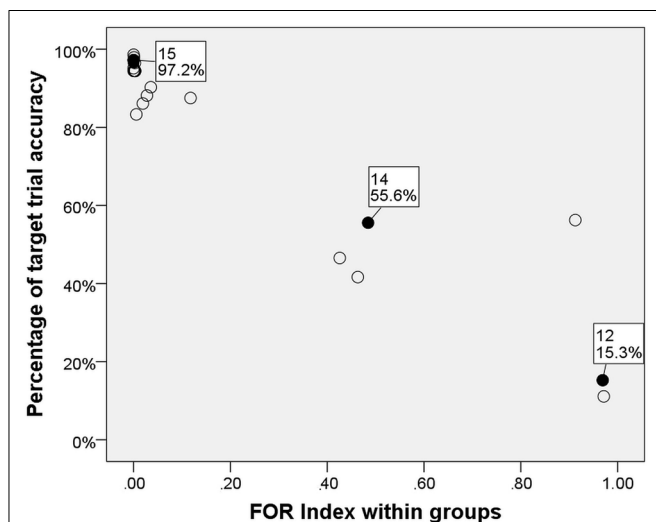


FIGURE 7 | Scatter plot of target accuracy and difference scores within groups.

QUALITATIVE ANALYSIS

In a qualitative analysis of the data, we investigated which qualitative strategies interlocutors applied to resolve ambiguities in target trials, considering only trials in which the relative or the intrinsic FOR were used. We analyzed the annotated dialog considering more than the first uninterrupted utterance for additional strategies to solve FOR ambiguities. Our analysis showed that additional information for disambiguation was provided in 23.3% of the target trials by the director even though the matcher indicated that the description was ambiguous with regard to the spatial FOR in only 1.5% of the target trials.

However, 11 out of 25 groups did not provide any additional information. Their results varied with regard to the matcher's accuracy (mean 78.8%, SD 30.6, ranging from 11 to 99%) and difference scores (mean 1841.7, SD 2690.5, ranging from 0 to 7448). The other 14 groups varied in their strategies, although we could classify three main approaches. The most common was a definition of the perspective (in 13.8%, e.g., "In front of the chair, as seen from my/the chair's perspective"). Other strategies were reference to specific intrinsic features of the object (8.3%, e.g., "The stool is in front of the front of the couch") or the use of specific verbs (1.3%) to express the position of the located object (e.g., "The plant disappears behind the sofa"). The latter was, however, only used in trials in which the located object was positioned relatively behind the reference object and was thus partly covered by it.

With regard to the quantity of strategies within each FOR, we found that more additional strategies were used within the intrinsic FOR (39.7%) than within the relative FOR (13.4%). Within the intrinsic FOR, 25.2% of the target trial descriptions contained additional information with regard to the perspective (7.3% within the relative FOR).

CONCLUSION

Our results reveal a general priming effect of the relative FOR (as shown by the comparison between the two studies) and a significant effect of the located object's position on FOR selection. There were, however, no significant differences between the three priming conditions.

Furthermore, our results show that participants adapt each other's strategies (as shown by the comparison between intra- vs. intergroup difference scores) and that target trial accuracy is influenced by the extent of this adaptation.

With regard to qualitative strategies, we found that even though FOR ambiguity was indicated in only 1.5% of the target trials, participants added further information in about a third of the target trials (27.8%). Strategies comprised perspective marking (17.2%), the reference to intrinsic features of the reference object (9.4%) or the use of verbs denoting a specific location (1.2%). Strategies were used more often within the intrinsic FOR (39.9%) than within the relative FOR (13.6%).

DISCUSSION

In the present study, we investigated effects of priming and preference on FOR selection in a dialog task. As the prime trials only allowed a description using the relative FOR (i.e., the intrinsic FOR was not available or both FOR were aligned), the priming account would predict a prevailing use of the relative FOR in the target trials (cf. Pickering and Garrod, 2004; Watson et al., 2004), even though the intrinsic FOR was available. The comparison of FOR selection (intrinsic vs. relative) between the first and the second study revealed significant differences indicating greater use of the relative FOR in the dialog study. This increase in the use of the relative FOR might reflect priming effects in target trials resulting from processing the relative FOR in the preceding prime trial. In any case, the preference for the intrinsic FOR, as found in the first study, was diminished, which indicates that this preferences cannot be considered robust and predominant (contra to Miller and Johnson-Laird, 1976). Interestingly, the choice of FOR had

no effect on target trial accuracy. In order to efficiently solve FOR ambiguities, we would have expected that participants negotiated which perspective should be used, comparable to a conceptual pact (Brennan and Clark, 1996) with regard to the spatial FOR. However, no group used this strategy and defined a consistent perspective. This indicates that the groups must have used other strategies.

Even though priming effects might explain the more frequent choice of the relative FOR, we would like to discuss the role of priming in communication. Priming leading to alignment has been claimed to be the key to successful communication (cf. to Pickering and Garrod, 2004). In our study, the primed relative FOR was used in target trial descriptions only half of the time. If priming was automatic and thus unavoidable, should we not expect a greater frequency of relative FOR selection in target trials? Given that half of the time, interlocutors did *not* use the primed FOR, the role of priming as the prevailing mechanism in communication might have been overestimated. Furthermore, FOR selection did not have an influence on target trial accuracy. If both interlocutors were primed to use the same FOR, this should be evident not only in their spatial descriptions but also in their interpretations of the other's descriptions. Thus, our findings indicate that, even though priming may have an influence on FOR selection in dialog, it may not be as automatic and comprehensive as has previously been assumed and does not necessarily lead to successful communication (measured here in terms of target trial accuracy; cf. Pickering and Garrod, 2004).

Despite the fact that there was a general priming effect, there were no differences between the three priming conditions (*neutral* vs. *same position* vs. *different position* of the located object) with regard to the FOR selection in the target trials. This suggests two things: on the one hand, it did not matter whether the relative position of the located object was the same in prime and target trial. Thus, accessing single components of the relative FOR (i.e., the front/back axis) leads to an activation of the whole FOR resulting in a priming effect in the subsequent trial. As the intrinsic FOR was either not available or aligned with the relative FOR in the prime trials, we can exclude an inhibition of the FOR as reported by Carlson and Van Deman (2008). However, due to the design of the experiment, our focus was on activation of FOR, which limits our conclusions about the nature of inhibition. On the other hand, the fact that there was no difference in effects of FOR selection between the *same* vs. *different position* condition reveals that there was no cumulative effect of lexical and FOR priming, a result that supports findings previously reported (Watson et al., 2004).

Independent of priming effects, we found effects on FOR selection resulting from the position of the located object in both studies. If the located object was positioned on the front/back axis of the FOR (relative or intrinsic), this made the choice of the respective FOR more likely. This suggests a general preference for localizing along the front–back axis and stands in line with related work. With regard to the egocentric FOR, this result coheres with the idea that the front/back relations are easier to process, due to the inherent asymmetric features, than are left/right relations, as has been reported before (e.g., Tversky, 1996). With regard to perspective taking, fundamental differences in processing

the front/back compared to the left/right axis have been reported (e.g., Kessler and Rutherford, 2010). Extending these findings with regard to the intrinsic FOR, our results emphasize the impact of the intrinsic front/back axis in spatial descriptions.

As we have shown that priming effects were less pronounced than we would have expected from an automatic process and that FOR selection did not have an effect on target trial accuracy, we assumed that the groups developed their own strategies to resolve FOR ambiguities. In order to investigate these strategies, we calculated difference scores within groups that represented how similar the descriptions of the two interlocutors were and compared them to difference scores that arose between random interlocutors. The significant difference between the groups revealed that within groups, interlocutors tended to adopt the same strategies, using either the relative or intrinsic FOR, both FOR at the same time or other descriptions which completely avoid spatial FOR. This indicates that interlocutors adapted to each other, but not necessarily by consistently using the primed relative FOR or the preferred intrinsic FOR. The efficiency of this mutual adaptation of strategies was measurable in terms of target trial accuracy: the more interlocutors adapted each other's strategies, the more accurate they were. More generally speaking, this reveals that communicative success depends on mutual adaptation. A comparable adaptation process of types of descriptions has been reported for players in a maze-game (Garrod and Anderson, 1987). Furthermore, Schober (1993) found that pairs of interlocutors in dialog varied idiosyncratically with regard to the perspective-setting strategies they used in their descriptions of spatial configurations. Under these conditions, a lot of variability between groups was possible without impairing the ultimate success of communication. This variability is necessarily reduced in dialog studies in which one of the interlocutors is a confederate. While the naïve participant can adapt to the confederate's strategies, the confederate's contributions are limited to scripted utterances. Thus, the collaborative aspect of communication that arises from the fact that “language use is really a form of joint action” (Clark, 1996, p. 3) becomes a unilateral process. This reduction in variability may explain why priming effects appear stronger in such studies.

Interestingly, there were five groups in our study that revealed a very low level of adaptation (i.e., very high difference scores) and a target trial accuracy equal to or below 56%. The low percentage of accuracy reveals that interlocutors misunderstood each other about half of the time (or even more often for lower numbers). By taking a closer look at the strategies of each participant, we found that all five groups showed the same pattern: one of the participants predominantly used the relative FOR whereas the other participant used the intrinsic FOR. This pattern may reflect individual preferences, as pointed out by Levelt (1982). Given that the experiment did not include feedback with regard to accuracy and that both target pictures could possibly be correct within different FOR interpretations, participants obviously did not realize that they used different FOR throughout the dialog. We avoided including feedback in order to allow participants to develop their own strategies for dealing with the problem of FOR ambiguity and to keep the dialog as natural as possible. Thus, misunderstandings resulting from different FOR interpretation may be common in natural language (20% of the groups experienced this problem).

Following this idea, we investigated the time course of dialog with regard to difference scores. When plotting the cumulative sum of these scores over the trial sequence, the slope of the resulting curve depends on the difference score: the higher the score, the steeper the curve. In general, groups that revealed a high level of adaptation showed a low slope of the resulting curve whereas groups that adapted each other's strategies to a lesser extent revealed a steeper curve. Taking three groups as examples that differed with regard to their target trial accuracy, we found that the more successful the group was (i.e., in maintaining overall high target trial accuracy), the lower the difference scores remained over time, indicating a constantly high level of strategy adaptation (see, group 10). As expected, the opposite pattern was found in unsuccessful groups (i.e., revealing overall low target accuracy) that showed high difference scores throughout the dialog, reflecting that participants consistently used different strategies. Group 12 (Figure 6), for example, showed this opposite pattern: constantly high difference scores arose due to different descriptions strategies between interlocutors, leading to a steep increase of the curve over time and low target trial accuracy (15.3%). Group 14 can be considered as being moderately successful with a target accuracy of 55.6%. Note that this percentage indicates that participants misunderstood each other nearly half of the time.

However, even though we can conclude that mutual adaptation of strategies seems to be strongly facilitating communicative success, an open question remains why some groups showed high levels of adaptation while other groups did not adapt at all. This question cannot be answered unambiguously but there may be two explanations. Haywood et al. (2005) have shown that speakers in a dialog study were sensitive to the ease of comprehension for their interlocutor, disambiguating their descriptions in visually ambiguous contexts. Thus, on the one hand, the lack of disambiguation in some of the groups in our study could reveal that participants erroneously assumed they were successful because they failed to notice the potential ambiguity. This stands in line with the claim that people are normally not aware of the fact that there are two alternative FOR (Grabowski and Miller, 2000, p. 526). On the other hand, we cannot exclude that participants may have deliberately chosen not to adapt to each other's strategies, for instance due to a lack of motivation for solving the task successfully. Thus, collaboration may well be a prerequisite for successful communication.

In a final step, we investigated the dialogs for qualitative strategies. Qualitative strategies consisted of explicitly adding a perspective to the FOR (e.g., "[...] as seen from my/the chair's perspective" or "[...] if you sit in the chair"), reference to intrinsic features of an object (e.g., "The plant is behind the backrest of the chair"), or the use of specific verbs (e.g., "The plant disappears behind the chair"). Qualitative strategies were used in nearly one third of the descriptions in addition to the intrinsic or relative FOR. The use of these strategies suggests that the director was aware of the ambiguity problem and tried to help the matcher by

adding unambiguous information to resolve it. The fact that this was done quite often may result from the role switching in the dialog. As both interlocutors were confronted with FOR ambiguities when they were the matcher, they were aware of the problem when they were director.

Interestingly, about a quarter of the descriptions within the intrinsic FOR contained information about the perspective and thus the Origo of the FOR. This stands in contrast to what has been assumed by Grabowski and Miller (2000, p. 521) who claimed that "the entity that constitutes the Origo is never expressed explicitly in the case of intrinsic relations." By contrast, the infrequent addition of explicit perspective in trials in which the relative FOR was used is surprisingly small given the prediction that "[...] if a deictic² interpretation is intended when an intrinsic interpretation is possible, the speaker will usually add explicitly 'from my point of view' [...]" (Miller and Johnson-Laird, 1976, p. 398). We assume that giving information on the Origo depends on the speaker's confidence about the listener's interpretation of the FOR and can be interpreted in terms of the Gricean maxims of conversation (Grice, 1975). If both interlocutors have adopted the same FOR consistently, mentioning the Origo would violate the Gricean maxim of quantity and make the contribution more informative than required (Grice, 1975, p. 308). However, when there is no such agreement on a specific FOR, providing no information on the Origo disregards the maxim of manner, i.e., avoiding ambiguity. Thus, we suggest that adding perspective reflects the speaker's degree of certainty of the listener's FOR interpretation independent of the type of FOR being used.

In conclusion, our results show that neither FOR preferences nor priming alone represent the key to successful communication in this domain. Intrinsic FOR preferences (as shown in the first study) were partly diminished by priming effects in the dialog study. However, priming effects could only account for half of the FOR selection in target trials. As groups varied widely with regard to their description strategies, priming of FOR leading to an alignment of situation models (Pickering and Garrod, 2004; Watson et al., 2004) does not provide a comprehensive account of successful communication. Rather, successful communication seems to depend on the adaptation of strategies between interlocutors: the more the interlocutors adapted to each other's strategies, the more successful they were.

ACKNOWLEDGMENTS

This work was funded by the German Research Foundation (DFG) within the Collaborative Research Centre 673 "Alignment in Communication." We acknowledge support for the Article Processing Charge by the Deutsche Forschungsgemeinschaft and the Open Access Publication Funds of Bielefeld University Library.

²In this case, the deictic FOR coincides with the relative FOR in our terminology

REFERENCES

- Bates, D., Maechler, M., and Boker, B. (2011). *lme4: Linear mixed-effects models using Eigen and R package version 0.999375-39*. Retrieved from <http://CRAN.R-project.org/package=lme4>
- Brennan, S. E., and Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 1482–1493. doi: 10.1037/0278-7393.22.6.1482
- Bürkle, B., Nirmaier, H., and Herrmann, T. (1986). "Von dir aus..." *Zur hörerbegleitenden Referenz. Bericht Nr. 10, Arbeiten der Forschergruppe "Sprechen und Sprachverstehen im sozialen Kontext", Heidelberg/Mannheim*. Retrieved

- from <http://www.psychologie.uni-heidelberg.de/institutsberichte/FG/FG10.pdf>
- Carlson, L. A. (1999). Selecting a reference frame. *Spat. Cogn. Comput.* 1, 365–379. doi: 10.1023/A:1010071109785
- Carlson, L. A., and Van Deman, S. R. (2008). Inhibition within a reference frame during the interpretation of spatial language. *Cognition* 106, 384–407. doi: 10.1016/j.cognition.2007.03.009
- Carlson-Radvansky, L. A., and Irwin, D. E. (1993). Frames of reference in vision and language: where is above? *Cognition* 46, 223–244. doi: 10.1016/0010-0277(93)90011-J
- Carlson-Radvansky, L. A., and Radvansky, G. A. (1996). The influence of functional relations on spatial term selection. *Psychol. Sci.* 7, 56–60. doi: 10.1111/j.1467-9280.1996.tb00667.x
- Clark, H. H. (1996). *Using Language*, Vol. 4. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511620539
- de Ruiter, J. P., Bangerter, A., and Dings, P. (2012). The interplay between gesture and speech in the production of referring expressions: investigating the tradeoff hypothesis. *Top. Cogn. Sci.* 4, 232–248. doi: 10.1111/j.1756-8765.2012.01183.x
- Duran, N. D., Dale, R., and Kreuz, R. J. (2011). Listeners invest in an assumed other's perspective despite cognitive cost. *Cognition* 121, 22–40. doi: 10.1016/j.cognition.2011.06.009
- Flavell, J. H. (1986). The development of children's knowledge about the appearance-reality distinction. *American Psychologist* 41, 418–425. doi: 10.1037/0003-066X.41.4.418
- Franklin, N., and Tversky, B. (1990). Searching imagined environments. *J. Exp. Psychol. Gen.* 119, 63–76. doi: 10.1037/0096-3445.119.1.63
- Friederici, A. D., and Levelt, W. J. M. (1990). Spatial reference in weightlessness: perceptual factors and mental representation. *Percept. Psychophys.* 47, 253–266. doi: 10.3758/BF03205000
- Galati, M., Michael, C., Mello, C., Greenauer, N. M., and Avraamides, M. N. (2013). The conversational partner's perspective affects spatial memory and descriptions. *J. Mem. Lang.* 68, 140–159. doi: 10.1016/j.jml.2012.10.001
- Garrod, S., and Anderson, A. (1987). Saying what you mean in dialogue: a study in conceptual and semantic coordination. *Cognition* 27, 181–218. doi: 10.1016/0010-0277(87)90018-7
- Grabowski, J., and Miller, G. (2000). Factors affecting the use of dimensional prepositions in German and American English: object orientation, social context, and prepositional pattern. *J. Psycholinguist. Res.* 29, 517–553. doi: 10.1023/A:1005124210205
- Graf, R., and Herrmann, T. (1989). *Zur sekundären Raumreferenz: Gegenüberobjekte bei nicht-kanonischer Betrachtungsposition*. Bericht Nr. 11, Arbeiten aus dem Sonderforschungsbereich 245 "Sprechen und Sprachverstehen im sozialen Kontext", Heidelberg/Mannheim.
- Grice, H. P. (1975). "Logic and conversation," in *Syntax and Semantics 3: Speech Arts*, ed. P. Cole (Amsterdam: Elsevier Science Ltd), 41–58.
- Haywood, S. L., Pickering, M. J., and Branigan, H. P. (2005). Do speakers avoid ambiguities during dialogue? *Psychol. Sci.* 16, 362–366. doi: 10.1111/j.0956-7976.2005.01541.x
- Herrmann, T. (1988). *Partnerbezogene Objektlokalisierung – ein neues sprachpsychologisches Forschungsthema*. Arbeiten der Forschergruppe "Sprechen und Sprachverstehen im sozialen Kontext", Bericht Nr. 25.
- Johannsen, K., and de Ruiter, J. P. (2013). The role of scene type and priming in the processing and selection of a spatial frame of reference. *Front. Psychol.* 4:182. doi: 10.3389/fpsyg.2013.00182
- Kessler, K., and Rutherford, H. (2010). The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions. *Front. Psychol.* 1:213. doi: 10.3389/fpsyg.2010.00213
- Levelt, W. J. (1982). "Cognitive styles in the use of spatial direction terms," in *Speech, Place and Action*, eds R. Jarvella and W. Klein (Chichester: John Wiley & Sons), 251–268.
- Levelt, W. J. (1989). *Speaking: From Intentions to Articulation*. Cambridge: MIT Press.
- Levelt, W. J. M. (1984). "Some perceptual limitations on talking about space," in *Limits in Perception*, eds A. J. van Doorn, W. de Grind, and J. Koenederink (Utrecht: VNU Science Press), 323–358.
- Levinson, S. C. (1996). "Frames of reference and Molyneux's question: crosslinguistic evidence," in *Language and Space*, 1st Edn, eds P. Bloom, M. A. Peterson, L. Nadel, and M. F. Garrett (Cambridge: MIT Press), 109–170.
- Levinson, S. C. (2003). *Space in Language and Cognition: Explorations in Cognitive Diversity*, Vol. 5. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511613609
- Linde, C., and Labov, W. (1975). Spatial networks as a site for the study of language and thought. *Lang. Cogn. Process.* 51, 924–939.
- Logan, G. D., and Sadler, D. D. (1996). "A computational analysis of the apprehension of spatial relations," in *Language and Space*, eds P. Bloom, M. Peterson, L. Nadel, and M. Garrett (Cambridge: MIT Press), 493–529.
- Miller, G. A., and Johnson-Laird, P. N. (1976). *Language and Perception*. Cambridge: Belknap Press of Harvard Univ. Press.
- Pickering, M. J., and Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behav. Brain Sci.* 27, 169–226. doi: 10.1017/S0140525X04000056
- Quené, H., and van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects, and with binomial data. *J. Mem. Lang.* 59, 413–425. doi: 10.1016/j.jml.2008.02.002
- R Core Team. (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>
- Schober, M. (1993). Spatial perspective-taking in conversation. *Cognition* 47, 1–24. doi: 10.1016/0010-0277(93)90060-9
- Taylor, H. A., Naylor, S. J., Faust, R. R., and Holcomb, P. J. (1999). "Could you hand me those keys on the right?" Disentangling spatial reference frames using different methodologies. *Spat. Cogn. Comput.* 1, 381–397. doi: 10.1023/A:1010035613419
- Tversky, B. (1996). "Spatial perspective in descriptions," in *Language and Space*, 1st Edn, eds P. Bloom, M. A. Peterson, L. Nadel, and M. F. Garrett (Cambridge: MIT Press), 463–492.
- Watson, M. E., Pickering, M. J., and Branigan, H. P. (2004). "Alignment of reference frames in dialogue," in *Proceedings of the 26th Annual Conference of the Cognitive Science Society*, Chicago (Mahwah: Erlbaum), 2353–2358.
- Watson, M. E., Pickering, M. J., and Branigan, H. P. (2006). "An empirical investigation into spatial reference frame taxonomy using dialogue," in *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (Mahwah: Erlbaum).
- Ziegler, L., Johannsen, K., Swadzba, A., de Ruiter, J. P., and Wachsmuth, S. (2012). Exploiting spatial descriptions in visual scene analysis. *Cogn. Process.* 13, 369–374. doi: 10.1007/s10339-012-0460-1

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 28 June 2013; accepted: 24 September 2013; published online: 16 October 2013.

Citation: Johannsen K and De Ruiter JP (2013) Reference frame selection in dialog: priming or preference? *Front. Hum. Neurosci.* 7:667. doi: 10.3389/fnhum.2013.00667

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2013 Johannsen and De Ruiter. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



What drives successful verbal communication?

Miriam de Boer^{1*}, Ivan Toni¹ and Roel M. Willems^{1,2}

¹ Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, Netherlands

² Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands

Edited by:

Klaus Kessler, University of Glasgow, UK

Reviewed by:

Ulrich Pfeiffer, University Hospital Cologne, Germany

Dale J. Barr, University of Glasgow, UK

*Correspondence:

Miriam de Boer, Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, P.O. Box 9101, 6500 HB, Nijmegen, Netherlands
e-mail: miriam.deboer@donders.ru.nl

There is a vast amount of potential mappings between behaviors and intentions in communication: a behavior can indicate a multitude of different intentions, and the same intention can be communicated with a variety of behaviors. Humans routinely solve these many-to-many referential problems when producing utterances for an Addressee. This ability might rely on social cognitive skills, for instance, the ability to manipulate unobservable summary variables to disambiguate ambiguous behavior of other agents ("mentalizing") and the drive to invest resources into changing and understanding the mental state of other agents ("communicative motivation"). Alternatively, the ambiguities of verbal communicative interactions might be solved by general-purpose cognitive abilities that process cues that are incidentally associated with the communicative interaction. In this study, we assess these possibilities by testing which cognitive traits account for communicative success during a verbal referential task. Cognitive traits were assessed with psychometric scores quantifying motivation, mentalizing abilities, and general-purpose cognitive abilities, taxing abstract visuo-spatial abilities. Communicative abilities of participants were assessed by using an on-line interactive task that required a speaker to verbally convey a concept to an Addressee. The communicative success of the utterances was quantified by measuring how frequently a number of Evaluators would infer the correct concept. Speakers with high motivational and general-purpose cognitive abilities generated utterances that were more easily interpreted. These findings extend to the domain of verbal communication the notion that motivational and cognitive factors influence the human ability to rapidly converge on shared communicative innovations.

Keywords: communication, language, individual differences, mentalizing, Raven's progressive matrices

INTRODUCTION

Daily human communication is surprisingly effective, even though it involves producing and understanding utterances that are inherently ambiguous. The potential mapping between behavior and intentions in communication is very large and many-to-many, such that similar behaviors can indicate different intentions and vice versa.

The ability of humans to map behavior to intentions has been labeled *interactive intelligence* (Levinson, 2006) and might be supported by motivational factors and cognitive abilities. The cognitive abilities implicated in understanding the intentions, feelings or thoughts of others, are often labeled as Theory of Mind or mentalizing abilities (Premack and Woodruff, 1978; Baron-Cohen and Wheelwright, 2004; Frith and Frith, 2012). Motivational factors refer to the drive to invest resources to understand another individual, the willingness and motivation to spend energy understanding the mental states of others (Levinson, 2006; Tomasello, 2009). In an alternative account, it is proposed that most of the time interlocutors would not have to infer the mental state of the other's mind at all. Automatic alignment of representations of the other's message-meaning mapping by tight coupling of production and comprehension (Pickering and Garrod, 2004) or the many cues generated during interaction (Shintel and Keysar, 2009) would suffice. Under most circumstances, no specific mentalizing skills would be needed to solve the many-to-many mapping problem.

In this perspective, communicative coordination relies on general-purpose cognitive abilities, as if communication would be similar to complex problem solving. The latter account gets credibility from the finding that, considering the speed of human communication, mentalizing as the only strategy to solve the multi-mapping problem is implausible as it would require extensive cognitive and temporal resources (Shintel and Keysar, 2009; Lin et al., 2010).

Here, we test whether motivational factors, mentalizing abilities, or general cognitive abilities in speakers predict successful tailoring of a message in a verbal communication game. For instance, an agent might have extremely sophisticated computational abilities and be able to store/retrieve a very large set of behavior/meaning mappings, but fail to do anything if not motivated to communicate, or fail to adjust a sophisticated behavior/meaning mapping to an Addressee and make it comprehensible. Different cognitive abilities involved in human communication might be differently sensitive to the expression of psychological traits across a group of individuals (Baron-Cohen et al., 2005; De Ruiter et al., 2010). Individual variation can help us understand the general principles of human communication (Levinson and Gray, 2012). In this study we investigate which psychometric scores indexing motivational factors and cognitive abilities, contribute most to a Communicator's success.

Previous research investigated individual sources of variation in subject pairs engaged in a non-verbal communication game

(Volman et al., 2012). The design in that study focussed on how pairs of Communicators establish communicative strategies, and how inter-subject differences influence communicative success. Communicators' motivation to solve complex tasks, as indexed by the Need for Cognition Scale (NCS; Cacioppo et al., 1996), predicted communicative success. General intelligence of the Addressees, as indexed by the Raven's advanced progressive matrices (RAPM; Raven et al., 1995) accounted for higher accuracy scores. Although attribution of mental states to another person (mentalizing) seems an important capacity for creating a new communication system that both Communicator and Addressee can comprehend, the speed and success with which such a new communicative system was established could not be explained by the participants' score on the empathy quotient (EQ; Baron-Cohen and Wheelwright, 2004), or a similar measure for empathy, the interpersonal reactivity index (IRI; Davis, 1983). In a related study using the same non-verbal communication game, the magnitude of communicative *adjustments* to a presumed Addressee was explained by the EQ (Newman-Norlund et al., 2009). Senders high in empathy put greater emphasis on crucial communicative elements when they believed their Addressee was a child compared to when they believed their Addressee was an adult. In contrast, individuals with high motivation for complex problems (NCS; Cacioppo et al., 1996) were less likely to adapt their communicative behavior toward their Addressee.

The picture that emerges from those studies on non-verbal communication systems is that empathic traits may be beneficial for *adapting* communicative behavior to another individual. In contrast, the ability to *generate* effective communicative acts might be mainly influenced by the motivation and ability to solve complex problems.

Here, we tested the role of trait variables on the ability to generate successful communicative interaction in the verbal domain by indexing individual differences in empathizing (IRI and EQ, respectively Davis, 1983; Baron-Cohen and Wheelwright, 2004) Need for Cognition Scale (NCS, Cacioppo et al., 1984), general intelligence (RAPM, Raven et al., 1995) and verbal intelligence (Groninger Intelligentie Test Matrix Reasoning, and the Wexler Adult Intelligence Scale Similarity and Vocabulary subscale, respectively Kooreman and Luteijn, 1987; WAIS-III, 1997). Abstract visuo-spatial abilities were indexed as part of the RAPM (Carpenter et al., 1990; Mackintosh and Bennett, 2005). We will examine how these factors in the Communicator contribute to successful communication, that is, generating an accurate and easy interpretable message for an Addressee (see Ickes et al., 2000 on the role of motivation on empathic accuracy in observers) in the context of an interactive word game.

In the interactive word game, both communicative setting and linguistic difficulty were independently manipulated. We used a paradigm called the Taboo game (Willems et al., 2010) where a Communicator had to describe a Target-word (e.g., "Beard") to an Addressee in one sentence without using Taboo-words (e.g., "man," "shave," "hair," "chin" and "mustache"; see Figure 1A). An indication of the Target-word description's communicative success was obtained by evaluation of these utterances by a new group of subjects (labeled as Evaluators, see Figure 1B). The data reported

in this manuscript relates the performance of these Evaluators to the psychometric scores of Communicators. We predict to find a similar pattern as described above: not mentalizing abilities *per se*, but the motivation or general cognitive ability to solve complex tasks will account for effective communication in an existing verbal communication system. This study aims to open the way for understanding variations in visual perspective-taking abilities during social interactions. Accordingly, we pay particular attention to the RAPM as an index of visuo-spatial abilities (Carpenter et al., 1990; Mackintosh and Bennett, 2005).

MATERIAL AND METHODS

SUBJECTS

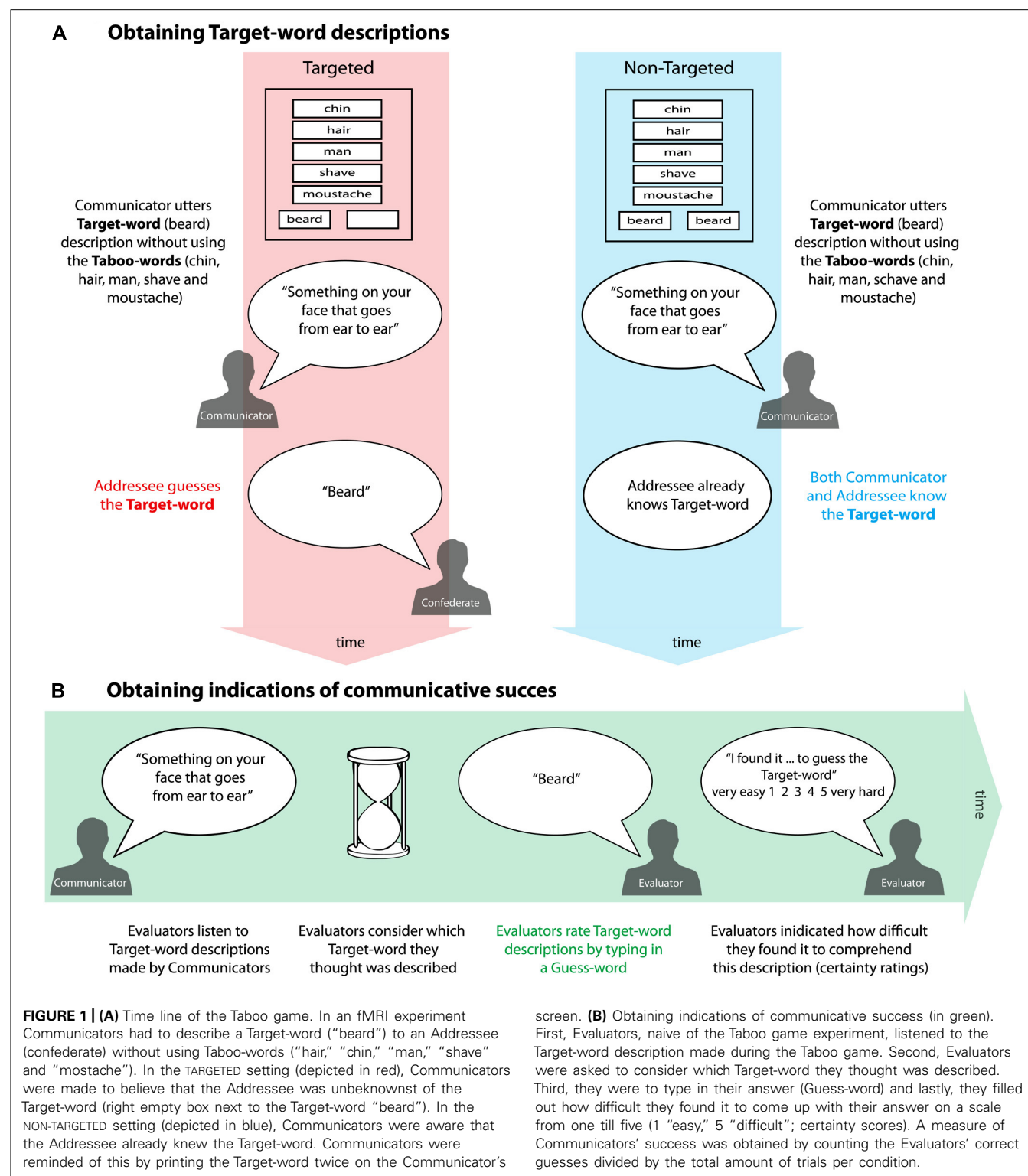
Sixteen participants (labeled as Communicators, four males, mean age = 21 years old, SD = 3 years) played the Taboo game in the context of an fMRI experiment (for further details, see Willems et al., 2010) and completed several psychometric tests. All had Dutch as their mother-tongue, and did not have a known neurological history, hearing problems, dyslexia, stuttering or other language-related problems. In a separate experimental session, sixteen subjects naive to the Taboo game evaluated the Target-word descriptions generated by the Communicators. These Evaluators (four males, mean age = 20 years old, SD = 3 years) did not have language, hearing or eyesight difficulties and had Dutch as their mother tongue. The data reported in this manuscript relates the performance of the Evaluators to the psychometric scores of Communicators.

PROCEDURE

Description from Communicators

Experimental material was obtained in the context of an fMRI study (for further details, see Willems et al., 2010). Communicators generated descriptions for a confederate (referred to as Addressee) after which we obtained their psychometric scores on various cognitive abilities and motivational factors (for details of the acquisition of the Communicators' psychometric scores, see Psychometric indexes of individual cognitive abilities of Communicators). In a separate study, a group of new participants labeled as Evaluators rated these descriptions' communicative success.

Communicators made descriptions of 60 concrete nouns (Target-words). They would for instance have to describe the Target-word "beard" without using five so called Taboo-words "hair," "chin," "man," "shave" and "mustache" (see Figure 1A). Communicator and Addressee could clearly hear each other's utterances via MR (Magnetic Resonance) compatible headphones, with the Addressee inferring the Target-word that the Communicator described. Since the Communicator was lying in the MR scanner, we filtered out scanner noise using the audacity noise reduction function (Audacity from <http://audacity.sourceforge.net/>) to increase the audibility of the Target-word descriptions. Descriptions lasted on average 5.14 s (SD = 0.68 s). In the Taboo game, two factors were manipulated: communicative setting and linguistic difficulty. Communicative setting was manipulated by changing the Communicator's belief of the Addressee's knowledge of the Target-word. In the TARGETED setting the Communicator generated the description for a specific other (a confederate), who gave wrong answers on a



prescribed set of trials (30% of the trials). In case of a wrong trial, Communicators were asked to generate a new Target-word description consecutively. These repeated trials were not rated by the Evaluators. In the NON-TARGETED setting, it was explained to Communicators that the Addressee was already aware

of the Target-word and that this person was only overhearing the Communicator's Target-word description. Communicators were reminded that the Addressee already knew the Target-word by printing the Target-word twice on the Communicator's screen (see Figure 1A). Linguistic difficulty was manipulated by varying

the semantic distance between Target-word and Taboo-words. During EASY trials, Communicators described Target-words without using Taboo-words that were loosely semantically related to the Target-word (e.g., Target-word “rainbow,” Taboo-words: “four-leaf-clover,” “violet,” “water,” “sound,” “fairy-tale”). During DIFFICULT trials, Communicators described Target-words without using Taboo-words that were closely semantically related to the Target-word (such as the “beard” example above).

During the TARGETED and the NON-TARGETED setting, half of the trials were EASY, and half of the trials were DIFFICULT. Lexical frequency of Taboo and Target-words was matched between all conditions (CELEX database, Baayen et al., 1995). Stimulus lists were pseudo-randomized in two sets such that participants did not describe the same Target-words in TARGETED and NON-TARGETED trials. Half of the Communicators described Target-words of set A in the TARGETED setting and Target-words of set B in the NON-TARGETED setting. The other half of the Communicators described Target-words in the opposite settings, meaning set B in the TARGETED setting and set A in the NON-TARGETED setting. More Communicators completed Set A during the TARGETED setting. To prevent Evaluators from hearing certain Target-word descriptions more often generated in the TARGETED or the NON-TARGETED setting, four out of the twenty Communicators of the original Taboo game experiment were excluded at random. With sixty Target-word descriptions of sixteen Communicators; there were a total of 960 unique Target-word descriptions.

Evaluators

In the current experiment, a new group of subjects evaluated these Target-word descriptions from the Willems et al. (2010) study to obtain an indication of the Communicator’s communicative success. After reading a written instruction, Evaluators completed three practice trials not used in the remainder of the experiment, and then performed the actual task in two blocks of approximately thirty minutes each. Trials were separated in different phases (see **Figure 1B**). At first, a black screen was presented in which a fixation cross appeared. The Evaluators heard a Target-word description made by one of the Communicators, e.g., “Something on your face that goes from ear to ear.” Evaluators planned their response with a cut-off time of twenty seconds and typed which Target-word they thought was described (Guess-word). Thereafter, Evaluators were asked to give a score from one to five on how difficult they found it to generate their answer with “1” meaning that they found this very difficult and “5” meaning that they found this very easy (from now on referred to as “certainty score”). After a randomized intertrial interval (mean = 4.5 s, SD = 0.93 s), the next trial was presented. The experiment was performed using Presentation software (Version 10.2, www.neurobs.com) and presented on a laptop computer via earphones. Stimulus presentation was pseudo-randomized such that each Communicator’s Target-word description was rated by two different Evaluators. In total, each Evaluator heard a total of 120 unique Target-word descriptions, eight from the same Communicator: two recorded during the TARGETED EASY condition, two recorded during the TARGETED DIFFICULT condition, two during the NON-TARGETED EASY and two during the NON-TARGETED DIFFICULT condition. Descriptions of the same Communicator or the same Target-word would never

be presented in immediate succession; neither would Evaluators hear a description of a particular Target-word more than once per block. For instance, in the first block, Evaluators would hear a recording of a Target-word description of “beard” by Communicator A, and in the second block they would hear a recording of a Target-word description of “beard” by Communicator B.

Psychometric indexes of individual cognitive abilities of Communicators

After playing the Taboo game, each Communicator completed psychometric tests to characterize their empathizing abilities (IRI and EQ, respectively Davis, 1983; Baron-Cohen and Wheelwright, 2004), motivation for complex tasks (NCS, Cacioppo et al., 1984), general intelligence (RAPM, Raven et al., 1995) and verbal intelligence (GIT matrices, WAIS Similarity and WAIS Vocabulary subscale, respectively Kooreman and Luteijn, 1987; WAIS-III, 1997). Since the focus of our paper was on the Communicator, no psychometric indexes of cognitive abilities or motivational factors were taken from the Evaluators.

The EQ indexes both cognitive and affective empathy. It characterizes cognitive empathy (mentalizing), reactivity and social skill but is not correlated with social desirability (Baron-Cohen and Wheelwright, 2004; Lawrence et al., 2004). Instead of calculating one scale, empathy can also be indexed in four subscales as is done in the IRI (Davis, 1983). The Perspective Taking subscale indexes the ease with which one can take the point of view or perspective of the other. The Fantasy subscale indexes how easily somebody can identify himself/herself with a fictional character. There are two subscales of emotional reactions: the Empathic Concern subscale indexes feelings of compassion and warmth, while the Personal Distress subscale indexes the tendency to feel discomfort when observing another person in distress. Motivation to be engaged in complex tasks, such as we assume the Taboo game is, was indexed with the NCS (Cacioppo et al., 1984). The EQ, IRI and NCS are self-report Likert scale type questionnaires. All three questionnaires were completed with paper and pencil.

Raven’s advanced progressive matrices (Raven et al., 1995) index general intelligence. Two separate factors underlie performance on the RAPM. Part of the items are solved by verbal-analytical rules, whereas other items tend to be solved using visual-spatial rules (Carpenter et al., 1990; DeShon et al., 1995). Communicators had to complete as many of the 36 items (RAPM set II) as possible within twenty minutes. The Communicator’s RAPM score was calculated by adding up the number of correctly completed items within that time.

Communicators high in verbal intelligence may have a larger vocabulary and, due to their increased word reasoning skills, have easier access to alternatives for Taboo-words. The WAIS Vocabulary subscale (WAIS-III, 1997) indexes word understanding and how well this word understanding can be expressed. Participants are asked to give definitions of words that become increasingly more unfamiliar. Word reasoning skills were indexed by the Groninger Intelligence Test Matrix Reasoning subscale (GIT Matrix Reasoning, Kooreman and Luteijn, 1987). Participants are asked to solve analogies, such as “if table is to wood, stove is to iron, thus shoe is to. . .” During the WAIS Similarity subscale (WAIS-III, 1997), participants are asked to describe how common objects or

concepts are similar, e.g., “what is the similarity between a bike and a car?” All the verbal intelligence subscales were taken orally and scored according to prescribed standards (Kooreman and Luteijn, 1987; WAIS-III, 1997).

Communicative success

Our measure of communicative success was based on the correct guesses of the Evaluators divided by the total amount of trials per condition. In the following cases, we rated the Evaluators' guesses as correct: if the Guess-word had exactly the same word form as the Taboo word, if the Guess-word was a compound instead of a head, or vice versa (for example “woonwijk” or “wijk” meaning “living district” and “neighborhood”), if it was a synonym (“leunstoel” by “fauteuil,” meaning “armchair” and “lounge chair”), or if it was a diminutive (e.g., “munt” by “muntje” meaning “coin” and “little coin”). In this manner, we were able to consider successful communication of word *meaning*.

STATISTICAL ANALYSIS

Accuracy and certainty scores of Evaluators were analyzed using a 2×2 within subjects ANOVA with factors setting (TARGETED and NON-TARGETED) and linguistic difficulty (EASY and DIFFICULT). First, to assess which psychometric indexes explained variance in description quality, we performed a regression analysis with communicative success in the TARGETED setting as a dependent variable. Second, to correct for the individual differences in general performance on the Taboo game, a second analysis was conducted comparing the TARGETED to the NON-TARGETED setting by subtracting the communicative success scores obtained from the TARGETED and the NON-TARGETED setting. Third, regression analyses were conducted to investigate which cognitive traits explained communicative success during our manipulation of linguistic difficulty (DIFFICULT, EASY and EASY subtracted from DIFFICULT). In each regression analysis, the Communicators' psychometric scores on all tests were entered as independent regressors in a stepwise fashion: a variation on the forward algorithm. Only those independent factors whose contribution was unique and significant were entered in the model ($p < 0.05$), while at each subsequent search step redundant factors were removed. Since questionnaires indexing the same cognitive ability may potentially correlate, e.g., mentalizing ability was indexed by both the EQ and the IRI), we considered whether predictors correlated strongly with one another, but Pearson's correlation coefficients were <0.8 across regressors. Only independent variables explaining unique variance are reported. All statistical analyses were conducted with IBM SPSS Statistics for Windows (Version 19.0).

RESULTS

REACTION TIMES, CERTAINTY RATINGS AND ACCURACY SCORES

Evaluators on average took 2.5 s (SD = 0.5 s) to generate a Guess-word. Evaluators found the task rather difficult (mean certainty rating = 2.25, SD = 0.29, 1–5 scale). However, Evaluators comprehended the Communicators' Target-word descriptions well (mean percentage correct = 73%, SD = 5%, minimum score 62% and maximum 83%). There was no interaction in reaction times, certainty ratings, or accuracy scores between communicative setting (TARGETED,

NON-TARGETED) and difficulty (EASY, DIFFICULT), neither was there a main effect of setting (TARGETED, NON-TARGETED). Evaluators planned shorter, were more certain and more accurate for Target-word descriptions made in the EASY condition (for statistics see Table 1).

COMMUNICATIVE SUCCESS AND INDIVIDUAL DIFFERENCES

Only those regressors explaining a statistically significant portion of variance are described here (for statistics see Table 2). Communicative success during the TARGETED setting was positively driven by the Communicators' motivation to solve complex tasks as indexed by the NCS (Table 2, see Figure 2A). No such effect was observed during the NON-TARGETED trials. Indexes of empathy (IRI, EQ) did not account significantly for variance in performance.

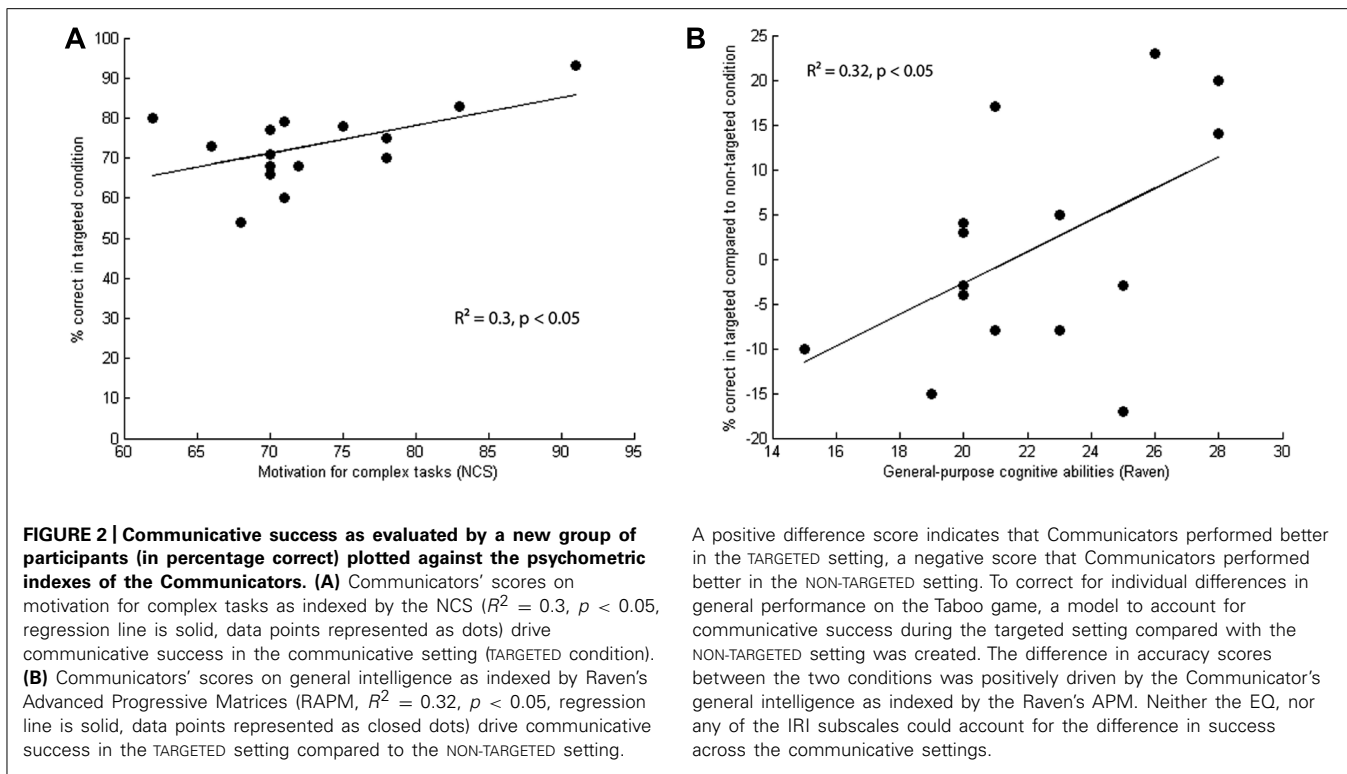
Table 1 | Repeated measures analysis of variance was applied on reaction times, certainty ratings and accuracy scores of Evaluators when listening to Target-word descriptions made by Communicators in an earlier conducted fMRI experiment.

	F(df)	MSe	p
Reaction times			
Communication	<1 (1,15)	32.2	0.73
Linguistic difficulty	11.25 (1,15)	30.66	<0.01
Communication × linguistic difficulty	1.48 (1,15)	53.37	0.24
Certainty ratings			
Communication	2.78 (1,15)	0.05	0.12
Linguistic difficulty	11.75 (1,15)	0.06	<0.01
Communication × linguistic difficulty	1.53 (1,15)	0.07	0.24
Accuracy			
Communication	<1 (1,15)	0.02	0.87
Linguistic difficulty	7.45 (1,15)	0	<0.05
Communication × linguistic difficulty	<1 (1,15)	0.01	0.91

The model contained the factors communicative setting (descriptions that Evaluators listened to were made in the TARGETED or the NON-TARGETED setting) and linguistic difficulty (descriptions were made in the EASY or DIFFICULT condition). Evaluators planned shorter ($F(1,15) = 11.25$, $p < 0.01$), were more certain ($F(1,15) = 11.75$, $p < 0.01$) and more accurate ($F(1,15) = 7.45$, $p < 0.05$) for Target-word description made in the easy condition.

Table 2 | Overview of psychometric indexes significantly accounting for communicative success in the different experimental conditions.

Experimental condition	Psychometric index	Beta	F(df)	R ²	p
Targeted	NCS	0.54	5.86 (1,14)	0.3	<0.05
Targeted - non-targeted	Raven's APM	0.56	6.43 (1,14)	0.32	<0.05
Difficult	WAIS vocabulary	0.65	9.98 (2,13)	0.61	<0.01
	IRI personal distress	0.46			
Easy	NCS	0.55	8.4 (2,13)	0.56	<0.05
	IRI personal distress	0.51			



To correct for individual differences in general performance on the Taboo game, a model to account for communicative success during the TARGETED setting compared with the NON-TARGETED setting was created. The difference in accuracy scores between the two conditions was positively driven by the Communicator's general intelligence as indexed by the Raven's APM (see Table 2, Figure 2B). Neither the EQ, nor any of the IRI subscales could account for the difference in success across the communicative settings.

Verbal abilities as indexed with the WAIS vocabulary subscale positively accounted for communicative success during DIFFICULT trials (collapsed across TARGETED and NON-TARGETED settings). Furthermore, the Communicator's score on the IRI personal distress subscale, which indexes the tendency to feel discomfort when observing somebody else's distress, was predictive of accuracy scores on DIFFICULT trials. For EASY trials, the same subscale (IRI personal distress) and the Communicator's NCS positively accounted for communicative success. None of the psychometric indexes explained variance of communicative success in DIFFICULT compared to EASY trials.

DISCUSSION

We have employed inter-subject differences in trait parameters and communicative performance to examine whether motivational factors, mentalizing skills, or general-purpose cognitive abilities preferentially accounted for communicative success. In an interactive verbal communication task, participants (Communicators) were asked to describe concepts without using a number of semantically related words (Willems et al., 2010). Successful communication was quantified by how frequently a

A positive difference score indicates that Communicators performed better in the TARGETED setting, a negative score that Communicators performed better in the NON-TARGETED setting. To correct for individual differences in general performance on the Taboo game, a model to account for communicative success during the targeted setting compared with the NON-TARGETED setting was created. The difference in accuracy scores between the two conditions was positively driven by the Communicator's general intelligence as indexed by the Raven's APM. Neither the EQ, nor any of the IRI subscales could account for the difference in success across the communicative settings.

group of new participants (Evaluators) would infer the correct concept. We found that motivational factors, as indexed by the Communicator's motivation to solve complex tasks (NCS), were positively driving successful communication in a communicative ("TARGETED") setting. These findings extend previous observations (Volman et al., 2012) to the domain of verbal communication, to show the importance of motivational factors in communicative behavior. Communicators high in need for cognition may make more effort to select the message/meaning mapping that is best comprehensible. They may be more flexible in finding alternatives, if the solution they generated turned out to be incomprehensible for their Addressee (Cacioppo et al., 1984; Evans et al., 2003). However, need for cognition did not explain variance in communicative success, when we directly compared the TARGETED versus the NON-TARGETED settings. That is, need for cognition was important in explaining performance during the communicative (TARGETED) trials overall, but not when directly comparing TARGETED versus NON-TARGETED trials. Comparing TARGETED versus NON-TARGETED settings directly revealed that communicative success was significantly predicted by Communicators' general-purpose cognitive ability as indexed by Raven's APM (Raven et al., 1995). A Communicator's high general intelligence may be beneficial for the generation of efficient messages in several ways. It may help storage of speaker history (Horton and Gerrig, 2005; Shintel and Keysar, 2009; Galati and Brennan, 2010), executive control (Ybarra and Winkielman, 2012), and working memory capacity (Lin et al., 2010). This idea fits with recent evidence showing tightly matched neural dynamics in subjects solving communicative and rule-based solo problems (Stolk et al., 2013).

From our findings, we can only speculate as to whether Communicator's success in this communication game is driven by general cognitive abilities, or more specifically by visuo-spatial abilities. Research on the underlying cognitive processes of the RAPM has suggested that some of Raven's matrices are solved using a visuo-spatial strategy (Carpenter et al., 1990; DeShon et al., 1995) for an alternative view see (Plaisted et al., 2011). This abstract visuo-spatial ability may positively drive effective search of alternatives for words that cannot be used to generate the Target-word description (Taboo-words). Communicators with a high RAPM score may be more skilful in finding words that can be easily interpreted by the Addressee, and as a consequence, be more effective in solving the message-to-meaning problem.

Given that the communication task used in this study relied on verbal material, it might appear surprising that the psychometric indexes of verbal ability (GIT or WAIS subscales, Kooreman and Luteijn, 1987; WAIS-III, 1997) did not significantly account for variation in communicative success. Yet, the verbal intelligence of the Communicator (WAIS) was important for solving trials where the Taboo-words were closely semantically related to the Target-word (DIFFICULT trials). This may be an indication that linguistic abilities accounted for communicative success in semantically difficult trials in general, but not for communicative trials specifically. These findings support the notion of a cognitive difference between linguistic and communicative abilities (Willems and Varley, 2010; Willems et al., 2011).

Importantly, mentalizing abilities, as indexed by general cognitive empathy, emotional reactivity, social skill (EQ, Lawrence et al., 2004) or as indexed by the Perspective Taking, Fantasy, Empathic Concern and Personal Distress subscales (IRI; Davis, 1983), were also not significantly related to communicative success as a function of the communicative setting. Yet, a Communicator's personal distress was important for solving trials where Taboo-words were closely semantically related to the Target-word (DIFFICULT trials). This result is not immediately compatible with the idea that mentalizing abilities are important for generating a comprehensible message. However, this does not preclude the possibility that mentalizing abilities are important for implementing communicative *adjustments* toward a specific Addressee, as previously shown in the context of non-verbal communication (Newman-Norlund et al., 2009). Nor does it preclude that mentalizing abilities are employed in communicative task settings. As a matter of fact, the fMRI data of the study from which our materials were taken, shows that participants activate mentalizing related brain areas when designing a communicative message for a specific other (Willems et al., 2010). The present findings add to this that the *individual differences* in mentalizing abilities are not indicative of communicative success, but this obviously does not mean that such abilities are not used in communication.

The current study is a first step in the direction to point out the role of motivational factors and cognitive abilities on verbal communicative success. Given that the main experiment was performed in an MR environment, the interaction was quite rigidly structured and, as a consequence, not all constituents of social interaction (De Jaegher et al., 2010; Schilbach et al.,

2012) were present during the game. For instance, the role of Communicator and Addressee was fixed, and there was a maximum duration of the time interval during which Communicator and Addressee were allowed to speak. Our task *was* interactive in the sense that Communicators were actively engaged in our verbal interaction game. Interlocutors' performance depended on the clarity of the description of the Communicator and the comprehension of the Addressee. The interlocutors could to some extent monitor and adjust their behavior on the basis of feedback (correct or incorrect), and on the timing of the on-line interaction (e.g., time interval required by a Communicator to organize an utterance, and by an Addressee to reply). In this study, we focussed on the role of the Communicator. Future research should study the effect of cognitive abilities and motivational factors on *both* interlocutors and should investigate additional factors that could be of influence on communicative success, such as the role of motivation to engage in social interaction or the extent of the pre-existing common ground (e.g., strangers or close friends). Not only should these factors be studied at the individual level, but also on the "second person" level, the level that comes about *between* interactors (Becchio et al., 2010; De Jaegher et al., 2010; Schilbach et al., 2012).

More generally, our data speak to the observation that if a Communicator has a global idea of her Addressee, she may not always need to employ mentalizing abilities immediately or exclusively (Shintel and Keysar, 2009). As Zaki and Ochsner (2012) put it, in communication it is not *either* mentalizing, *or* general cognitive abilities, but more a question of "when/how" the one system is used and when/how the other system is used.

CONCLUSION

We have employed individual variation to examine whether motivational factors, mentalizing skills, or general cognitive abilities preferentially accounted for communicative success. We found that motivational factors ("need for cognition") and general-purpose cognitive abilities (Raven's matrices) were positively driving successful communication in an interactive communication task. These findings extend previous observations (Volman et al., 2012) to the domain of verbal communication and stress the importance of motivation and general-purpose cognitive abilities in communicative success. Mentalizing or empathy scores did not explain communicative success in the paradigm that we employed here. Future research should be directed toward understanding under which circumstances communicative behavior is most driven by motivational and general cognitive factors, and when differences in mentalizing abilities between individuals do make a difference.

ACKNOWLEDGMENTS

The authors thank Iris van Rooij and Mark Blokpoel for their valuable contribution to the interpretation of the data. This paper is supported by a grant from the European Union Joint-Action Science and Technology Project (IST-FP6-003747) and by a grant from the Netherlands Organisation for Scientific Research (VICI grant #453-08-002).

REFERENCES

- Baayen, H. R., Piepenbrock, R., and Gulikers, L. (1995). *The CELEX Lexical Database*, 2nd Edn. Philadelphia: Linguistic Data Consortium.
- Baron-Cohen, S., Knickmeyer, R. C., and Belmonte, M. K. (2005). Sex differences in the brain: implications for explaining autism. *Science* 310, 819–823. doi: 10.1126/science.1115455
- Baron-Cohen, S., and Wheelwright, S. (2004). The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *J. Autism Dev. Disord.* 34, 163–175. doi: 10.1023/B:JADD.0000022607.19833.00
- Becchio, C., Sartori, L., and Castiello, U. (2010). Towards you: the social side of actions. *Curr. Dir. Psychol. Sci.* 19, 183–188. doi: 10.1177/0963721410370131
- Cacioppo, J. T., Petty, R., and Kao, C. (1984). The efficient assessment of need for cognition. *J. Pers. Assess.* 48, 306–307. doi: 10.1207/s15327752jpa4803_13
- Cacioppo, J. T., Petty, R. E., Feinstein, J. A., Blair, W., and Jarvis, G. (1996). Dispositional differences in cognitive motivation: the life and times of individuals varying in need for cognition. *Psychol. Bull.* 119, 197–253. doi: 10.1037/0033-2909.119.2.197
- Carpenter, P. A., Just, M. A., and Shell, P. (1990). What one intelligence test measures: a theoretical account of the processing in the Raven progressive matrices Test. *Psychol. Rev.* 97, 404–431. doi: 10.1037/0033-295X.97.3.404
- Davis, M. H. (1983). Measuring individual-differences in empathy – evidence for a multidimensional approach. *J. Pers. Soc. Psychol.* 44, 113–126. doi: 10.1037/0022-3514.44.1.113
- De Jaegher, H., Di Paolo, E., and Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends Cogn. Sci.* 14, 441–447. doi: 10.1016/j.tics.2010.06.009
- De Ruiter, J. P., Noordzij, M. L., Newman-Norlund, S. E., Newman-Norlund, R. D., Hagoort, P., Levinson, S. C., et al. (2010). Exploring human interactive intelligence. *Interact. Stud.* 11, 51–77. doi: 10.1075/is.11.1.05rui
- DeShon, R. P., Chan, D., and Weissbein, D. A. (1995). Verbal overshadowing effects on Raven's advanced progressive matrices: evidence for multidimensional performance determinants. *Intelligence* 21, 135–155. doi: 10.1016/0160-2896(95)90023-3
- Evans, C., Kirby, J., and Fabrigar, L. (2003). Approaches to learning, need for cognition, and strategic flexibility among university students. *Br. J. Educ. Psychol.* 73, 507–528. doi: 10.1348/000709903322591217
- Frith, C. D., and Frith, U. (2012). Mechanisms of social cognition. *Annu. Rev. Psychol.* 287–313. doi: 10.1146/annurev-psych-120710-100449
- Galati, A., and Brennan, S. E. (2010). Attenuating information in spoken communication: for the speaker, or for the addressee? *J. Mem. Lang.* 62, 35–51. doi: 10.1016/j.jml.2009.09.002
- Horton, W. S., and Gerrig, R. J. (2005). The impact of memory demands on audience design during language production. *Cognition* 96, 127–142. doi: 10.1016/j.cognition.2004.07.001
- Ickes, W., Gesn, P. R., and Graham, T. (2000). Gender differences in empathic accuracy: differential ability or differential motivation? *Pers. Relat.* 7, 95–110. doi: 10.1111/j.1475-6811.2000.tb00006.x
- Kooreman, A., and Luteijn, F. (1987). Groninger Intelligentie Test GIT: Schriftelijke verkorte vorm.
- Lawrence, E., Shaw, P., Baker, D., Baron-Cohen, S., and David, A. (2004). Measuring empathy: reliability and validity of the Empathy Quotient. *Psychol. Med.* 34, 911–920. doi: 10.1017/S0033291703001624
- Levinson, S. C. (2006). "On the human interactional engine," in *Roots of Human Sociality: Culture, Cognition and Interaction*, eds N. J. Enfield and S. C. Levinson (New York: Berg), 39–69.
- Levinson, S. C., and Gray, R. D. (2012). Tools from evolutionary biology shed new light on the diversification of languages. *Trends Cogn. Sci.* 16, 167–173. doi: 10.1016/j.tics.2012.01.007
- Lin, S., Keysar, B., and Epley, N. (2010). Reflexively mindblind: using theory of mind to interpret behavior requires effortful attention. *J. Exp. Soc. Psychol.* 46, 551–556. doi: 10.1016/j.jesp.2009.12.019
- Mackintosh, N., and Bennett, E. (2005). What do Raven's matrices measure? An analysis in terms of sex differences. *Intelligence* 33, 663–674. doi: 10.1016/j.intell.2005.03.004
- Newman-Norlund, S. E., Noordzij, M. L., Newman-Norlund, R. D., Volman, I. A. C., De Ruiter, J. P., Hagoort, P., et al. (2009). Recipient design in tacit communication. *Cognition* 111, 46–54. doi: 10.1016/j.cognition.2008.12.004
- Pickering, M. J., and Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behav. Brain Sci.* 27, 169–190. doi: 10.1017/S0140525X04000056
- Plaisted, K., Bell, S., and Mackintosh, N. (2011). The role of mathematical skill in sex differences on Raven's matrices. *Pers. Individ. Diff.* 51, 562–565. doi: 10.1016/j.paid.2011.05.005
- Premack, D., and Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behav. Brain Sci.* 1, 515–526. doi: 10.1017/S0140525X00076512
- Raven, J., Raven, J. C., and Court, J. H. (1995). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. Oxford: Psychologists Press.
- Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., et al. (2012). Toward a second-person neuroscience. *Behav. Brain Sci.* 36, 393–414. doi: 10.1017/S0140525X12000660
- Shintel, H., and Keysar, B. (2009). Less is more: a minimalist account of joint action in communication. *Topics Cogn. Sci.* 1, 260–273. doi: 10.1111/j.1756-8765.2009.01018.x
- Stolk, A., Verhagen, L., Schoffelen, J. M., Oostenveld, R., Blokpoel, M., Hagoort, P., et al. (2013). Neural mechanisms of human communicative innovations. *Proc. Natl. Acad. Sci. U.S.A.* 110, 14574–14579. doi: 10.1073/pnas.1303170110
- Tomasello, M. (2009). *Why We Cooperate*. Cambridge, MA: MIT.
- Volman, I., Noordzij, M. L., and Toni, I. (2012). Sources of variability in human communicative skills. *Front. Hum. Neurosci.* 6:310. doi: 10.3389/fnhum.2012.00310
- WAIS-III. (1997). *WAIS-III Manual (Dutch version)*. Lisse, The Netherlands: Swets and Zeitlinger.
- Willems, R. M., Benn, Y., Hagoort, P., Toni, I., and Varley, R. (2011). Communicating without a functioning language system: implications for the role of language in mentalizing. *Neuropsychologia* 49, 3130–3135. doi: 10.1016/j.neuropsychologia.2011.07.023
- Willems, R. M., De Boer, M., De Ruiter, J. P., Noordzij, M. L., Hagoort, P., and Toni, I. (2010). A dissociation between linguistic and communicative abilities in the human brain. *Psychol. Sci.* 21, 8–14. doi: 10.1177/0956797609355563
- Willems, R. M., and Varley, R. (2010). Neural insights into the relation between language and communication. *Front. Hum. Neurosci.* 4:203. doi: 10.3389/fnhum.2010.00203
- Ybarra, O., and Winkielman, P. (2012). On-line social interactions and executive functions. *Front. Hum. Neurosci.* 6:75. doi: 10.3389/fnhum.2012.00075
- Zaki, J., and Ochsner, K. (2012). The neuroscience of empathy: progress, pitfalls and promise. *Nat. Neurosci.* 15, 675–680. doi: 10.1038/nn.3085

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 29 June 2013; accepted: 09 September 2013; published online: 01 October 2013.

Citation: de Boer M, Toni I and Willems RM (2013) What drives successful verbal communication? *Front. Hum. Neurosci.* 7:622. doi: 10.3389/fnhum.2013.00622

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2013 de Boer, Toni and Willems. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Flexible spatial perspective-taking: conversational partners weigh multiple cues in collaborative tasks

Alexia Galati^{1*} and Marios N. Avraamides^{1,2*}

¹ Department of Psychology, University of Cyprus, Nicosia, Cyprus

² Centre for Applied Neuroscience, University of Cyprus, Nicosia, Cyprus

Edited by:

Sarah H. Creem-Regehr, University of Utah, USA

Reviewed by:

Jan M. Wiener, Bournemouth University, UK

Jennifer M. Roche, University of Rochester, USA

*Correspondence:

Alexia Galati and Marios N. Avraamides, Department of Psychology, University of Cyprus, 65 Kallipoleos Avenue, PO Box 20537, 1678 Nicosia, Cyprus
e-mail: galati@ucy.ac.cy; mariosav@ucy.ac.cy

Research on spatial perspective-taking often focuses on the cognitive processes of isolated individuals as they adopt or maintain imagined perspectives. Collaborative studies of spatial perspective-taking typically examine speakers' linguistic choices, while overlooking their underlying processes and representations. We review evidence from two collaborative experiments that examine the contribution of social and representational cues to spatial perspective choices in both language and the organization of spatial memory. Across experiments, speakers organized their memory representations according to the convergence of various cues. When layouts were randomly configured and did not afford intrinsic cues, speakers encoded their partner's viewpoint in memory, if available, but did not use it as an organizing direction. On the other hand, when the layout afforded an intrinsic structure, speakers organized their spatial memories according to the person-centered perspective reinforced by the layout's structure. Similarly, in descriptions, speakers considered multiple cues whether available a priori or at the interaction. They used partner-centered expressions more frequently (e.g., "to your right") when the partner's viewpoint was misaligned by a small offset or coincided with the layout's structure. Conversely, they used egocentric expressions more frequently when their own viewpoint coincided with the intrinsic structure or when the partner was misaligned by a computationally difficult, oblique offset. Based on these findings we advocate for a framework for flexible perspective-taking: people weigh multiple cues (including social ones) to make attributions about the relative difficulty of perspective-taking for each partner, and adapt behavior to minimize their collective effort. This framework is not specialized for spatial reasoning but instead emerges from the same principles and memory-dependent processes that govern perspective-taking in non-spatial tasks.

Keywords: perspective-taking, spatial memory, intrinsic structure, audience design, common ground, spatial descriptions

SPATIAL PERSPECTIVE-TAKING IN COLLABORATIVE TASKS

When coordinating in joint activities, people routinely have to retrieve spatial information from memory and convey it to others. They do so in a range of tasks, such as describing places they have visited, providing driving directions over the phone, or arranging where to meet. Since in many such socially-embedded tasks people often occupy a different spatial viewpoint from their conversational partner, one important empirical question centers on how readily they take their partner's viewpoint into account.

Much of our understanding of people's ability to adopt or maintain non-egocentric spatial perspectives stems from studies using non-interactive tasks (e.g., Carlson-Radvansky and Irwin, 1993; Carlson-Radvansky and Logan, 1997; Mou et al., 2004b). Studies using interactive tasks, on the other hand, identify factors that influence the perspective from which people tend to produce or interpret spatial descriptions (e.g., Schober, 1993; Mainwaring et al., 2003; Tenbrink et al., 2011). However, such studies typically don't examine directly the underlying off-line spatial representations or on-line processes that support perspective-taking (but see

Shelton and McNamara, 2004; Duran et al., 2011; Galati et al., 2013).

Thus far, the findings emerging from both interactive and non-interactive tasks suggest that when people select a spatial perspective—whether to describe spatial information or to organize spatial information in memory—they consider various sources of information, including cognitive, contextual, and social factors. For instance, such factors have been shown *individually* to constrain how people organize and maintain spatial information in memory. When learning and remembering a spatial layout, people appear to interpret it in terms of a reference system that maintains spatial relations around a preferred direction (e.g., McNamara, 2003; Mou et al., 2004a), a process analogous to determining its "top." This preferred direction is influenced by egocentric information, such as one's initially experienced viewpoint (Shelton and McNamara, 2001), representational or environmental information, such as the environment's geometry (Shelton and McNamara, 2001), the symmetry or intrinsic structure of the spatial configuration (Mou and McNamara, 2002; Li

et al., 2011), functional features of landmarks in the configuration (Taylor and Tversky, 1992), and even social information available from the visual context, such as their conversational partner's viewpoint (Shelton and McNamara, 2004; Galati et al., 2013).

In this article, we present a framework for how people spontaneously recruit both social and representational information in spatial reasoning. Our view is that people consider all available cues and, upon gauging the relative cognitive demands of perspective-taking on each partner, weigh these cues accordingly to select the perspective that would minimize the pair's joint effort. This view differs from earlier proposals that have acknowledged the contribution of multiple sources of information in spatial reasoning, but have given precedence to representational cues (namely, egocentric experience, Shelton and McNamara, 2001, and the intrinsic structure of the layout, Mou and McNamara, 2002). Our framework emerges from our own work examining the perspectives people adopt in joint tasks both in their spatial descriptions and in their underlying representations supporting those descriptions. This framework allows for predictions for both linguistic and memory performance and can account for how people adapt their encoding, description and coordination strategies as incoming cues become available in range of spatial tasks.

Since our focus here is on collaborative tasks, we begin by reviewing in the next section studies demonstrating the contribution of social cues in spatial perspective-taking. Then, we present evidence from our own experimental work demonstrating that people integrate social cues (the availability of partner's viewpoint) with representational ones (their misalignment from their partner's viewpoint or from the configuration's intrinsic structure) to determine the perspective from which to organize information in memory and subsequently describe this information to a partner. Since an assumption of our framework is that partners jointly aim for efficient communication, in a subsequent section, we address whether people's perspective choices are in fact effective, as reflected by the pairs' efficiency and accuracy in the joint task. In the final section, we flesh out in more detail the characteristics of our proposed framework, addressing along the way some of the predictions it affords for a range of spatial perspective-taking tasks. We conclude that people weigh multiple cues to determine the task's cognitive demands on themselves and their partners, and as a result select strategies that are generally effective in facilitating coordination.

SOCIAL CUES INFLUENCE SPATIAL PERSPECTIVE-TAKING

A growing body of evidence suggests that, when adopting a spatial perspective, people consider different sources of social information, including attributional and contextual cues about their conversational partner. Attributional cues about the partner include pre-existing beliefs or attributions made about the partner based on prior experience, expectations or a stereotype (e.g., believing that the partner is unfamiliar vs. familiar with the environment, believing that the partner is a child vs. an adult). Such cues, if not available in advance, can also be accumulated during the course of the interaction and may even be used to update initial beliefs about the partner (e.g., Brennan et al., 2010). Whereas attributional cues pertain to the partner's cognitive or

other intrinsic abilities, contextual cues are not intrinsic to the partner, but are instead visually available in the physical environment and concern the partner's visibility, relative position in space, misalignment, or other relevant external features.

In line with other researchers (Clark and Wilkes-Gibbs, 1986; Schober, 1995; Duran et al., 2011), we propose that, on the basis of social cues, people make inferences about their conversational partner's ability to contribute to the joint task and adapt their perspective-taking behavior accordingly. This view follows from the proposal that, when collaborating, people share responsibility for ensuring mutual understanding and try to minimize their collective effort. This shared responsibility requires one partner to invest greater cognitive effort when appraising that the other partner is likely to find the interaction difficult; such behavioral adjustments are said to follow *the principle of least collaborative effort* (Clark and Wilkes-Gibbs, 1986; Clark, 1996). The evidence we report in the next two subsections, on how individual social cues influence perspective-taking, are broadly compatible with this view.

ATTRIBUTIONAL CUES ABOUT THE PARTNER INFLUENCE SPATIAL PERSPECTIVE-TAKING

As we have mentioned, one source of social information that shapes perspective-taking arises from the attributions people make about the partner's ability to contribute to the joint spatial task. Such attributions can depend on the status of the partner—for instance, whether the partner is believed to be real, imaginary, or simulated. There is evidence that with imaginary partners, or with partners with whom they cannot interact contingently, speakers are more likely to invest in adopting the partner's perspective. That is, speakers are more likely to use descriptions from the partner's perspective (e.g., “to your right” or “in front of you”) and less likely to use egocentric ones when describing spatial layouts to imaginary partners than to real ones (Schober, 1993). Speakers are also less likely to disambiguate the spatial descriptions they produce when they suspect that their partner is a confederate and does not have real informational needs (vs. a naïve participant) (Roche et al., 2010). This adaptation in perspective holds not only for the production of spatial expressions but for their interpretation as well. When listeners believe that their partner is real (vs. simulated) they are more likely to interpret ambiguous spatial descriptions egocentrically than from their partner's perspective (Duran et al., 2011). Comparable adaptation is found in non-linguistic communication strategies as well: in a “tacit communication game” in which participants could convey their intentions only through graphical means, they spent more time signaling the location of critical information to their partner when they believed they were interacting with a child than with an adult (Newman-Norlund et al., 2009).

Thus, when people believe that their partner cannot coordinate with them contingently or is otherwise less able to, they are more likely to adopt the partner's perspective and invest the effort to convey spatial information to them. And conversely, when people believe that their partner is real and able to coordinate with them contingently, they are more likely to shift the burden of mutual understanding to the partner, producing or interpreting spatial descriptions egocentrically.

Similarly, speakers adapt their spatial descriptions according to their beliefs about the partner's familiarity with the environment pertinent to the task. When speakers describe landmarks to a partner who is likely to be unfamiliar with them (e.g., Washington Square Park to a non-New Yorker), speakers use more detailed descriptions and are less likely to refer to the landmarks by their proper names than when they interact with partners who are natives of the city (Isaacs and Clark, 1987). Speakers also adapt how they plan and describe routes within environments. When describing routes to a partner who is presumed to be unfamiliar with the environment (vs. for themselves), speakers elaborate their descriptions by using more words and details, refer to more landmarks for orienting, and simplify the routes by navigating along fewer, larger and more prominent streets (Hölscher et al., 2011).

As this last study suggests, the framing of the task as collaborative or intended for an audience, as opposed to a monologic activity, can shape spatial descriptions. In a related study, Mainwaring et al. (2003) demonstrated that speakers were more likely to adopt their own perspective when describing spatial information for themselves, thus bearing the cognitive burden exclusively, whereas they were more likely to adopt their partner's perspective when describing spatial information to a misaligned imaginary partner who presumably bore more of the cognitive burden (see also Schober, 1993).

Moreover, people adapt their perspective-taking behavior not only on the basis of their beliefs about the partner's ability to contribute to the spatial task, but also on the basis of so-called "second order beliefs" about the partner: the speaker's beliefs about what the partner believes about the speaker's viewpoint or abilities. For example, when people believe that their partner doesn't know their spatial viewpoint (and therefore cannot consider their perspective), they are more likely to interpret spatial descriptions from the partner's perspective (Duran et al., 2011).

In addition to social cues that are available a priori (e.g., by being told in advance that the partner is unfamiliar with the environment, or that the partner is a child), people can often discover such cues by accruing relevant evidence as the interaction unfolds. For example, based on their partner's performance and feedback, people can make attributions about their relative spatial skills and thus their ability to advance the joint goals of the task. Schober (2009) demonstrated that perspective adaptation can occur on the basis of local, incremental cues, using preselected pairs of participants that had matched or mismatched spatial abilities. As expected, high-ability speakers were overall more likely to use partner-centered descriptions whereas low-ability speakers were more likely to use egocentric ones. But critically, during the course of the interaction, as high-ability speakers in mixed pairs formed attributions about their low-ability partners they increased their use of partner-centered descriptions, and conversely, low-ability speakers describing to high-ability partners decreased their use of partner-centered descriptions. Similarly, incremental visual cues about progress on the task, such as errors indicating the partner's misunderstanding, can contribute to updating attributions about the partner and lead to appropriate adaptation, such as disambiguating spatial descriptions (Roche et al., 2010). Thus, along with a priori information, incoming information about the

partner's ability to contribute to the spatial task can influence perspective-taking.

CONTEXTUAL CUES RELATING TO THE PARTNER INFLUENCE SPATIAL PERSPECTIVE-TAKING

In assessing the relative cognitive demands of perspective-taking, people consider not only attributional cues pertaining to the partner's knowledge or ability, but also contextual cues concerning the partner's spatial relation to themselves and other features of the environment. Information that is visually available in the shared environment, what is termed as the partners' physical co-presence, is one of the principal heuristics that people assess in order to establish what they have in common ground and to tailor their behavior accordingly (Clark and Marshall, 1981; Clark and Brennan, 1991).

The visibility between partners is, most obviously, one factor that shapes what is physically co-present and thus influences how people interact in the context of spatial tasks. For example, in a task where partners reconstructed arrangements of lego blocks, pairs who could see each other coordinated differently than those who couldn't: speakers adapted their descriptions contingently as their addressees exhibited, poised, pointed at and oriented blocks, which resulted in more accurate and efficient performance on the task (Clark and Krych, 2004). Similarly, in a task in which pairs were trying to align icons on identical maps displayed on networked computer screens, when the person giving directions lacked visual evidence about their partner's icon movements, they left the initiative to move to the next trial to their partner and went through a lengthier process of verbally checking that they have achieved mutual understanding (Brennan, 2005).

The misalignment between partners is another salient contextual cue that shapes people's attributions about the partner's ability to contribute to the task, making it perhaps the factor most often manipulated in interactive studies of spatial perspective-taking. There's evidence that when pairs jointly reconstruct layouts that are simple or randomly configured, the degree of misalignment between partners influences the perspective of speakers' descriptions. For instance, speakers are more likely to use partner-centered descriptions than egocentric ones when describing layouts to partners who are misaligned rather than aligned with them (Schober, 1993, 1995). A caveat here is that, in these and other experiments (e.g., Mainwaring et al., 2003; Duran et al., 2011), partners were misaligned exclusively by orthogonal offsets (i.e., were at 90°, 180°, or 270°). Because orthogonal perspectives are aligned with one's canonical axes, they may be privileged and thus relatively easily adopted or maintained (see McNamara, 2003; Avraamides et al., 2013). The facilitation of the canonical axes may, thus, account for the similarities in participants' description preferences across different offsets. In our own studies (Galati et al., 2013; Galati and Avraamides, in revision), which will be subsequently described in detail, we have addressed this possibility by including oblique offsets between partners in order to determine when, in fact, perspective-taking is most computational demanding for speakers.

The misalignment between partners also influences the shape of their shared space, which can consequently influence spatial descriptions, even when these descriptions are embedded in a

narrative. For instance, speakers adapt the directionality of their gestures that accompany spatial prepositions, such as *in* and *out*, as a function of the shape of the space they share with their conversational partners (Özyürek, 2002). When partners are seated face-to-face *in* and *out* are mapped onto a sagittal axis with respect to the speaker's body, whereas when the partner is seated to the side *in* and *out* are mapped onto a lateral axis. These findings are taken to suggest that, in interaction, spatial concepts are encoded with respect to the partners' shared space (i.e., with *in* corresponding to the "inside" of the shared space).

In sum, people consider various aspects of their relation to their partner within the physical environment, including their partner's visibility, their degree of misalignment, and the shape of their shared space. Upon considering these contextual cues, they adapt their descriptions or coordination strategies accordingly.

BEYOND THE INFLUENCE OF ISOLATED SOCIAL CUES

So far we have considered evidence that people consider social cues, either available from the onset of the interaction (e.g., through advance instructions or through the physical environment) or accrued during the course of the interaction (e.g., through the partner's feedback) to make attributions about their partner's ability to contribute to the task.

Overall, when people perceive their partner to be limited in some way, they invest the effort to adopt their partner's perspective or to convey information in a more accessible way. This is the case when they believe that the partner is imaginary (Schober, 1993), a child (Newman-Norlund et al., 2009), or unfamiliar with the environment (Hölscher et al., 2011), when they believe the partner does not know (Duran et al., 2011) or does not share their viewpoint (Schober, 1993, 1995; Mainwaring et al., 2003), when they discover that the partner has worse spatial abilities than they do (Schober, 2009), or cannot provide feedback during the interaction (Shelton and McNamara, 2004; Duran et al., 2011). On the other hand, when people perceive their partner to be less limited, as for example when they interact with a real (or assumed to be real) partner or a partner who can contribute contingently to the interaction, they may not invest as much effort in adopting the partner's perspective and instead rely on that partner to request clarifications, as needed. Together, these studies serve as a compelling demonstration of the principle of least collaborative effort (Clark and Wilkes-Gibbs, 1986; Clark, 1996) in spatial tasks.

However, few studies have addressed directly how social cues are considered alongside other sources of information pertinent to spatial tasks, such as information about the intrinsic structure of the configuration. Since real-world environments are often systematically organized, having axes of symmetry or salient landmarks, when selecting the perspective from which to describe them, speakers likely consider not only their partner's viewpoint but also other representational cues intrinsic to the configurations. Indeed, some of the reviewed studies allude to the possibility that people integrate multiple sources of information, even if this is not examined directly. For example, in Hölscher et al.'s study (2011), a social cue (the partner's assumed familiarity with the environment) influenced the extent to which representational cues (landmarks and other salient features of the environment) were incorporated in route descriptions.

Our research agenda has focused on elucidating precisely how social cues about the conversational partner interact with other sources of information during spatial reasoning. This approach extends the principle of least collaborative effort (Clark and Wilkes-Gibbs, 1986; Clark, 1996), insofar as it places an emphasis on the probabilistic weighing and interaction of social and other cues when assessing collaborative effort. Moreover, our approach focuses not only on clarifying how people combine various sources of information to adapt how to coordinate in spatial tasks, but also how this behavior is supported by the cognitive infrastructure—namely, by spatial memory representations. Relating perspective-taking choices in descriptions to their underlying spatial representations would further bolster the view that partner-specific adaptation in dialog is supported by ordinary cognitive processes acting on memory representations (e.g., Horton and Gerrig, 2002, 2005; Metzger and Brennan, 2003; Pickering and Garrod, 2004). Thus, in our work, we examine how social and representational cues interact to influence, not only speakers' spatial descriptions, but also the preferred perspective around which they organize spatial information in memory (see the next section).

Others have shared our view that, at least with respect to organizing and maintaining spatial information in memory, a number of cues are taken into account. For instance, McNamara and his colleagues have proposed that learning and remembering a spatial layout involves interpreting it in terms of a reference system, whose selection depends on spatial and non-spatial properties of the objects, the structure of the surrounding environment, the observer's egocentric viewpoint, and even verbal instructions (Shelton and McNamara, 2001; Mou and McNamara, 2002). But contrary to these proposals, which ascribe precedence to certain cues as being dominant, such as egocentric experience (Shelton and McNamara, 2001) or the intrinsic structure of the layout (Mou and McNamara, 2002), we consider all available cues to be probabilistically combined upon being weighted according to task-specific demands.

In the context of collaboration, task-specific demands arise from aiming for effective coordination. On the basis of such demands, our framework affords predictions for how different cues are weighted and ultimately whose perspective is selected, whether for organizing spatial information in memory or for descriptions. Our framework also affords predictions about how people make use of cues that become available at different time points of a spatial task, as for example when discovering the partner's viewpoint relative to a configuration only *after* the configuration has been learned. Specifically, our framework assumes a great deal of flexibility in incorporating incoming cues to select a spatial perspective (see also Li et al., 2011). For instance, it predicts that when having to describe from memory a learned configuration, people won't simply select the perspective according to which their spatial memory is organized, but will also take into account new perceptually available cues from the interactive situation.

In the next two sections, we present some of our experimental work, which demonstrates that partners consider multiple cues to assess each other's cognitive demands when encoding and communicating spatial information. In the final section of this article,

we describe in more detail our framework for flexible perspective-taking, which qualitatively accounts for our experimental results and affords predictions for other perspective-taking tasks.

WEIGHING SOCIAL AND REPRESENTATIONAL CUES IN SPATIAL PERSPECTIVE-TAKING

In our work, we have focused on one contextual social cue—the a priori visual availability of the partner’s misaligned viewpoint. Our goal was to examine the conditions under which this a priori social cue influences how speakers spontaneously organize spatial information in memory and how they describe it to their partner. In the first study (Galati et al., 2013), we examined whether knowing the partner’s viewpoint in advance is, on its own, a sufficient cue to influence speakers’ memory and descriptions. In the second study (Galati and Avraamides, in revision), we examined how the availability of the partner’s viewpoint may be used in conjunction with another representational cue, the intrinsic structure of the spatial layout, to shape memories and descriptions.

In both studies, in order to clarify how memory representations support perspective-taking behavior, we dissociated the learning of spatial layouts from their description: speakers first learned a spatial layout, had their memory of the layout assessed, and then described it from memory to a partner. Most earlier studies don’t address the relationship between memory representations and linguistic choices, as they involve situations in which speakers can see the spatial information they describe (e.g., Schober, 1993, 1995, 2009; Mainwaring et al., 2003), learn the spatial information while simultaneously describing it (Shelton and McNamara, 2004), are instructed to describe spatial information from a particular perspective before their memories are assessed (Shelton and McNamara, 2004), or describe familiar environments whose underlying memory representation is not directly assessed (Hölscher et al., 2011). Dissociating the encoding of spatial information of from its description enables us to determine not only whether advance knowledge of the partner’s viewpoint influences speakers’ memories and descriptions, but also the extent to which speakers rely on their memories when describing spatial information.

THE INFLUENCE OF THE AVAILABILITY OF THE PARTNER’S MISALIGNED VIEWPOINT

In Galati et al. (2013), we asked whether knowing the partner’s viewpoint in advance influences speakers’ memory and descriptions. In 18 pairs, one participant (the Director) first studied a randomly configured tabletop layout of seven objects (see Figure 1). They later described it from memory to another participant (the Matcher), seated at a separate round table, who reconstructed the layout by following the Director’s descriptions (see Figure 2). This took place across three blocks that varied in terms of what Directors knew about their Matcher’s viewpoint when studying the layout. In the first block, Directors didn’t know that they would later describe the layout to a Matcher, whereas in the subsequent blocks, whose order was counterbalanced across pairs, they either knew they would describe the layout to a Matcher but didn’t know the Matcher’s viewpoint, or knew the Matcher’s viewpoint because the Matcher was co-present in the room during learning, seated at the position they

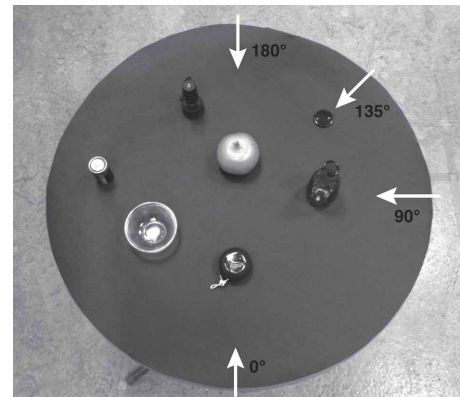


FIGURE 1 | One of the three seven-object layouts used in Galati et al. (2013), whose configuration was designed to appear seemingly random. It comprised a battery, a flashlight, a bowl, an orange, a yoyo, a button, and a vase. The arrows represent the Director’s viewpoint (0°), and the Matcher’s viewpoint when offset by 90°, 135°, and 180°.

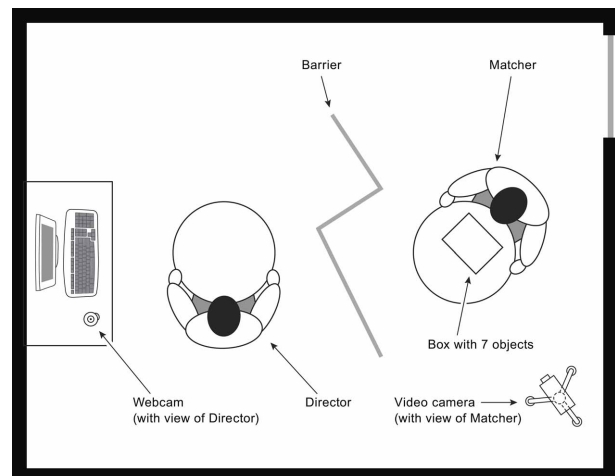


FIGURE 2 | Set-up of our studies, showing the Director’s and Matcher’s working stations, and the locations of recording devices. This example of a description phase illustrates the conditions of Galati et al. (2013) in which Directors and Matchers were misaligned by 135°, and the conditions of Galati and Avraamides (in revision) in which Directors were aligned with the intrinsic structure (with Matchers misaligned by 135°).

would occupy at the description phase. The degree of misalignment between partners during the description phase, was 90°, 135°, or 180°, and was counterbalanced across the three blocks.

After studying the spatial layout, the Director’s memory of it was assessed through two tasks. The first involved *judgments of relative direction* (JRDs), which required imagining a specific location and orientation, and pointing with a joystick to another object from that imagined perspective (e.g., *Imagine being at the vase, facing the orange. Point to the button.*) These JRD trials included eight imagined headings (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°), whose order was randomized. Performance was assessed in terms of Directors’ orientation latency (the time

taken to adopt the imagined perspective of the first instruction) and their response latency (the time taken to point to the target identified in the second instruction). Performance on JRDs permits determining the preferred direction participants use to organize the spatial relations in memory (e.g., Kelly et al., 2007). The rationale is that spatial relations specified with respect to the preferred direction can be retrieved from memory more readily than those relations that are not explicitly specified and therefore have to be inferred. Thus, judgments from headings aligned with that preferred direction should show facilitation in terms of the orientation and response latencies.

We found that, when the Matcher's viewpoint was unavailable at study (whether on the first or a subsequent block), Directors encoded spatial layouts egocentrically: they were faster to imagine orienting to and to respond from perspectives aligned with their own. On the other hand, when the Matcher's viewpoint was known in advance, it was encoded in memory, showing distinctive processing, at least when Matchers were known to be misaligned by 90° or 135°. When knowing that the Matcher would be at these offsets, Directors took longer to imagine orienting to headings aligned with these known viewpoints. This slower orienting may seem counterintuitive in light of previous findings that speakers can show facilitation for the partner's viewpoint (Shelton and McNamara, 2004). However, in our study, Directors knew that during the description phase they could interact freely with their Matchers and that their respective viewpoints would be mutually known (cf. Shelton and McNamara, 2004), such that the Matcher could bear some of the cognitive burden of perspective-taking. We therefore proposed that our Directors may have not invested the cognitive effort at study to organize spatial relations from their Matcher's viewpoint, but instead encoded their Matcher's viewpoint to use it later, as needed. The longer orientation latencies may therefore reflect a reconstructive process, whereby Directors recalled an episodic representation of their experience at study, which included the location of the Matcher in space, and linked the Matcher's viewpoint to their representation of the layout.

The second memory task provided corroborating evidence that Directors represented the partner's viewpoint in memory. In this task, the Directors drew the spatial layout by indicating the position of each object on a grid circle representing their table. These array drawings allowed us to assess the Directors' memory for the relative positioning of objects and for systematic biases (e.g., Friedman and Kohler, 2003). We found that when Directors knew their Matcher's viewpoint in advance, their drawings showed a reliable rotational bias by approximately 5° toward the Matcher's viewpoint.

Following these memory tasks, Directors described the layout from memory to their Matcher. We examined the distribution of Directors' egocentric (e.g., "in front of me is the bracelet") and partner-centered (e.g., "the battery is to your right") expressions. The distribution of these types of expressions allows for inferences concerning whether an egocentric or partner-centered perspective was predominately in use, and thus reflect Directors' overall description strategies. We found that Directors did adapt their spatial expressions according to what they had known about their Matchers at study (see Table 1). However, knowing the

Table 1 | Means (and standard deviations) of the proportions of Director-centered and Matcher-centered expressions produced by Directors describing layouts that were randomly configured (Galati et al., 2013) or with an intrinsic structure (Galati and Avraamides, in revision).

	Director-centered	Matcher-centered
CONFIGURATION IS RANDOM		
Matcher's viewpoint unavailable^a		
Misaligned by 90°	0.16 (0.22)	0.28 (0.17)
Misaligned by 135°	0.16 (0.14)	0.23 (0.21)
Misaligned by 180°	0.22 (0.23)	0.18 (0.14)
Matcher's viewpoint available		
Misaligned by 90°	0.15 (0.20)	0.26 (0.17)
Misaligned by 135°	0.42 (0.23)	0.08 (0.13)
Misaligned by 180°	0.16 (0.11)	0.24 (0.16)
CONFIGURATION HAS INTRINSIC STRUCTURE		
Matcher's viewpoint unavailable		
Aligned with Director	0.34 (0.20)	0.09 (0.16)
Aligned with Matcher	0.00 (0.00)	0.45 (0.05)
Aligned with Neither	0.27 (0.25)	0.10 (0.13)
Matcher's viewpoint available		
Aligned with Director	0.17 (0.09)	0.18 (0.16)
Aligned with Matcher	0.05 (0.07)	0.31 (0.31)
Aligned with Neither	0.00 (0.00)	0.29 (0.20)

For both studies, the distribution of these expressions is shown across the availability of the Matcher's viewpoint at learning (unavailable vs. available) and across the relative positioning of the conversational partners (in Galati et al., 2013, with respect to their misalignment, whereas in Galati and Avraamides, in revision, with respect to their alignment with the intrinsic structure.).

^aThis combines the two conditions from Galati et al. (2013), in which the Matcher's viewpoint was unavailable: the first block in which Directors didn't know there would be a description phase and a subsequent block in which they did know about the description phase but did not know the Matcher's viewpoint.

Matcher's viewpoint in advance did not determine on its own the perspective of Directors' descriptions. For instance, when Directors knew their Matcher's viewpoint at study, they didn't simply use more partner-centered expressions during the description. Instead, they made strategic choices upon considering the demands of perspective-taking on themselves and their Matchers.

When perspective-taking was relatively easy (at the small offset of 90°), Directors used Matcher-centered expressions more frequently than egocentric ones. When pairs were counter-aligned and thus shared a canonical axis, Directors mixed perspectives more frequently, suggesting that they could alternate flexibly between their own and their partner's perspective. When perspective-taking was known to be more computationally demanding for Directors, at the oblique offset of 135°, they were more likely to describe layouts egocentrically, as shown in Table 1. That is, since Directors presumably bore more of the cognitive burden in this task, having to recall spatial relations and convey them to their partner, they opted for their own perspective when perspective-taking was especially demanding for them, letting their partners unpack the spatial mappings of

their egocentric descriptions. Explicit agreements between partners to do so did, indeed, happen most often when partners had known in advance they would be offset by 135° relative to the other offsets. Thus, the availability of the partner's viewpoint enabled both interlocutors to mutually recognize when the cognitive demands would be taxing for the person carrying the greatest cognitive load and to adapt their communication strategies in ways that facilitated their coordination (for evidence for this facilitation see the next section, on the Coordination in Spatial Perspective-Taking).

Thus, speakers do not spontaneously use their partner's viewpoint as an organizing direction for their memories when it is available; in our study, Directors didn't show facilitation for their partner's viewpoint (cf. Shelton and McNamara, 2004). But despite not using the partner's viewpoint as an organizing direction, speakers do represent that viewpoint in memory; this was evidenced by the Directors' array drawings and the distinctive processing, in JRDs, of perspectives aligned with the partner (at least when they were misaligned, though not counteraligned, with their partner). Finally, when describing this spatial information, speakers don't merely rely on their initial representations, but are able to use information perceptually available in the task (i.e., their degree of misalignment from their partners) to adapt descriptions appropriately.

The flexible adaptation of speakers' perspective choices, here, is consistent with the principle of least collaborative effort (Clark and Wilkes-Gibbs, 1986; Clark, 1996) in that partners shared the burden of ensuring mutual understanding and shifted their cognitive effort appropriately. When recognizing that one of them was especially likely to find the perspective-taking difficult (e.g., the Director describing the layout from a 135° offset), the other readily invested greater effort (e.g., the Matcher agreed to interpret descriptions from the Director's viewpoint).

INTEGRATING THE PARTNER'S MISALIGNED VIEWPOINT WITH REPRESENTATIONAL CUES

So far, we have seen that when speakers are not instructed to adopt their partner's viewpoint and can interact freely with their partners, they may not have sufficient pragmatic motivation to organize spatial relations around a non-egocentric viewpoint. Organizing spatial relations non-egocentrically presumably requires investing cognitive effort, at least when there aren't any other spatial cues, as with the randomly configured layouts in Galati et al. (2013). In such circumstances, as we've seen, speakers can represent the partner's viewpoint relative to the spatial layout and use it later as needed. In our next study (Galati and Avraamides, in revision), we wanted to establish whether speakers *would have* sufficient pragmatic motivation to organize spatial relations around the partner's viewpoint, when that viewpoint is reinforced by additional spatial cues.

The overall procedure of this study was similar: Directors first studied a spatial layout, which now had an intrinsic orientation (seven real objects were configured across a bilateral axis of symmetry, as shown in Figure 3), while either knowing their misaligned Matcher's viewpoint or not. Then, as with Galati et al. (2013), the Directors' memory of the layout was assessed through JRDs and array drawings, and finally they described the layout to

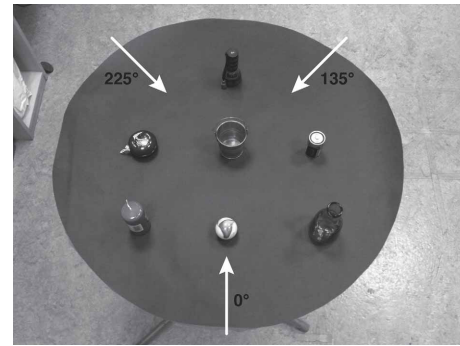


FIGURE 3 | The seven-object layout used in Galati and Avraamides (in revision), whose configuration had an intrinsic structure. It comprised a flashlight, a yoyo, a bucket, a battery, a candle, a marble, and a vase. The arrows at 0° , 135° , and 225° represent the viewpoint that Directors and Matchers occupied at the different conditions of their relative alignment with the intrinsic structure.

their Matcher. In this experimental design, across the 24 pairs, the Director's and the Matcher's relation to the intrinsic structure of the layout differed, such that the structure was aligned with the Director, the Matcher, or neither partner. A third of the Directors studied arrays while aligned with the intrinsic structure (from 0°), and later described it to a Matcher who was offset by 135° (measured counterclockwise from 0°). Another third of the Directors studied arrays from 225° and later described it to a Matcher who was aligned with the structure (at 0°). And a final third of the Directors studied arrays again from 225° and later described to a Matcher who was offset by 135° , such that neither partner was aligned with the structure. For each group, half the Directors had known at study their Matcher's subsequent viewpoint and half of them did not. By dissociating the study from the description of the spatial layout and varying systematically the convergence of cues (i.e., whose viewpoint was aligned with the structure), we aimed to clarify how people integrate these cues as they become available.

The memory tests revealed that the preferred direction around which Directors organized spatial relations in memory depended on the convergence of cues—i.e., on whose viewpoint was reinforced by the layout's intrinsic orientation. This was most obvious by how Directors oriented their array drawings. When Directors had studied layouts while aligned with their intrinsic structure, they always drew them from their own viewpoint; it did not matter whether they knew their Matcher's subsequent viewpoint or not. When they had studied layouts while misaligned with the intrinsic structure, knowing the Matcher's subsequent viewpoint did influence how they oriented their drawings. Specifically, they were more likely to use the structure's axes (vs. their own viewpoint) as the organizing direction when knowing in advance that the Matcher would be aligned with the structure. And conversely, they were more likely to use their own viewpoint (vs. the structure's axes) when not knowing their Matcher's subsequent viewpoint. When knowing in advance that the Matcher would also be misaligned with the structure, Directors were equally likely to draw arrays from their own viewpoint or from an axis of the

structure, perhaps because the intrinsic structure became more salient (relative to not knowing the Matcher's viewpoint) upon considering objects from a second oblique viewpoint.

Performance in the JRD task corroborated that Directors had indeed organized spatial relations in memory according to the orientation of their drawings. Directors who had drawn layouts aligned with the structure were faster to orient to and respond from headings aligned with the structure's axes (0° , 90° , 180° , 270°), whereas Directors who had drawn layouts misaligned with the structure (specifically from their study viewpoint of 225°), were faster to orient to and respond to from headings aligned with that viewpoint and its canonical axes (i.e., 315° , 45° , 135°).

Directors also selected perspectives strategically in their descriptions. In this study, we examined the distribution of three types of spatial expressions of theoretical interest: Director-centered, Matcher-centered, and Structure-centered ones. The latter category involved expressions that were from headings aligned with the intrinsic structure and were not person-centered (e.g., "*On the perpendicular*. You're supposed to be on one side *on the left*, and I'm on the one side of the table *on the right*."). Overall, as shown in **Table 1**, Directors used reliably more Matcher-centered expressions than other types of expressions when the Matcher was aligned with the structure, and used numerically but not reliably more Director-centered expressions than Matcher-centered ones when they were the ones aligned with the structure. As with Galati et al. (2013), speakers didn't merely rely on the organization of their memories to choose the perspective of their descriptions, but rather took into account information that was perceptually available during the description phase. Although Directors used overall more Matcher-centered expressions when knowing their Matcher's viewpoint in advance, the preferred direction of Directors' memory (as reflected by their drawings) did not reliably influence their distribution of egocentric or partner-centered expressions. For example, even though most Directors who had studied layouts from 225° while not knowing their Matcher's viewpoint organized spatial information egocentrically in memory, they used overwhelmingly Matcher-centered expressions when they interacted with a Matcher who was aligned with the structure (at 0°) (see **Table 1**). In other words, when the convergence of available cues at the interaction strongly biases a particular perspective (e.g., when the partner's viewpoint and the structure's intrinsic alignment coincide), speakers override their initial memory representation to select the perspective of their descriptions.

Nevertheless, the advance availability of a social cue, such as the partner's viewpoint, and its relation to other cues (e.g., the intrinsic structure) can influence perspective-taking, when it highlights alternative and potentially useful perspectives for encoding and describing a spatial layout. As we have mentioned, Directors who studied layouts from 225° were relatively more likely to use the structure's axis as an organizing direction when knowing that the Matcher would be at 135° compared to not knowing the Matcher's viewpoint. Knowing in advance that neither partner was aligned with the structure influenced descriptions as well: when Directors at 225° had known in advance that Matchers would be at 135° , they used more Matcher-centered than egocentric expressions, and used numerically more

Structure-centered expressions compared to not knowing the Matcher's viewpoint. Thus, knowing the partner's viewpoint while studying a layout from an oblique viewpoint can make its intrinsic organization more apparent and can influence both how speakers organize spatial information in memory and how they describe it.

Together, the findings of these two studies set the stage for a framework for how people use multiple cues, including social and representational ones, in spatial perspective-taking. Upon considering all available cues jointly and weighing them according to their salience and relevance to the task, people select the perspective reinforced probabilistically by most cues, and organize spatial information in memory or describe it to a partner accordingly. One assumption here is that people consider the perspective reinforced by multiple cues to be optimally effective in minimizing the pair's collective effort. In the next section, we examine whether in fact the perspectives pairs select make their coordination more effective.

COORDINATION IN SPATIAL PERSPECTIVE-TAKING

We have so far suggested that collaborating partners select the perspective reinforced by all available cues in an effort to minimize their collective effort and maximize their efficiency of coordination—their efficiency at behaving contingently to achieve shared goal. To determine whether, in fact, people are adept at gauging which perspective would be most effective in the task, we examined two aspects of collaborative performance in our previously described studies.

The first, which tapped into pairs' efficiency on the task, was the number of conversational turns—uninterrupted stretches of speech by a Director or a Matcher—that pairs took to reconstruct a spatial layout. We took conversational turns to reflect the pairs' degree of *grounding*, or exchanging evidence about what they do or do not understand (e.g., Clark and Brennan, 1991; Clark, 1996; Brennan, 2005). Examining the pairs' turn-taking patterns enables us to identify the circumstances and description strategies that contribute to facilitated grounding: when perspective-taking strategies facilitate grounding, pairs should interact over fewer conversational turns.

The second collaborative outcome tapped into pairs' accuracy on the task: it assessed the accuracy with which Matchers, having followed the Directors' descriptions, reconstructed the spatial layouts with real objects on top of their table. Using bidimensional regression analyses we compared the Matcher's reconstruction (photographed from a bird's eye view at the end of the session) to the veridical coordinates of the original configuration. Again, when the pairs' perspective-taking strategies are effective, the Matchers' reconstructions should be less distorted.

It is important to note that what conversational partners consider to be an effective strategy is task-dependent, rather than strictly defined in terms of efficiency and accuracy. In our studies, the pairs' goal was to reconstruct layouts as accurately as possible despite lacking visual access to each other's work stations. These task-specific goals and constraints must have influenced the criterion that pairs adopted to reach the mutual belief that they had understood each other well enough for their current

purposes. According to Clark and Brennan (1991), this “grounding criterion” depends both on the goals of communication (here, emphasizing accuracy) and the affordances of the communicative situation (here, lacking visibility). Thus, although an effective strategy is ideally one that maximizes efficiency in terms of turn-taking while also yielding high accuracy on the resulting reconstruction, in some circumstances, efficiency and accuracy may be dissociated if weighted differently by the task’s goals.

SOCIAL AND REPRESENTATIONAL CUES SHAPE GROUNDING

For the pairs who reconstructed randomly configured layouts (Galati et al., 2013), knowing the partner’s viewpoint at learning helped their subsequent efficiency in some circumstances. Specifically, pairs took numerically fewer turns to complete the reconstruction of the layout when they had mutually known in advance that they would be misaligned by the oblique 135° than by the other, orthogonal offsets (see Galati and Avraamides, 2012). This counterintuitive pattern makes sense insofar as the Directors’ description strategies were suitable. As we’ve reported in the previous section, in Galati et al. (2013), at the oblique and more computationally demanding offset of 135°, Directors showed a strong preference for describing layouts from their own perspective, frequently upon their Matcher’s prompting. This strategy turned out to be beneficial in alleviating their collective effort, as reflected by their conversational turns. In fact, as Directors used greater proportions of egocentric expressions when knowing their Matcher’s viewpoint in advance, pairs took reliably fewer turns to reconstruct the layout.

In this article, we present some analyses not reported elsewhere, on the efficiency of pairs who reconstructed layouts that had an intrinsic structure (from the corpus of Galati and Avraamides, in revision). In these circumstances, the number of turns that pairs took to complete the task was determined primarily by the alignment of the two partners’ viewpoints relative to the intrinsic structure at the description, $F_{(2, 18)} = 4.44$, $p < 0.05$. Here, advance knowledge of the partner’s viewpoint did not influence significantly the number of turns pairs took to reconstruct the layout ($p = 0.99$) and did not interact with the partners’ alignment with the structure ($p = 0.98$). Partners were the least efficient when neither of them was aligned with the intrinsic structure during the description: in that scenario they took an average of 259.25 turns ($SD = 125.29$), whereas they took an average of 114.88 turns ($SD = 66.87$) when the Matcher was aligned with the intrinsic structure and 166.75 turns ($SD = 68.27$) when the Director was aligned with it. Indeed, compared to when neither partner was aligned with the structure, pairs took significantly fewer turns when the Matcher was aligned with it, 95% CI ($-247.49, -41.26$), $p < 0.01$, and marginally more so when the Director was, 95% CI ($-195.62, 10.62$), $p = 0.08$.

Together, the patterns of turn-taking in the two experiments underscore the flexibility with which people use all available cues to select a spatial perspective in a joint task. When the spatial layout does not afford any representational cues (as when it is randomly configured), the a priori availability of a social cue, such as the partner’s subsequent viewpoint, can enable partners

to recognize when coordinating a perspective would be difficult for the partner bearing greater responsibility for mutual understanding and to agree on a perspective that alleviates that partner’s cognitive demands. With turns as a proxy of partners’ collaborative effort, these mutually agreed-upon strategies can make their interactions more efficient. On the other hand, when the spatial layout affords an intrinsic organization, its alignment relative to each partner during the interaction is what influences most the efficiency of coordination. In general, interactions are more efficient when the orientation of structure of the layout converges with one of the partner’s viewpoints than when it does not. Even though pairs were misaligned by a smaller offset when neither of them was aligned with the structure (by 90°) compared to when one of them was aligned with the structure (in which case their offset was 135°), the process of coordination was lengthier: thus, it was their relation to the intrinsic structure, not their misalignment from each other, that influenced their efficiency. We will return to this point in the final section of our article.

Thus far, we have seen that pairs generally adopt strategies that make their coordination more efficient in terms of the number of conversational turns they take to complete their joint task. When the layout provides intrinsic cues that coincide with a given partner’s perspective, speakers describe the layout from that person-centered perspective and this strategy is effective. When the layout does not provide such intrinsic cues, a priori information about the partners’ relative viewpoints helps determine which perspective is optimal for the speaker—adopting that perspective is an effective strategy.

SOCIAL AND REPRESENTATIONAL CUES SHAPE THE PAIRS’ ACCURACY ON THE TASK

Although the availability and convergence of various cues facilitated performance in terms of the efficiency of dialogs as reflected by turn-taking, it didn’t facilitate performance in the same way in terms of accuracy on the task. We assessed accuracy by examining the bidimensional regression coefficient (BDr), which estimates the goodness-of-fit between the tabletop reconstructions and the actual coordinates of the arrays, thus capturing unsystematic error in reconstructions when systematic biases are accounted for. We also examined the rotation parameter (θ), which indicates the degree to which the tabletop reconstruction was rotated relative to the studied array, thus capturing a potential systematic bias in the reconstructions.

In our study with randomly configured layouts, the only reliable finding from examining the Matcher’s tabletop reconstructions was that the relationships among objects became more distorted as Directors used more Matcher-centered expressions (see Galati and Avraamides, 2013). This could be due to Directors inadvertently introducing more inaccuracies in descriptions when computing spatial relations and selecting spatial terms from a non-egocentric perspective. This possibility is supported by the fact that, when partners were offset by 180° and Directors could more easily map egocentric spatial terms to partner-centered ones (e.g., *my left = your right*), the reconstructed layouts were less distorted than at the offsets of 90° and 135°.

In our study with layouts with an intrinsic structure, our new analyses reported here reveal a somewhat different pattern. Although the *BDr* was not reliably correlated with any of the three main types of expressions (Director-centered, Matcher-centered, Structure-centered), pairs reconstructed were less distorted layouts as Directors used greater proportions of Matcher-centered expressions with Matcher's aligned with the layout's intrinsic structure (Pearson's $r = 0.83$, $p < 0.05$). As we have shown in Galati and Avraamides (in revision), in this alignment condition Directors adopted the strategy of describing layouts from their Matcher's viewpoint, using overwhelmingly Matcher-centered expressions. This strategy was therefore effective, not only in terms of reducing the number of turns (see previous subsection), but also in terms of yielding less distorted reconstructions, underscoring that there was no speed-accuracy tradeoff in pair's efficiency. In general, reconstructions did not become more distorted as pairs interacted over fewer turns (Pearson's $r = -14$, $p = 0.52$), suggesting that pairs upheld the goal of the task to reconstruct layouts that were as accurate as possible.

Nevertheless, pairs demonstrated a systematic bias in rotating the spatial layout when its intrinsic structure was aligned with the Matcher during the description. For reconstructions in that condition, the average rotation parameter was $\theta = 1.94$, $r = 0.24$. Individual θ 's were not uniformly distributed around 0° ($V = 1.88$, $p = 0.17$). On the other hand, for reconstructions in the aligned-with-Director condition, the average rotation parameter was $\theta = 9.93$, $r = 0.97$, 95% CI $[-3.22, 23.08]$, and individual scores were uniformly distributed around 0° , $V = 6.66$, $p < 0.001$). This was also the case for reconstructions in the aligned-with-Neither condition, $\theta = -8.07$, $r = 0.94$, 95% CI $[-24.23, 8.09]$, $V = 7.47$, $p < 0.001$.

To summarize, although collaborating partners are successful at selecting perspectives that increase their efficiency, by minimizing their collective effort in terms of the length of their interaction, these perspectives don't always make them accurate on the task. In particular, decrements in accuracy seem to arise when speakers describe spatial information from the partner's viewpoint, especially when the configuration does not afford an intrinsic structure. When the configuration does afford an intrinsic structure, adopting the partner's perspective when it is reasonable to do so (when the partner is aligned with the structure) may be effective in some ways (e.g., reducing the length of the interaction, reducing distortion in the reconstructions) but not others (e.g., eliminating systematic rotational biases).

A FRAMEWORK FOR FLEXIBLE PERSPECTIVE-TAKING IN SPATIAL TASKS

Our findings contribute to a framework for flexible perspective-taking that captures several of the nuanced ways in which speakers reason and coordinate in spatial tasks. In our framework, perspective-taking is flexible insofar as speakers consider all available cues—both social and representational—and weigh them according to their salience and relevance to the task to select the most effective perspective. This probabilistic weighing of cues distinguishes our framework from others that ascribe precedence to egocentric experience (Shelton and McNamara, 2001) or intrinsic structure (Mou and McNamara, 2002). Another consequence

of this simultaneous weighing of multiple cues is that a single cue, such as the partner's viewpoint, may require further reinforcement from other cues to be adopted as an organizing direction in spatial memory. Pragmatic motivation from explicit instructions (Shelton and McNamara, 2004) or from the intrinsic structure (Galati and Avraamides, in revision) can supply such reinforcement. It also suggests that the misalignment between partners does not on its own reflect the computational demands of perspective-taking; instead, as we will argue, misalignment can lead to appropriate attributions about each partner's cognitive demands only in conjunction with other cues. Our framework's proposal that people use multiple, weighted cues extends to non-social spatial perspective-taking as well, affording predictions for which perspective or organizing direction they will select, even when reasoning for themselves.

That interacting partners take into account take each other's relative cognitive demands when selecting a perspective further underscores the flexibility of perspective-taking. Through this process, they determine the most effective perspective to use both for organizing information in memory and for describing it to one another. As we will discuss, in determining their relative cognitive demands, people take into account the collective effort invested across all phases of their joint task, from learning to the interaction.

In our framework, perspective-taking is also flexible in the sense that speakers don't rely blindly on their memories when selecting the perspective of their descriptions. Instead, they use perceptual information from the communicative information (e.g., about their partner's viewpoint), even if this hadn't been available in advance. In other words, they use both a priori and incrementally unfolding cues to update their attributions about which perspective would be optimal. Their assessment for what constitutes an effective perspective that would minimize their collective effort and maximize their performance depends on the grounding criterion they adopt in light of task's goals and constraints.

Finally, perspective-taking is flexible insofar as reflects the general flexibility of the cognitive system. Our framework considers partner-specific adaptation to emerge from ordinary cognitive processes acting on ordinary memory representations, whether spatial or episodic ones. As such, the principles of our framework—that speakers consider a confluence of cues, whether available perceptually or a priori, aiming to minimize collective effort—hold not just for spatial perspective-taking, but for conversational perspective-taking more broadly.

Below we expound further on the main characteristics of this framework and the insights that follow from it.

PEOPLE CONSIDER SIMULTANEOUSLY SOCIAL AND REPRESENTATIONAL CUES

During the course of perspective-taking, people consider various sources of information, including social cues (e.g., the availability of the partner's viewpoint), representational spatial cues (e.g., the layout's intrinsic structure), and egocentric biases (e.g., based one's own learning viewpoint). When multiple cues are available, people consider their confluence, weighing them according to their salience and relevance to the task.

In weighing multiple cues, people in collaborative tasks have to appraise the relative cognitive demands on each partner in order to select the perspective that minimizes their collective effort. An assumption here is that the perspective that is reinforced by the greatest number of cues or by the most salient cues is the most effective and thus preferable for encoding spatial information in memory and in language. Indeed, as we have shown, converging social and representational cues (e.g., the alignment of the layout with a given partner's viewpoint) motivate the use of a given perspective as the preferred orientation in memory and in descriptions.

Critically, a social cue, such as the availability of the partner's viewpoint, may not be sufficient on its own to shape the organizing direction of spatial memories (Galati et al., 2013) since organizing spatial relations around that viewpoint is costly and is unnecessary when pairs can interact freely and can correct misunderstandings (cf., Shelton and McNamara, 2004). As we have shown, in free dialogs, the partner's viewpoint may simply be encoded in memory. However, when this social cue converges with other cues (e.g., the layout's intrinsic structure), it can be used as the preferred direction of spatial representations at no discernible cost, despite being non-egocentric.

The intrinsic orientation of a spatial configuration is therefore one factor that contributes to adopting a non-egocentric viewpoint around which to organize spatial relations in memory. Related findings have led other researchers to propose that the intrinsic orientation of a layout is the dominant factor determining the preferred direction around which to organize information in memory (Mou and McNamara, 2002). However, in our framework, rather than ascribing precedence to particular cues, all available cues are weighted probabilistically according to task-specific demands. (Indeed, in Galati and Avraamides, in revision, Directors didn't invariably organize information in memory according to the configuration's intrinsic structure.) When multiple cues that are relevant to the task reinforce a particular viewpoint, that viewpoint is more likely to be adopted. Thus, when the orientation of the structure converges with one's own viewpoint, people opt for that egocentric viewpoint, whereas when it converges with their partner's viewpoint, they opt for their partner's viewpoint.

A final observation is that social cues can be combined not only with other types of non-social information (e.g., representational cues), but also with other types of social cues. A contextual cue concerning the partner (e.g., his misalignment from the speaker) can interact with an attributional cue about the partner (e.g., concerning his spatial abilities). For example, when a speaker describes a spatial layout to a partner misaligned by a relatively difficult offset (e.g., the oblique 135°), she may use more partner-centered expressions if she perceives him to have relatively poor spatial abilities, but more egocentric ones if she perceives him to have relatively good spatial abilities. Such predictions following from our framework can be explored in future research.

PEOPLE CONSIDER THE COGNITIVE DEMANDS OF PERSPECTIVE-TAKING FOR BOTH PARTNERS

Our framework accommodates and is compatible with *the principle of least collaborative effort* (Clark and Wilkes-Gibbs, 1986;

Clark, 1996)—the view that, in sharing responsibility for mutual understanding, conversational partners adapt their behavior in ways that aim to minimize their collective effort and facilitate their coordination.

In collaborative spatial tasks, the relative cognitive demands of perspective-taking on each partner motivate the perspective from which people encode or describe spatial information. Critically, the partner's viewpoint can influence the process of estimating their respective perspective-taking demands, as soon as it becomes available—whether at encoding or at the interaction.

In several real-world scenarios, people first have to commit certain spatial information to memory and convey it to someone else later, as for example, on a road trip when the co-pilot studies the route to the destination on a map and then gives directions from memory to the driver. In such situations, our framework posits that, to gauge their and their partner's relative cognitive demands, speakers must consider the cognitive effort they would invest in total, both when encoding the information and when describing it. Speakers must therefore estimate whether investing additional cognitive effort at encoding would yield savings in the effort they would expend later, when coordinating with their partner.

Having information about the upcoming interaction available in advance enables speakers to better anticipate the perspective most effective during the interaction and to adapt their encoding strategies accordingly. In our work, when speakers knew in advance that their partner's viewpoint was aligned with the layout's intrinsic orientation, they were more likely to adopt it as an organizing direction at encoding. Organizing spatial relations according to the partner's viewpoint made sense in terms of minimizing subsequent effort: speakers judged that this would be an effective perspective from which to describe the layout since the partner would not have to unpack the mappings of spatial expressions. Indeed, when the partner was aligned with the structure, speakers used overwhelmingly partner-centered expressions and pairs were the most efficient, at least in terms of their conversational turns.

Nonetheless, the availability of the partner's viewpoint alone, without the reinforcement of intrinsic spatial cues is not sufficient motivation, in free dialogs, to invest in organizing spatial relations around their partner's viewpoint. As we have seen, when speakers studied randomly configured layouts, they simply represented that viewpoint in memory in order to use it later, as needed. Despite not having invested the effort to encode such layouts from their partner's viewpoint, speakers could still adapt their descriptions upon considering the relative cost of perspective-taking based on their misalignment (see the subsection on the right column for a more detailed discussion of the factors contributing to the cost of perspective-taking). For instance, speakers could still adopt their partner's viewpoint in descriptions when perspective-taking was relatively easy for them (e.g., at small or canonical offsets). And when perspective-taking was relatively difficult (e.g., at oblique offsets), speakers would opt for their own perspective in descriptions. Their partner's endorsement of this strategy indicates that pairs mutually agree to reduce the cognitive demands of the

speaker, who in this context was encumbered by the greatest effort due to having to retrieve and describe spatial relations from memory.

People's dynamic and sophisticated adaptation of perspective choices suggests that they seek perspectives that are optimally effective in minimizing their effort, not just when collaborating, but also when investing cognitive resources in preparation for that collaboration. This is a novel elaboration of the principle of least collaborative effort.

PEOPLE USE FLEXIBLY A PRIORI AND PERCEPTUALLY AVAILABLE INFORMATION

The above discussion, regarding the cognitive demands at encoding and at the interaction, underscores the dissociation between the perspective of spatial descriptions and of the spatial memories supporting those descriptions. Our work demonstrates that speakers don't merely rely on the organization of their memories to select how to describe spatial relations, but instead also use information that is perceptually available in the interaction. A contextual social cue, such as the partner's viewpoint, can shape descriptions even if it had been unavailable at encoding and thus not incorporated in speakers' memory representations.

For example, in Galati et al. (2013), when the partner's viewpoint wasn't available at study speakers didn't necessarily use more egocentric expressions at the description, and conversely, when the partner's viewpoint was available at study speakers didn't necessarily use more partner-centered descriptions. Instead, speakers' description strategies were guided by contextual cues they encountered at the interaction: seeing that the partner was misaligned by a relatively small offset led to more frequent use of partner-centered descriptions, whereas seeing that the partner was misaligned by an oblique offset led to more frequent use of egocentric expressions.

Similarly, in Galati and Avraamides (in revision), the contextual social cue of the partner's viewpoint shaped descriptions even when its relation to the layout's structure was unknown at encoding. Overall, the organization of speakers' memories (as reflected by the orientation of their array drawings) didn't reliably influence their descriptions. For instance, despite most frequently encoding a spatial layout egocentrically when having studied it from a viewpoint oblique to its structure (225°) without knowing the partner's viewpoint, speakers overwhelmingly used partner-centered expressions upon encountering a partner aligned with the structure at the description.

Together, these findings suggest that speakers carefully attend to contextual social cues—partner-specific information that is perceptually available in the social situation—and use this information readily. As a result, they may override their perspective preferences for encoding the spatial information. This view is compatible with findings that people don't always adhere to the organizing direction of their memories when it conflicts with perceptual evidence, but use instead both sources of information to select the perspective of their descriptions (Li et al., 2011). Thus, the organization of spatial memories does not dictate how spatial information is subsequently described. Descriptions are also guided by perceptual information, which

partners use to determine the optimal perspective for the collaborative task.

PEOPLE DON'T ASSESS THE RELATIVE DIFFICULTY OF PERSPECTIVE-TAKING ONLY BASED ON THEIR MISALIGNMENT

There have been some incongruent findings concerning the offsets at which spatial perspective-taking is most difficult in collaborative tasks. In a study that involved interpreting another's spatial descriptions, listeners incurred a greater processing cost as the degree of misalignment from their partner increased (Duran et al., 2011). On the other hand, some studies focusing on production reported similarities in speakers' descriptions across misaligned offsets: with misaligned partners, speakers used partner-centered expressions with comparable frequency, regardless of the degree of misalignment (Schober, 1993, 1995; Mainwaring et al., 2003). This was taken as evidence against a mental rotation model of perspective-taking, and in favor of a categorical distinction between reasoning from an egocentric vs. a non-egocentric perspective (Schober, 1995). However, in all of these studies, real or assumed partners were misaligned by orthogonal offsets (90°, 180°, 270°). This methodological feature may limit our understanding of when perspective-taking is most demanding since, according to McNamara (2003), perspectives aligned with canonical axes can be facilitated relative to oblique ones.

Our findings are line with McNamara (2003) view, since when no intrinsic cues were available speakers opted for egocentrism when they were misaligned by 135° from their partners: they were more likely to use egocentric expressions at 135° than at 90°, but no more likely (and, in fact, marginally less likely) to do so at the maximum offset 180°. These findings suggest that this oblique viewpoint is more computationally demanding, at least when producing spatial descriptions (though we find converging evidence from the interpretation of spatial descriptions in ongoing work in our lab).

Our findings offer a further caveat: it is not misalignment alone that ultimately determines the difficulty of perspective-taking, but its combination with other cues. In our study with layouts with an intrinsic structure, speakers made different description choices depending on the alignment of the structure with either partner, despite the partners' misalignment remaining the same. Directors who were at 0° with Matchers at 135° overall opted for their own perspective in descriptions, presumably because reasoning from a perspective oblique to the structure (and their own) was computationally more difficult. However, Directors who were at 225° with Matchers at 0° (also a 135° offset) readily opted for their partner's perspective.

In sum, people do not simply mentally rotate a spatial configuration in order to consider their partner's viewpoint. It is not the case that as the misalignment between partners increases perspective-taking becomes more difficult. Adopting the partner's viewpoint when the partner is misaligned by an oblique offset is generally more difficult than canonical offsets, though not when it is reinforced by other representational cues. The misalignment between partners determines the relative difficulty of perspective-taking for each partner in conjunction with other cues.

PEOPLE SELECT PERSPECTIVES THAT LEAD TO MORE EFFICIENT BUT NOT ALWAYS MORE ACCURATE PERFORMANCE

As we have noted, the adaptation we documented in our studies is consistent with a principle governing human interaction, whereby conversational partners seek to minimize their collective effort and maximize the efficiency of their coordination (Clark and Wilkes-Gibbs, 1986; Clark, 1996). Overall, attributions about the partner's ability to contribute to mutual understanding, enabled by either a priori or perceptual information, lead to strategies that improve task performance. In our studies, recognizing which perspective would be optimal for a particular set of circumstances led to description strategies that were successful at reducing collective effort. Despite the high grounding criterion that pairs had to adopt, given that instructions emphasized accuracy and that speakers could not visually monitor their partner's progress in reconstructing the layout, speakers still managed to select strategies that made interactions efficient.

For instance, pairs took fewer turns to reconstruct randomly configured layouts when they knew in advance that they would be misaligned by an oblique and presumably computationally demanding offset, compared to other orthogonal offsets (Galati and Avraamides, 2012). Under those circumstances, pairs recognized that adopting the perspective of the speaker would be beneficial and were more likely to explicitly agree on that perspective in advance. Thus, when the spatial layout does not afford intrinsic cues, a priori information about the partners' cognitive demands (derived from their relative viewpoints) helps pairs select strategies that make the interaction efficient.

When the layout does afford intrinsic cues, considering the relation of those spatial cues to social cues was critical to determining the optimal perspective. As we've found, interactions took longer in terms of turn-taking when intrinsic cues were not aligned with either partner compared to when they were. And when intrinsic cues converged with the perspective of the partner (vs. the speaker), interactions were somewhat more efficient. This is likely because it was easier for partners to interpret partner-centered expressions (which speakers used almost exclusively when the structure was aligned with the partner) than speaker-centered expressions (which speakers used at greater proportions when they were the ones aligned with the structure).

Nevertheless, although partners made reasonable assumptions about which perspective would be optimal to adopt and although these perspectives minimized their collective effort in terms of their conversational turns, they didn't necessarily improve all aspects of performance on the task. In terms of accuracy, we've found that when the partner was aligned with the layout's structure, reconstructions exhibited a significant rotational bias relative to the other alignment conditions, despite being significantly less distorted the more partner-centered expressions were used. Thus, adopting the partner's perspective in this scenario was an effective strategy in most but not all outcomes.

Adopting the partner's perspective when layouts did not afford an intrinsic structure was actually detrimental to accuracy: reconstructions were more distorted as speakers used more partner-centered expressions. This distortion was curbed somewhat when partners were counteraligned, perhaps because the straightforward mappings of egocentric to other-centered expressions (e.g.,

my left = your right) made it easier for speakers to provide more accurate descriptions, or for partners to interpret speakers' descriptions in the intended way.

Altogether, even though in our studies accuracy was prioritized in pairs' joint goal, it wasn't always achieved perfectly. Whether the source of inaccuracies resides in the speakers' descriptions or the addressees' interpretations remains unresolved. Future research could clarify this by examining task performance against the qualitative content and structure of the pairs' dialogs, beyond just the proportions of speakers' spatial expressions (e.g., high-level description strategies, such separating the table in quadrants). Another methodological consideration for future studies would be to include measures of spatial ability for both collaborating partners. Accounting for some of the variability arising from individual differences in spatial ability can help distinguish whether decrements in accuracy are due to speakers' poor recall and inadequate descriptions or due to partners' misinterpretation of otherwise accurate descriptions. Such efforts would inform the dynamic coupling of partners behaving contingently in joint spatial tasks.

PERSPECTIVE-TAKING BEYOND SPATIAL TASKS

Our framework for spatial perspective-taking reflects the general flexibility of the cognitive system; it is not intended as a framework specialized for or limited to spatial perspective-taking. Our view is that coordination in spatial perspective-taking is governed by some of the same principles as non-spatial perspective-taking—when people consider their conversational partner's conceptual construal, their knowledge, or agenda (see Schober, 1998).

To determine the similarity of their conceptual perspectives, people routinely have to consider what they have in common ground with their conversational partner and to tailor how to produce or interpret utterances. Discrepancies in perspective are especially apparent when there are asymmetries in the partners' respective knowledge or ability, as when one interacts with a non-native speaker (Bortfeld and Brennan, 1997) or a novice (Isaacs and Clark, 1987). Indeed, when people share the same perspective (whether conceptual or physical), it can be trivially easy to adopt the partner's perspective; people can perform generic linguistic or behavioral adjustments (benefiting themselves), rather than adjustments that are specifically designed for their partner (Brown and Dell, 1987; Dell and Brown, 1991). Investigations of partner-specific adaptation should therefore dissociate the perspectives of speakers and their partners (see Keysar, 1997).

Our empirical undertaking to unveil the relation between linguistic perspective choices and the underlying spatial memories that support them is compatible with a memory-based view of partner-specific adaptation. This view considers linguistic and behavioral adjustments to the partner to emerge from cognitive constraints acting on memory-dependent processes (Metzing and Brennan, 2003; Pickering and Garrod, 2004; Horton and Gerrig, 2005). Specifically, shared experiences with a partner and partner-specific associations are considered to be represented in memory and accessed through ordinary processes, such as resonance with combinations of cues in working memory, influencing behavior

accordingly (Horton and Gerrig, 2005). In this view, failures in perspective-taking occur when relevant information about the partner isn't available early enough (Kraljic and Brennan, 2005), when complex inferences about the partner have not yet been made (Gerrig et al., 2000), when executive functioning is taxed (Brown-Schmidt, 2009), or under time pressure (Epley et al., 2004).

Our own findings underscore that simple but relevant cues about the partner (e.g., the partner's location in space, their relation to a configuration's intrinsic structure) can indeed be represented and used to compute the relative difficulty of reasoning from their perspective, consequently determining linguistic choices. This is also in agreement with proposals that when information about the partner is readily available, can be represented simply or computed unambiguously, it can influence language processing at no discernible cost, relative to egocentric processing (Brennan and Hanna, 2009; Galati and Brennan, 2010, 2013).

Our framework is also in line with constraint-based models of language processing (e.g., MacDonald, 1994; Tanenhaus and Trueswell, 1995; McRae et al., 1998). According to constraint-based models, information from various sources, including the discourse context, within-sentence structural, lexical biases, and even information about the partner (e.g., Hanna et al., 2003; Brown-Schmidt and Hanna, 2011), is integrated probabilistically and in parallel to shape the interpretation of utterances, and presumably also speech plans. Similarly, in computational models of perspective-taking, attributions about the partner can be represented as control parameters that can alter behavior (e.g., Duran and Dale, 2013). Other computational accounts also underscore that language processing is adaptive by demonstrating that language users update probability distributions of relevant discourse features (e.g., syntactic structures) as new linguistic evidence becomes available (Fine et al., 2010).

In our work, we've demonstrated that in spatial tasks people indeed use all relevant information from various sources, whether it becomes available at encoding or at collaboration, to form attributions about each partner's relative cognitive effort, which they can update during the course of the interaction and tailor their behavior. This relevant information can include contextual social cues, such as the partner's location in space, or attributional cues, such as beliefs or expectations about the partner's spatial abilities. Such social cues may combine with other cues—intrinsic or functional properties of the objects, the intrinsic structure of the layout or the surrounding environment, one's egocentric viewpoint, and explicit instructions—to determine perspective choices in a constraint-based fashion.

Such an approach departs from proposals that have, on the one hand, acknowledged that the organization of spatial memories depends on the contribution of several cues, but on the other hand, held that certain cues are dominant (Shelton and McNamara, 2001; Mou and McNamara, 2002). Neither egocentric experience (Shelton and McNamara, 2001) nor the intrinsic structure of a spatial configuration (Mou and McNamara, 2002) necessarily need to carry the greatest weight across all tasks. Instead they interact with other weighted parameters, including attributional and contextual cues about the partner.

Finally, our framework for the flexible processing of multiple cues can be extended to non-interactive spatial perspective-taking tasks. We propose that even in non-social situations where people have to imagine adopting different perspectives in space (as when imagining how our redecorated living room would look from different vantage points), the preference for or ease of adopting particular perspectives depends on the confluence of weighted relevant cues.

CONCLUSION

In this article, we have emphasized the centrality of social cues in spatial perspective-taking and have outlined a framework for flexible adaptation of memory and behavior in collaborative spatial tasks. Studying spatial perspective-taking by focusing entirely on individual processes overlooks the ubiquitous and remarkable ability with which people coordinate with one another in a range of everyday activities. The findings emerging from our experimental work underscore people's ability to appraise both social and other representational cues to select the perspective that would be optimal for minimizing their collective effort. Thus, information about the partner (whether derived from the visual context, or from inferences or prior expectations), alongside other cues, can shape how spatial relations are organized in memory and whose perspective is adopted in descriptions. We have argued that this adaptation involves weighing cues according to their relevance and salience to the task, similar to constraint-based approaches, and selecting the perspective most reinforced by the summated contribution of those cues. Moreover, cues are factored into this process whenever they become available—whether through perceptual evidence or advance knowledge. This highlights the flexibility with which people convey information accessed from spatial memory: rather than merely relying on their memory's organization, their assessment of task-specific demands is updated by incoming cues.

Partner-specific adaptation in spatial tasks emerges from processes comparable to those governing non-spatial perspective-taking. This holds both for the principles that regulate the social dynamics of interacting partners (e.g., the principle of least collaborative effort), and for the general cognitive architecture that supports adopting spatial and conceptual perspectives other than one's own. When executive functioning is overloaded, or when relevant cues aren't readily available or easily computed, the ability to appraise the optimal perspective for the joint task is compromised. Partners in perspective-taking tasks—spatial and non-spatial—consider multiple sources of information to make attributions about their respective ability to contribute to mutual understanding. According to these attributions, they adapt how they represent partner-specific information in memory and how they coordinate in dialog.

ACKNOWLEDGMENTS

This material is based upon work supported by the European Research Council under grant 206912-OSSMA. We are grateful to Christina Michael, Chrystalleni Nicolaou, Eleni Xenophontos, and Margarita Antoniou for assistance with data collection and coding, and to Nathan Greenauer and Catherine Mello for useful discussions.

REFERENCES

- Avraamides, M. N., Theodorou, M., Agathokleous, A., and Nicolaou, A. (2013). Revisiting Perspective-taking: can people maintain imagined perspectives? *Spat. Cogn. Comput.* 13, 50–78. doi: 10.1080/13875868.2011.639915
- Bortfeld, H., and Brennan, S. E. (1997). Use and acquisition of idiomatic expressions in referring by native and non-native speakers. *Discourse Process.* 23, 119–147. doi: 10.1080/01638537709544986
- Brennan, S. E. (2005). “How conversation is shaped by visual and spoken evidence,” in *Approaches to Studying World-Situated Language Use: Bridging the Language-as-Product and Language-Action Traditions*, eds J. C. Trueswell and M. K. Tanenhaus (Cambridge, MA: MIT Press), 95–129.
- Brennan, S. E., and Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Top. Cogn. Sci.* 1, 274–291. doi: 10.1111/j.1756-8765.2009.01019.x
- Brennan, S. E., Galati, A., and Kuhlen, A. K. (2010). “Two minds, one dialog: coordinating speaking and understanding,” in *The Psychology of Learning and Motivation*, Vol. 53, ed B. H. Ross (Burlington, MA: Academic Press/Elsevier), 301–345. doi: 10.1016/S0079-7421(10)53008-1
- Brown, P. M., and Dell, G. S. (1987). Adapting production to comprehension: the explicit mention of instruments. *Cogn. Psychol.* 19, 441–472. doi: 10.1016/0010-0285(87)90015-6
- Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychon. Bull. Rev.* 16, 893–900. doi: 10.3758/PBR.16.5.893
- Brown-Schmidt, S., and Hanna, J. E. (2011). Talking in another person's shoes: incremental perspective-taking in language processing. *Dialog Discourse* 2, 11–33. doi: 10.5087/dad.2011.102
- Carlson-Radvansky, L. A., and Irwin, D. E. (1993). Frames of reference in vision and language: where is above? *Cognition* 46, 223–244. doi: 10.1016/0010-0277(93)90011-J
- Carlson-Radvansky, L. A., and Logan, G. D. (1997). The influence of reference frame selection on spatial template construction. *J. Mem. Lang.* 37, 411–437. doi: 10.1006/jmla.1997.2519
- Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.
- Clark, H. H., and Brennan, S. E. (1991). “Grounding in communication,” in *Perspectives on Socially Shared Cognition*, eds L. B. Resnick, J. M. Levine, and S. D. Teasley (Washington, DC: APA Books), 127–149.
- Clark, H. H., and Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *J. Mem. Lang.* 50, 62–81. doi: 10.1016/j.jml.2003.08.004
- Clark, H. H., and Marshall, C. E. (1981). “Definite reference and mutual knowledge,” in *Elements of Discourse Understanding*, eds A. J. Joshi, B. L. Webber, and I. A. Sag (Cambridge: Cambridge University Press), 10–63.
- Clark, H. H., and Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition* 22, 1–39. doi: 10.1016/0010-0277(86)90010-7
- Dell, G. S., and Brown, P. M. (1991). “Mechanisms for listener-adaptation in language production: limiting the role of a model of the listener,” in *Bridges Between Psychology and Linguistics*, eds D. J. Napoli and J. Kegl (San Diego, CA: Academic Press), 105–129.
- Duran, N. D., and Dale, R. (2013). Perspective-taking in dialogue as self-organization under social constraints. *New Ideas Psychol.* 1–16. doi: 10.1016/j.newideapsych.2013.03.004
- Duran, N. D., Dale, R., and Kreuz, R. J. (2011). Listeners invest in an assumed other's perspective despite cognitive cost. *Cognition* 121, 22–40. doi: 10.1016/j.cognition.2011.06.009
- Epley, N., Keysar, B., Van Boven, L., and Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *J. Pers. Soc. Psychol.* 87, 327–339. doi: 10.1037/0022-3514.87.3.327
- Fine, A. B., Qian, T., Jaeger, T. F., and Jacobs, R. A. (2010). “Is there syntactic adaptation in language comprehension,” in *Proceedings of ACL: Workshop on Cognitive Modeling and Computational Linguistics*, ed J. T. Hale (Stroudsburg, PA: Association for Computational Linguistics), 18–26.
- Friedman, A., and Kohler, B. (2003). Bidimensional regression: assessing the configural similarity and accuracy of cognitive maps and other two-dimensional data sets. *Psychol. Methods* 8, 468–491. doi: 10.1037/1082-989X.8.4.468
- Galati, A., and Avraamides, M. N. (2012). “Collaborating in spatial tasks: partners adapt the perspective of their descriptions, coordination strategies, and memory representations,” in *Lecture Notes in Artificial Intelligence: Spatial Cognition*, eds C. Stachniss, K. Schill, and D. Uttal (Heidelberg: Springer), 182–195. doi: 10.1007/978-3-642-32732-2_12
- Galati, A., and Avraamides, M. N. (2013). Collaborating in spatial tasks: how partners coordinate their spatial memories and descriptions. *Cogn. Process.* 14, 193–195. doi: 10.1007/s10339-013-0541-9
- Galati, A., and Brennan, S. E. (2010). Attenuating information in spoken communication: for the speaker, or for the addressee? *J. Mem. Lang.* 62, 35–51. doi: 10.1016/j.jml.2009.09.002
- Galati, A., and Brennan, S. E. (2013). Speakers adapt gestures to addressees' knowledge: implications for models of co-speech gesture. *Lang. Cogn. Process.* doi: 10.1080/01690965.2013.796397
- Galati, A., Michael, C., Mello, C., Greenauer, N. M., and Avraamides, M. N. (2013). The conversational partner's perspective affects spatial memory and descriptions. *J. Mem. Lang.* 68, 140–159. doi: 10.1016/j.jml.2012.10.001
- Gerrig, R. J., Brennan, S. E., and Ohaeri, J. O. (2000). What can we conclude from speakers behaving badly? *Discourse Process.* 19, 173–178.
- Hanna, J. E., Tanenhaus, M. K., and Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *J. Mem. Lang.* 49, 43–61. doi: 10.1016/S0749-596X(03)00022-6
- Hölscher, C., Tenbrink, T., and Wiener, J. M. (2011). Would you follow your own route description? Cognitive strategies in urban route planning. *Cognition* 121, 228–247. doi: 10.1016/j.cognition.2011.06.005
- Horton, W. S., and Gerrig, R. J. (2002). Speakers' experiences and audience design: knowing when and knowing how to adjust utterances to addressees. *J. Mem. Lang.* 47, 589–606. doi: 10.1016/S0749-596X(02)00019-0
- Horton, W. S., and Gerrig, R. J. (2005). The impact of memory demands on audience design during language production. *Cognition* 96, 127–142. doi: 10.1016/j.cognition.2004.07.001
- Isaacs, E. A., and Clark, H. H. (1987). References in conversation between experts and novices. *J. Exp. Psychol.* Gen. 116, 26–37. doi: 10.1037/0096-3445.116.1.26
- Kelly, J. W., Avraamides, M. N., and Loomis, J. M. (2007). Sensorimotor alignment effects in the learning environment and in novel environments. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 1092–1107. doi: 10.1037/0278-7393.33.6.1092
- Keysar, B. (1997). Unconfounding common ground. *Discourse Process.* 24, 253–270. doi: 10.1080/01638539709545015
- Kraljic, T., and Brennan, S. E. (2005). Prosodic disambiguation of syntactic structure: for the speaker or for the addressee? *Cogn. Psychol.* 50, 194–231. doi: 10.1016/j.cogpsych.2004.08.002
- Li, X., Carlson, L. A., Mou, W., Williams, M. R., and Miller, J. E. (2011). Describing spatial locations from perception and memory: the influence of intrinsic axes on reference object selection. *J. Mem. Lang.* 65, 222–236. doi: 10.1016/j.jml.2011.04.001
- MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Lang. Cogn. Process.* 9, 157–201. doi: 10.1080/01690969408402115
- Mainwaring, S. D., Tversky, B., Ohgishi, M., and Schiano, D. J. (2003). Descriptions of simple spatial scenes in English and Japanese. *Spat. Cogn. Comput.* 3, 3–42. doi: 10.1207/S15427633SC0301_2
- McNamara, T. P. (2003). “How are the locations of objects in the environment represented in memory?” in *Lecture Notes in Artificial Intelligence: Spatial Cognition*, eds C. Freksa, W. Brauer, C. Habel, and K. F. Wender (Berlin: Springer-Verlag), 174–191.
- McRae, K., Spivey-Knowlton, M. J., and Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *J. Mem. Lang.* 38, 283–312. doi: 10.1006/jmla.1997.2543
- Metzing, C., and Brennan, S. E. (2003). When conceptual pacts are broken: partner-specific effects on the comprehension of referring expressions. *J. Mem. Lang.* 49, 201–213. doi: 10.1016/S0749-596X(03)00028-7
- Mou, W., and McNamara, T. P. (2002). Intrinsic frames of reference in spatial memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 162–170. doi: 10.1037/0278-7393.28.1.162
- Mou, W., McNamara, T. P., Valiquette, C. M., and Rump, B. (2004a). Allocentric and egocentric updating of spatial memories. *J. Exp. Psychol.*

- Learn. Mem. Cogn.* 30, 142–157. doi: 10.1037/0278-7393.30.1.142
- Mou, W., Zhang, K., and McNamara, T. P. (2004b). Frames of reference in spatial memories acquired from language. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 171–180. doi: 10.1037/0278-7393.30.1.171
- Newman-Norlund, S. E., Noordzij, M. L., Newman-Norlund, R. D., Volman, I. A. C., De Ruiter, J. P., Hagoort, P., et al. (2009). Recipient design in tacit communication. *Cognition* 111, 46–54. doi: 10.1016/j.cognition.2008.12.004
- Özyürek, A. (2002). Do speakers design their co-speech gestures for their addressees? The effects of addressee location on representational gestures. *J. Mem. Lang.* 46, 688–704. doi: 10.1006/jmla.2001.2826
- Pickering, M. J., and Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behav. Brain Sci.* 27, 169–226. doi: 10.1017/S0140525X04000056
- Roche, J., Dale, R., and Kreuz, R. J. (2010). “The resolution of ambiguity during conversation: more than mere mimicry,” in *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society* (Portland, OR).
- Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition* 47, 1–24. doi: 10.1016/0010-0277(93)90060-9
- Schober, M. F. (1995). Speakers, addressees, and frames of reference: whose effort is minimized in conversations about location? *Discourse Process.* 20, 219–247. doi: 10.1080/01638539509544939
- Schober, M. F. (1998). “Different kinds of conversational perspective-taking,” in *Social and Cognitive Psychological Approaches to Interpersonal Communication*, eds S. R. Fussell and R. J. Kreuz (Mahwah, NJ: Lawrence Erlbaum), 145–174.
- Schober, M. F. (2009). “Spatial dialogue between partners with mismatched abilities,” in *Spatial Language and Dialogue*, eds K. R. Coventry, T. Tenbrink, and J. A. Bateman (Oxford: Oxford University Press), 23–39.
- Shelton, A. L., and McNamara, T. P. (2001). Visual memories from nonvisual experiences. *Psychol. Sci.* 12, 343–347. doi: 10.1111/1467-9280.00363
- Shelton, A. L., and McNamara, T. P. (2004). Spatial memory and perspective taking. *Mem. Cogn.* 32, 416–426. doi: 10.3758/BF03195835
- Tanenhaus, M. K., and Trueswell, J. C. (1995). “Sentence comprehension,” in *Handbook of Cognition and Perception*, eds J. L. Miller and P. D. Eimas (San Diego, CA: Academic Press), 217–262.
- Taylor, H. A., and Tversky, B. (1992). Descriptions and depictions of environments. *Mem. Cogn.* 20, 483–496. doi: 10.3758/BF03199581
- Tenbrink, T., Coventry, K. R., and Andonova, E. (2011). Spatial strategies in the description of complex configurations. *Discourse Process.* 48, 237–266. doi: 10.1080/0163853X.2010.549452
- was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 28 June 2013; paper pending published: 16 July 2013; accepted: 08 September 2013; published online: 26 September 2013.

Citation: Galati A and Avraamides MN (2013) Flexible spatial perspective-taking: conversational partners weigh multiple cues in collaborative tasks. *Front. Hum. Neurosci.* 7:618. doi: 10.3389/fnhum.2013.00618

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2013 Galati and Avraamides. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Conflict of Interest Statement: The authors declare that the research



How culture influences perspective taking: differences in correction, not integration

Shali Wu^{1,2*}, Dale J. Barr^{3*}, Timothy M. Gann⁴ and Boaz Keysar⁵

¹ School of Economics and Management, Tsinghua University, Beijing, China

² School of Management, Kyung Hee University, Seoul, South Korea

³ Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, UK

⁴ University of California at Merced, Merced, CA, USA

⁵ Department of Psychology, The University of Chicago, Chicago, IL, USA

Edited by:

Antonia Hamilton, University of Nottingham, UK

Reviewed by:

Elizabeth Sheppard, University of Nottingham Malaysia Campus, Malaysia

Jessica Wang, University of Birmingham, UK

*Correspondence:

Shali Wu, School of Management, Kyung Hee University, 1 Hoegi-Dong, Seoul, 130-701, South Korea
e-mail: shalichina@gmail.com;
Dale J. Barr, Institute of Neuroscience and Psychology, University of Glasgow, 58 Hillhead Street, Glasgow G12 8QB, UK
e-mail: dale.barr@glasgow.ac.uk

Individuals from East Asian (Chinese) backgrounds have been shown to exhibit greater sensitivity to a speaker's perspective than Western (U.S.) participants when resolving referentially ambiguous expressions. We show that this cultural difference does not reflect better integration of social information during language processing, but rather is the result of *differential correction*: in the earliest moments of referential processing, Chinese participants showed equivalent egocentric interference to Westerners, but managed to suppress the interference earlier and more effectively. A time-series analysis of visual-world eye-tracking data found that the two cultural groups diverged extremely late in processing, between 600 and 1400 ms after the onset of egocentric interference. We suggest that the early moments of referential processing reflect the operation of a universal stratum of processing that provides rapid ambiguity resolution at the cost of accuracy and flexibility. Late components, in contrast, reflect the mapping of outputs from referential processes to decision-making and action planning systems, allowing for a flexibility in responding that is molded by culturally specific demands.

Keywords: perspective taking, comprehension, cultural differences, ambiguity, reference

INTRODUCTION

The human language comprehension system is shaped by informational demands related to communication that are relatively universal, as well as by demands of a more social nature that can vary widely across cultures. On the universal side, spoken language is inherently ambiguous at multiple levels, from lexical processing all the way up to the identification of speech acts and resolution of referential ambiguity. In addition, the speech signal itself is evanescent, requiring language comprehenders to rapidly commit to specific parsing decisions and interpretations. On the culturally specific side, cultures vary in the underlying norms and values that regulate social behavior, including norms for participation in conversational interaction. Do the cultural norms governing language and social interaction impact language processing as immediately and as powerfully as the universal demands for rapid ambiguity resolution? Or do they mainly determine how outputs from relatively universal processes are mapped onto later decisions and actions?

One way of addressing these general questions is to compare language users from different cultures in terms of how they integrate social and linguistic information during the online processing of referring expressions. In this study, we investigated cultural differences in how Chinese vs. Western (U.S.) language users take into account a speaker's diverging perspective when they resolve ambiguous references such as *the candle*. Referring expressions are of theoretical interest not only because they are ubiquitous in conversation, but also because they require listeners to go beyond the input – an expression such as the *candle*

denotes a particular class of object, not any particular individual object, and so listeners must access further information to determine which candle is being spoken about. When speakers and listeners have different visual perspectives, reference resolution will only be consistently successful if listeners take these differences into account. It is also methodologically convenient to study visual perspective taking during reference resolution, because a listener's eye gaze during the search for a referent provides an external index of the moment-by-moment process of language interpretation (Cooper, 1974; Tanenhaus et al., 1995). The waxing and waning of referential alternatives during processing will be reflected in moment-by-moment changes in the probability distribution of eye gaze over these alternatives.

Studies of perspective taking during reference resolution have experimentally created differences in perspective between speakers and listeners, and monitored listeners' interpretations as they interpret speakers' instructions to manipulate objects (Keysar et al., 2000; Nadig and Sedivy, 2002; Hanna et al., 2003). These studies suggest that listeners momentarily experience egocentric interference, with listeners considering "privileged" information that they know is unavailable to the speaker. For example, when searching for a referent for the expression *the candle*, listeners will temporarily consider a candle that is hidden from the speaker's view, in spite of their knowledge that the speaker does not know about it and therefore could only be referring to another, mutually visible candle (Keysar et al., 2000, 2003). Ultimately, listeners tend to eventually choose the mutually visible candle, although sometimes they may exhibit signs of confusion. For example, listeners

frequently ask the speaker to clarify the reference, even though if they took the speaker's point of view, they would realize the reference was perfectly clear.

Although the basic phenomenon of egocentric interference has been replicated in numerous studies, recent evidence suggests that it might be specific to the Western (European and North American) populations that have been the traditional object of study (Wu and Keysar, 2007). Cultures differ in the extent to which they emphasize the thoughts and beliefs of the individual versus those of the larger group, with cultures of East Asia exhibiting a more "collectivist" character relative to Western cultures, which tend to be more "individualist" in nature (Triandis et al., 1988; Markus and Kitayama, 1991; Ross et al., 2002). Lifelong membership in a particular culture may shape one's tendency or ability to take another's perspective into account while comprehending language. If so, then people from East Asian backgrounds should show more reliable and effective perspective taking than Westerners in resolving references.

To test this prediction, Wu and Keysar (2007) conducted an eye-tracking study using the basic visual perspective-taking task of Keysar et al. (2000), comparing the performance of Mandarin-speaking Chinese to English-speaking North Americans (from the U.S.). Each group performed the task in its participants' native language. Participants played the role of "listener," sitting across a table from a confederate "director," with a set of shelves placed between them. The contents of some of the shelves were visible from both sides, while others were hidden from the speaker's view. The director had a picture of how the objects in the shelves should be arranged, and told the listener which objects to move and where to move them. Embedded within the interaction were certain pre-scripted test instructions designed to be ambiguous from the listener's perspective, in that they could refer either to a mutually visible "target" object, or a privileged "competitor" object that was visible only to the listener. For example, in one such instruction the director told the listener to "move the candle to the top row," in a context where the listener saw two identical candles, only one of which was visible to the speaker. Listeners' eyes were tracked as they interpreted these test instructions. To provide a baseline, in a control condition, the competitor object was replaced with a non-competitor (an object that did not match the description of the target, such as a toy truck for the "candle" instruction). Egocentric interference would lead to an elevated probability of looking at the hidden competitor (candle) relative to the hidden non-competitor (toy truck), as well as in a delayed latency to fixate on the competitor.

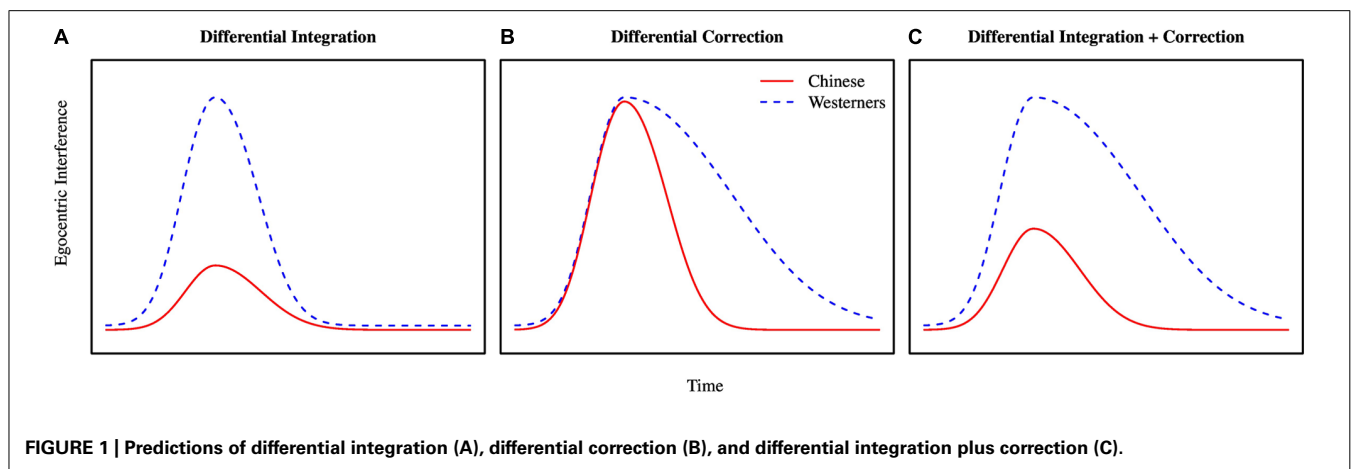
Wu and Keysar (2007) found that while Western participants showed the typical pattern of strong egocentric interference, Chinese participants showed virtually no interference. Unlike their American counterparts, Chinese participants were far less likely to fixate on privileged objects or to ask the speaker to clarify a reference that was ambiguous from their own perspective. In short, the Chinese participants were much more effective overall at taking the speaker's perspective into account.

How might these cultural differences be explained in terms of underlying cognitive processing? Wu and Keysar (2007) measured egocentric interference in terms of first fixation latency and

fixation duration, measures that can detect overall differences between groups, but that do not provide information about when such differences might emerge. To gain further insight into the underlying processes, we reanalyzed the data from Wu and Keysar (2007) using a more time-sensitive analysis in order to investigate the time-course of these cultural differences. Our analysis sought to test whether cultural differences emerged early or late relative to the onset of referential processing. On the one hand, cultural differences in egocentric interference may be present from the earliest moments of referential processing, suggesting that Chinese are able to more effectively use information about perspective to constrain the online processing of referring expression. On the other hand, it is possible that cultural differences emerge late, with both groups showing similar levels of egocentric interference early on, and only diverging later. This latter pattern would imply that the earliest moments of processing are unaffected by social information, and are driven largely by egocentric heuristics that enable rapid ambiguity resolution. Under this view, cultural differences would emerge late because participants from a Chinese background would be faster and more effective than Westerners at suppressing the pragmatically inappropriate information. In other words, cultural differences would not reflect differences in the ability to integrate social information into language processing, but instead would reflect differences in how listeners connect the outcome of basic referential processes to further thought and action.

Having laid out these possibilities in general terms, let us now consider in more detail the nature of the analysis, the possible outcomes, and their implications for theories of language processing and social cognition. Our analysis focused on the temporal profile of egocentric interference across the two cultural groups. We define egocentric interference as the difference in the likelihood of gazing at a hidden competitor (e.g., candle) versus gazing at a hidden non-competitor (e.g., toy truck). Note that we expect interference to show a curvilinear effect over time as shown by the curves in **Figure 1**, climbing from zero up to a peak from which it will eventually drop (as the listener will ultimately ignore the competitor and select the target).

Based on previous literature, we identify three different effect profiles that would be consistent with three different theoretical accounts. The first account, which we term the *differential integration* account, assumes that the cultural difference reflects the enhanced ability of Chinese to integrate information about the speaker's perspective with incoming linguistic information. This account would be consistent with constraint-based models of perspective use in language comprehension (Nadig and Sedivy, 2002; Hanna et al., 2003), as these models assume that information about a speaker's perspective is one of many cues that are simultaneously and interactively integrated during processing. Critically, the account does not differentiate between different types of cues, assuming that any available cue can influence any level of processing from its earliest moments, regardless of its source (e.g., whether it is derived from the unfolding syntax or semantics of the utterance or from situational pragmatics); the influence of a given cue depends only on its salience and reliability. Under this view, the shared perspective between the speaker and listener is a more salient and reliable cue for Chinese than for Westerners. Thus,



Chinese should show less egocentric interference than Westerners from the earliest moments of referential processing – in other words, the onset of the cultural difference should be simultaneous with the onset of the overall effect of egocentric interference (Figure 1A).

An alternative possibility is suggested by the autonomous activation hypothesis of Barr (2008) which, in contrast to constraint-based accounts, assumes that information about a speaker's perspective is a kind of situational cue that influences comprehension through anticipatory or post-lexical decision processing, but is not integrated into online lexical processing. Anticipatory processing refers to those steps taken by the listener in preparation for a referring expression, such as increasing attention to shared (mutually visible) objects. Barr (2008) found that comprehenders strongly *anticipated* that speakers would refer to referential candidates that were shared with the speaker, as evidenced by a higher probability of fixating shared than privileged objects. However, supporting autonomous activation, while interpreting the referring expression, listeners did not show any less interference from privileged than from shared competitors: the probability of gazing at a privileged competitor increased from its (lower) baseline at the same rate as the increase in probability for a shared competitor. Strikingly, in one experiment Barr (2008) found that unlike information about the speaker's perspective, listeners could very efficiently integrate contextual constraints derived from verb semantics. Based on these findings, Barr (2008) argued that lexical processing is encapsulated from high-level information about a speaker's perspective, and perhaps from other kinds of situational information, but is not strictly modular in the sense of being completely cognitively impenetrable.

The autonomous activation account would predict that Chinese participants might be more sensitive overall to a speaker's perspective, but without showing any greater ability to integrate this information with the linguistic input. Under this view, they should experience comparable levels of egocentric interference to Westerners, at least during the earliest moments of comprehension. In the current paradigm, this difference would be expressed as *differential correction*: Chinese participants would not initially experience less egocentric interference, but would be faster and

more effective at suppressing this interference than Westerners (Figure 1B)¹. Of course, the integration and correction accounts are not mutually exclusive. A third possibility would be that the groups differ in both integration and correction, such that not only do Chinese participants experience lower interference from the earliest moments of comprehension, but they also are more efficient at suppressing this interference (Figure 1C). This pattern would be consistent with constraint-based models.

MATERIALS AND METHODS

Additional details regarding experimental and data collection procedures are available in the original report (Wu and Keysar, 2007).

Our analyses considered looks to the competitor/non-competitor object from 250 ms after the onset of the critical word (e.g., the word “candle” in the phrase “move the candle...”) until 3000 ms. Observations for a given trial were terminated when listeners touched the target. These points varied from trial to trial, with a median of 3306 ms (2808 vs. 3844 for Chinese vs. U.S. participants, respectively), and a standard deviation of 4729 ms. For those trials that were terminated before 3000 ms, we replaced the missing frames with 0 s (representing the absence of a look to the competitor/non-competitor object).

Our goal was to test whether there was a time-lag between the onset of egocentric interference and the onset of cultural differences. To give an overview of our analysis method, we applied the *cluster randomization* method that has become popular in neuroimaging for determining the spatial and temporal extent of experimentally induced effects (Bullmore et al., 1999; Maris and Oostenveld, 2007; for prior adaptation of the solution to the analysis of visual-world data, see Barr et al., 2013). This approach is attractive for localizing effects in time in a visual-world study because it takes advantage of temporal correlations among adjacent data points to overcome the problem of multiple comparisons. The approach proceeds as follows. First, a significance test is performed at each time slice for a given effect (e.g., main effect

¹Unlike Barr (2008), the current study does not offer the possibility of determining whether listeners were attentionally biased toward shared referential alternatives before hearing the referring expression. This would require additional conditions in which the competitor or non-competitor would be shared with the speaker.

or interaction). Then, “clusters” are defined by identifying adjacent time slices where the effect reaches significance, and where all effects are in the same direction. For example, consider tests performed at six subsequent time slices, t_1 , t_2 , t_3 , t_4 , t_5 , and t_6 , with tests significant at the 0.05 level only at t_2 , t_3 , t_5 , and t_6 . If t_2 and t_3 have effects in the same direction, then they form a cluster; likewise, if t_5 and t_6 are in the same direction, they also form a cluster. There are two separate clusters rather than a single one because of the intervening non-significant test at t_4 . Once the clusters have been identified, a “cluster mass statistic” is calculated for each one, typically the sum of all of the individual test statistics (e.g., t values) for that cluster. One obtains a null-hypothesis distribution for this cluster mass statistic through randomization (permutation tests); i.e., by randomly shuffling the condition labels across trials to create a large number of new datasets, repeating the above procedure on these datasets, and then storing the maximum obtained cluster mass statistic for each one. Finding a significant cluster between t_i and t_j with, say, 1000 additional randomized datasets and $p = 0.002$ means that the cluster mass statistic for the original data was matched or exceeded in only 2 of the 1000 randomly created datasets.

We did this procedure twice, once to test for the main effect of Competition (competitor vs. non-competitor, e.g., egocentric interference), and once to test for the Culture-by-Competition interaction. The cluster randomization procedure provides only p -values; however, we were also interested in defining confidence limits for our effects. To obtain these confidence limits we used bootstrapping (details below). The remainder of this section provide further technical details regarding how these analyses were implemented.

Rather than comparing the observed probabilities at each time point, we fit a time-series model to the data and compared predictions from the model, following Barr et al. (2013). The time-series model smooths the data over time, thus minimizing noise and facilitating the detection of clusters (see **Figure 2B**). In the model, time was represented as a 7th order polynomial. We determined the order of the polynomial using a model search procedure, in which we calculated the Akaike Information Criterion (AIC) value for all models ranging from a 3rd order to a 16th order polynomial, and then selected the model with the lowest AIC, which was a 7th order polynomial. This was done on the grand-averaged data (i.e., without any predictors for Competition or Culture) so as not to bias the cluster randomization procedure.

Logistic regression models were fit to the data using the `multinom()` procedure of the `nnet` package (Venables and Ripley, 2002) of R statistical software (R Core Team, 2013), treating the outcome for each sample as binary. The cluster randomization procedure was performed twice, once treating subjects as random and items as fixed (p_1), and once treating items as random and subjects as fixed (p_2). For simplicity, we describe the procedure treating subjects as random factors. In addition to the parameter estimates from a fit to the original data, we created 999 additional data sets by randomly permuting the condition labels (competitor vs. control) independently for each unit (subject or item). To obtain orthogonality between the main effects and the interaction, the relabeling followed a “synchronized” permutation logic

(Pesarin, 2001). For a given culture group, a permutation was created by randomly choosing, with equal probability, whether or not to block-exchange all competitor and non-competitor labels for each subject. The same number of exchanges was then performed for the other culture group (with the units undergoing the exchange chosen at random). The “synchronization” of exchanges across groups (i.e., ensuring that the same number of exchanges of Competition occurs at each of the two levels of Group) ensures that the tests for the main effects and interaction are orthogonal (Pesarin, 2001). The parameter estimates for the model fit to each of these data sets were stored as a row in the matrix.

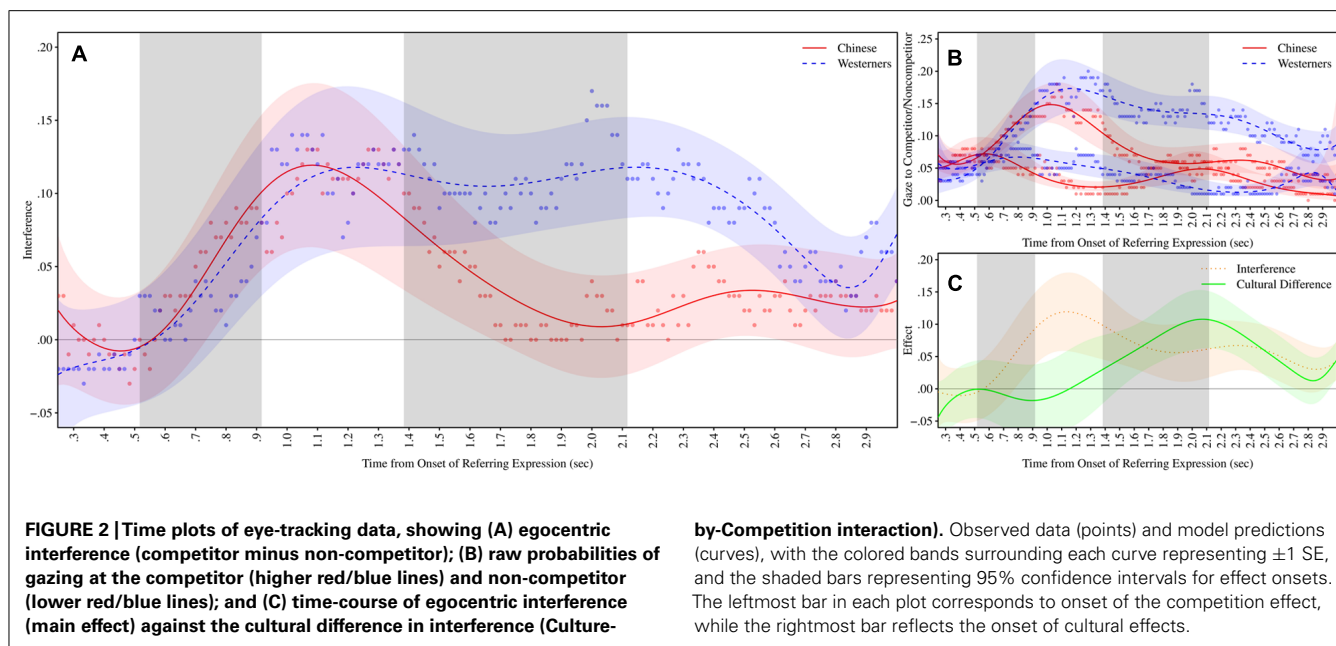
After all datasets were created, we then calculated the predicted log odds of a gaze to the competitor/non-competitor object at each 1/60 of a second (i.e., for each frame of data sampled at 60 Hz), deriving main effects and interactions at each time point. The predicted effects for each of the 1000 datasets (including the original) were stored as separate rows in a matrix. The p -value for each effect (main effect of competition or interaction) at each time point was given as the number of rows in the effect matrix exceeding the original value divided by the number of rows in the matrix. Then, we identified clusters by grouping together all temporally adjacent time-frames where the effect reached significance. A cluster mass statistic (Bullmore et al., 1999) was calculated for each cluster by summing together the negative (natural) logarithm of each p -value belonging to the cluster, such that smaller p -values would contribute to a larger cluster mass statistic; for example, for 0.05 the negative log is 3, and for 0.0001 it is 9.21. This cluster mass statistic was calculated for each cluster in the original data. Then, a null-hypothesis distribution for the statistic was derived by treating each permuted data set as if it was the “original” data, calculating p -values and cluster mass statistics in the manner described above, and storing the maximum observed statistic for each permutation. This allowed us to identify the onset of the first significant cluster for both the main effect as well as for the interaction.

To obtain confidence limits, we repeated the complete analysis described above for 999 bootstrapped versions of the data set, wherein we sampled subjects with replacement from each group at random. If for a given bootstrapped dataset, no significant cluster for the main effect or interaction was detected at $\alpha = 0.05$, the α level was progressively lowered until a cluster was detected, stopping at $\alpha = 0.2$. Although the confidence limits derived from bootstrapping provide useful information, the main inferential focus is on the results of the cluster randomization on the original data.

RESULTS

The time-course data appear in **Figure 2**. Note that 0 ms does not correspond to the onset of the utterance (e.g., “move” in “move the candle”), but to the onset of the referring expression within the utterance (e.g., “candle”). Thus any differences in timing between groups cannot be attributed to possible linguistic differences between Chinese and English in the duration of the material preceding the referring expression.

The cluster randomization procedure detected significant overall egocentric interference from 750 to 2800 ms, with the 95% confidence interval for the onset of interference ranging from 517



to 917 ms ($p_1 < 0.001$, $p_2 = 0.003$)². As **Figure 2** clearly shows, there was a large time-lag between the onset of egocentric interference and the onset of a cultural difference in this interference (as given by a Culture-by-Competition interaction). There was little evidence that Chinese participants experienced any less interference than U.S. participants until 1767 ms, approximately 1000 ms after the onset of interference. The Culture-by-Competition interaction was significant from 1767 to 2483 ms ($p_1 = 0.009$, $p_2 = 0.049$), with the 95% confidence interval for the onset ranging from 1383 to 2117 ms. Note that there was no overlap between the confidence interval for the onset of the cultural difference (1383–2117) with that for the onset of egocentric interference (517–917). Furthermore, we directly computed the delay between the onsets for each bootstrapped sample, which yielded a 95% confidence interval for the lag between 600 and 1400 ms.

DISCUSSION

Overall, our findings support the hypothesis that language users from different cultures share a common stratum of referential processing, with cultural variation in how the products of these early referential processes are used in the higher-level processes governing thought and action. Specifically, whereas neither Chinese nor Western participants were able to integrate the situational cue of the speaker's perspective into lexical processing, Chinese participants were better able to suppress the interference.

Could our findings of common interference and differential correction be alternatively explained in terms of linguistic differences between Mandarin Chinese and English? One potentially

relevant difference is that Mandarin lacks definite marking, such that the Mandarin version of the English expression “move the candle” might be glossed in English as “move candle.” It might be argued that the Chinese participants were interpreting the descriptions as if the speaker had said, “move any candle.” This would indeed predict that the Chinese participants would experience less interference than the U.S. participants because they would not need to decide between the two possible referents, but could pick either one. However, if this were the case, then Chinese participants should have shown a stronger tendency than U.S. participants to move the hidden candle, since any candle would suffice. However, the data showed the exact opposite. While the U.S. participants sometimes moved the occluded candle, the Chinese participants never did.

One possible concern might be that the later correction for Chinese participants reflects shorter referring expressions in Chinese, or more rapid speech when the confederate spoke Chinese. Although we lack the data to directly address this question, the overall patterns shown in **Figure 2** make this explanation seem unlikely. First, if the earlier correction occurred because the Chinese expressions were briefer or spoken more rapidly, then not only would the correction process take place earlier, but so would the egocentric interference; specifically, the initial rising slope of the curve should have been much steeper for the Chinese group than for the Western group, and should have reached its peak much earlier. However, egocentric interference seems to rise at similar rates for both groups, and both seem to initially reach their maximum values at roughly the same time (1000–1200 ms). Second, whereas the correction process seems to begin at around 1000 ms for the Chinese group, it seems delayed until about 2200 ms for the American group. This is far too great of a disparity to be explained by differences in the spoken expressions, given that expressions in these types of experiments typically last no more than 1 s. Finally, the groups differ not only in the timing of the correction, but also

²Although an onset of 750 ms is quite late relative to typical visual-world studies (250–350 ms), this is not surprising given that our paradigm presented participants with a more demanding search task than in a typical study. Whereas a typical grid in Wu and Keysar (2007) contained nine alternatives appearing in any of 16 possible locations, a typical visual-world task presents no more than four referential alternatives in fixed locations (Huetting et al., 2011).

in the *efficacy* of the correction, with a sudden sharp decline for the Chinese group, and more of a lingering pattern for the Western group. Thus, these patterns seem less likely to be driven by differences in the stimuli, and more likely to reflect true cultural differences in linguistic interpretation.

Constraint-based views would have difficulty accounting for the extreme delay in the emergence of cultural differences relative to the onset of egocentric interference. If, as constraint-based views predict, language users can integrate perspective information from the earliest moments of processing, and Chinese participants attend more strongly to the shared perspective than Westerners, then Chinese participants should have shown less egocentric interference from the very earliest moments of processing. Our view, then, is that despite attending more strongly to shared information, Chinese participants are no better at integrating it into referential processing. However, an alternative view must be considered, which is that perhaps the late emergence does not reflect a standalone correction process, but simply reflects delayed activation of shared information relative to other kinds of information. Under this view, had the shared knowledge become activated earlier, perhaps we would have seen its effects earlier in processing. However, it is unclear what would account for the delayed activation of shared knowledge within the current paradigm. For one, in the current experimental situation, listeners knew well before hearing the referring expression which items their partner could see and which they could not see. In other words, information about what was shared was available to participants even before any referential information became available. It is therefore not clear why listeners would wait for a referring expression to activate the shared knowledge, rather than using it to predict potential referents in advance. It is not possible to tell whether listeners in fact made such predictions, because this requires comparing shared to privileged objects, and our analysis only considered privileged objects. However, experiments using a similar setup have found that in the interval preceding the onset of the referring expression, listeners are more likely to look at shared objects (Keysar et al., 2000). Furthermore, recent experiments including conditions where competitors/non-competitors are shared show that listeners spontaneously access shared knowledge prior to the onset of referring expressions, but are unable to integrate this information into early referential processes (Barr, 2008). Specifically, listeners attend less overall to privileged objects than to shared objects, but nonetheless experience similar levels of interference from competitors regardless of whether they are shared or not. It would be of interest to repeat these experiments with East Asian participants. Our account predicts greater access to shared knowledge among East Asians, but without any reduction in the size of the interference produced by competitors.

Our view that information about perspective is involved in correction is consistent with an anchoring and adjustment view of perspective taking (Keysar et al., 2000), in which listeners anchor interpretation in their own perspectives, and use information about the speaker's perspective to incrementally adjust away from the anchor. However, distinct from Keysar et al.'s (2000) original formulation, our findings, together with those of Barr (2008), suggest that listeners do not strategically "anchor" in their own

egocentric perspective as a kind of reasoning heuristic; rather, their anchoring is forced upon them by the autonomous activation of referents by low-level interpretation processes that are blind to information about the speaker's perspective (Barr, 2008). Under this view, the noted egocentrism of listeners might be best characterized as a form of "mental contamination" – i.e., the result of rapid, automatic processes that are beyond control and possibly even awareness (Wilson and Brekke, 1994).

Consistent with the use of common ground in correction, other research shows that perspective taking involves cognitive effort (Rosnagel, 2000; Brown-Schmidt, 2009; Nilsen and Graham, 2009; Lin et al., 2010), and recent neuroimaging evidence suggests a role for the medial pre-frontal cortex in the adjustment process (Tamir and Mitchell, 2010). Furthermore, the correction account is also consistent with dual process views of perspective taking, which assume that social judgments reflect the combination of both efficient but inflexible processing that uses limited information and more flexible but effortful processing that can draw upon a broader set of information (Apperly and Butterfill, 2009). However, the current data offer no insight into why the adjustment process might differ across the groups. One possibility, consistent with the collectivist vs. individualist distinction, is that information about a speaker's perspective is simply more available to people from a collectivist background, since their cultures require greater attunement to one another's knowledge. Another is that perhaps Chinese participants are more motivated to perform the task "correctly" due to heightened concerns about self-presentation. A further possibility is that membership in a Chinese culture, where self-control is valued, results in better executive control abilities. This explanation is supported by research that finds enhanced executive control abilities among Chinese as opposed to North American children (Sabbagh et al., 2006), who nonetheless showed comparable performance on a belief reasoning task. As we have argued here and elsewhere (Keysar et al., 2003; Barr, 2008) listeners' difficulty in identifying the intended referent in conversational perspective-taking tasks is unlikely to be the result of a failure to have the appropriate beliefs about what is shared with the speaker. Instead, it seems to reflect difficulty using this information to constrain the processing of the linguistic input. To the extent that early referential processes are not guided by beliefs about the speaker, these processes will boost activation of referents that are pragmatically implausible, even in spite of correct and accessible representations of shared knowledge. Because suppressing this knowledge will involve executive control, it is here where we would expect to see strong individual (and cultural) differences. Although in this respect our view is consistent with Sabbagh et al.'s (2006) developmental findings, it is important to note that it is not yet known whether the differences in executive function that Sabbagh et al. (2006) noted extend into adulthood.

Whatever the explanation for the cultural differences, a recent study suggests that it might be possible to induce cultural effects through priming. Luk et al. (2012) replicated Wu and Keysar's (2007) study but with Chinese-Westerner bi-cultural individuals. Participants primed by images from Western culture committed more egocentric errors on the perspective-taking task relative to participants who were primed by images from Chinese culture.

The fact that cultural differences can be situationally induced in bicultural individuals suggests that they arise from flexible modes of processing. This flexibility is consistent with our explanation of such differences in terms of differential correction – it would seem easier to override a deliberative and effortful correction process than an integration process that is largely routinized and automatic.

In sum, our data suggest that people from different cultures share a common core of ambiguity resolution processes, but differ in how the output from these processes is linked to higher-level systems governing thought and action. The two cultures we have studied show systematic differences in how they prioritize the individual vs. the social (Triandis et al., 1988; Markus and Kitayama, 1991; Ross et al., 2002). Finding equivalent interference from privileged information in spite of such differences suggests that such egocentrism might be a universal consequence of rapid ambiguity resolution during spoken language comprehension.

ACKNOWLEDGMENTS

This research was supported by National Science Foundation of China Grants 71002014 and 71110107027 to Shali Wu. Boaz Keysar also received partial support from a grant from the University of Chicago's Wisdom Research Project and the John Templeton Foundation.

REFERENCES

- Apperly, I. A., and Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychol. Rev.* 116, 953–970. doi: 10.1037/a0016923
- Barr, D. J. (2008). Pragmatic expectations and linguistic evidence: listeners anticipate but do not integrate common ground. *Cognition* 109, 18–40. doi: 10.1016/j.cognition.2008.07.005
- Barr, D. J., Jackson, L., and Phillips, I. (2013). Using a voice to put a name to a face: the psycholinguistics of proper name comprehension. *J. Exp. Psychol. Gen.* doi: 10.1037/a0031813 [Epub ahead of print].
- Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychon. Bull. Rev.* 16, 893–900. doi: 10.3758/PBR.16.5.893
- Bullmore, E. T., Suckling, J., Overmeyer, S., Rabe-Hesketh, S., Taylor, E., and Brammer, M. J. (1999). Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. *IEEE Trans. Med. Imaging* 18, 32–42. doi: 10.1109/42.750253
- Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: a new methodology for the real-time investigation of speech perception, memory, and language processing. *Cogn. Psychol.* 6, 84–107. doi: 10.1016/0010-0285(74)90005-X
- Hanna, J. E., Tanenhaus, M. K., and Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *J. Mem. Lang.* 49, 43–61. doi: 10.1016/S0749-596X(03)00022-6
- Huetting, F., Rommers, J., and Meyer, A. S. (2011). Using the visual world paradigm to study language processing: a review and critical evaluation. *Acta Psychol.* 137, 151–171. doi: 10.1016/j.actpsy.2010.11.003
- Keysar, B., Barr, D. J., Balin, J. A., and Brauner, J. S. (2000). Taking perspective in conversation: the role of mutual knowledge in comprehension. *Psychol. Sci.* 11, 32–38. doi: 10.1111/1467-9280.00211
- Keysar, B., Lin, S., and Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition* 89, 25–41. doi: 10.1016/S0010-0277(03)00064-7
- Lin, S., Keysar, B., and Epley, N. (2010). Reflexively mindblind: using theory of mind to interpret behavior requires effortful attention. *J. Exp. Soc. Psychol.* 46, 551–556. doi: 10.1016/j.jesp.2009.12.019
- Luk, K. K. S., Xiao, W. S., and Cheung, H. (2012). Cultural effect on perspective taking in Chinese–English bilinguals. *Cognition* 124, 350–355. doi: 10.1016/j.cognition.2012.05.016
- Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190. doi: 10.1016/j.jneumeth.2007.03.024
- Markus, H., and Kitayama, S. (1991). Culture and the self: implications for cognition, emotion, and motivation. *Psychol. Rev.* 98, 224–253. doi: 10.1037/0033-295X.98.2.224
- Nadig, A. S., and Sedivy, J. C. (2002). Evidence of perspective-taking constraints on children's on-line reference resolution. *Psychol. Sci.* 13, 329–336. doi: 10.1111/j.0956-7976.2002.00460.x
- Nilsen, E., and Graham, S. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cogn. Psychol.* 58, 220–249. doi: 10.1016/j.cogpsych.2008.07.002
- Pesarin, F. (2001). *Multivariate Permutation Tests: With Applications in Biostatistics*. Chichester: Wiley.
- R Core Team. (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Ross, M., Xun, W., and Wilson, A. (2002). Language and the bicultural self. *Pers. Soc. Psychol. Bull.* 28, 1040–1050. doi: 10.1177/01461672022811003
- Rossnagel, C. (2000). Cognitive load and perspective taking: applying the automatic-controlled distinction to verbal communication. *Eur. J. Soc. Psychol.* 30, 429–445. doi: 10.1002/(SICI)1099-0992(200005/06)30:3<429::AID-EJSP3>3.0.CO;2-V
- Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., and Lee, K. (2006). The development of executive functioning and theory of mind: a comparison of Chinese and U.S. preschoolers. *Psychol. Sci.* 17, 74–81. doi: 10.1111/j.1467-9280.2005.01667.x
- Tamir, D. I., and Mitchell, J. P. (2010). Neural correlates of anchoring-and-adjustment during mentalizing. *Proc. Natl. Acad. Sci. U.S.A.* 107, 10827. doi: 10.1073/pnas.1003242107
- Tanenhaus, M. K., Spivey, M. J., Eberhard, K. M., and Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science* 268, 1632–1634. doi: 10.1126/science.7777863
- Triandis, H. C., Bontempo, R., Villareal, M. J., Asai, M., and Lucca, N. (1988). Individualism and collectivism: cross-cultural perspectives on self-in-group relationships. *J. Pers. Soc. Psychol.* 54, 323–338. doi: 10.1037/0022-3514.54.2.323
- Venables, W. N., and Ripley, B. D. (2002). *Modern Applied Statistics with S*, 4th Edn. New York: Springer. doi: 10.1007/978-0-387-21706-2
- Wilson, T. D., and Brekke, N. (1994). Mental contamination and mental correction: unwanted influences on judgments and evaluations. *Psychol. Bull.* 116, 117–142. doi: 10.1037/0033-2909.116.1.117
- Wu, S., and Keysar, B. (2007). The effect of culture on perspective taking. *Psychol. Sci.* 18, 600–606. doi: 10.1111/j.1467-9280.2007.01946.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 30 June 2013; accepted: 13 November 2013; published online: 02 December 2013.

Citation: Wu S, Barr DJ, Gann TM and Keysar B (2013) How culture influences perspective taking: differences in correction, not integration. *Front. Hum. Neurosci.* 7:822. doi: 10.3389/fnhum.2013.00822

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2013 Wu, Barr, Gann and Keysar. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Developmental changes in mental rotation ability and visual perspective-taking in children and adults with Williams syndrome

Masahiro Hirai^{1*†}, Yukako Muramatsu¹, Seiji Mizuno², Naoko Kurahashi², Hirokazu Kurahashi² and Miho Nakamura^{1*}

¹ Department of Functioning Science, Institute for Developmental Research, Aichi Human Service Center, Aichi, Japan

² Department of Pediatrics, Central Hospital, Aichi Human Service Center, Aichi, Japan

Edited by:

Antonia Hamilton, University of Nottingham, UK

Reviewed by:

Massimiliano Conson, Second University of Naples, Italy
Danielle Ropar, University of Nottingham, UK

*Correspondence:

Masahiro Hirai, Center for Development of Advanced Medical Technology, Jichi Medical University, 3311-1 Yakushiji, Shimotsuke, Tochigi 329-0498, Japan
e-mail: hirai@jichi.ac.jp;
Miho Nakamura, Department of Functioning Science, Institute for Developmental Research, Aichi Human Service Center, 713-8 Kagiya-cho, Kasugai, Aichi 480-0392, Japan
e-mail: mihon@inst-hsc.jp

† Present address:

Masahiro Hirai, Center for Development of Advanced Medical Technology, Jichi Medical University, Shimotsuke, Tochigi, Japan

Williams syndrome (WS) is a genetic disorder caused by the partial deletion of chromosome 7. Individuals with WS have atypical cognitive abilities, such as hypersociability and compromised visuospatial cognition, although the mechanisms underlying these deficits, as well as the relationship between them, remain unclear. Here, we assessed performance in mental rotation (MR) and level 2 visual perspective taking (VPT2) tasks in individuals with and without WS. Individuals with WS obtained lower scores in the VPT2 task than in the MR task. These individuals also performed poorly on both the MR and VPT2 tasks compared with members of a control group. For the individuals in the control group, performance scores improved during development for both tasks, while the scores of those in the WS group improved only in the MR task, and not the VPT2 task. Therefore, we conducted a second experiment to explore the specific cognitive challenges faced by people with WS in the VPT2 task. In addition to asking participants to change their physical location (self-motion), we also asked them to adopt a third-person perspective by imagining that they had moved to a specified location (self-motion imagery). This enabled us to assess their ability to simulate the movement of their own bodies. The performance in the control group improved in both the self-motion and self-motion imagery tasks and both performances were correlated with verbal mental age. However, we did not find any developmental changes in performance for either task in the WS group. Performance scores for the self-motion imagery task in the WS group were low, similar to the scores observed for the VPT2 in this population. These results suggest that MR and VPT2 tasks involve different processes, and that these processes develop differently in people with WS. Moreover, difficulty completing VPT2 tasks may be partly because of an inability of people with WS to accurately simulate mental body motion.

Keywords: Williams syndrome, visual perspective taking, mental rotation, developmental trajectory, children, developmental disorder, reference frame

INTRODUCTION

Williams syndrome (WS) is a rare genetic disorder caused by the deletion of approximately 25 genes on chromosome 7. The prevalence of WS is between 1:20000 and 1:7500 (Stromme et al., 2002; Meyer-Lindenberg et al., 2006). Although there is heterogeneity in the cognitive domains that are affected by WS (Porter and Coltheart, 2005), several specific cognitive strengths and weaknesses have been consistently reported in this population (Bellugi et al., 2000; Meyer-Lindenberg et al., 2006; Martens et al., 2008; Riby and Porter, 2010; Jarvinen et al., 2013). For instance, the literature suggests that while language and auditory abilities are generally preserved (Bellugi et al., 1990; Karmiloff-Smith et al., 1997; Jordan et al., 2002; Brock, 2007), elements of visuospatial cognition, such as perceptual grouping, mental imagery, and global motion processing, are impaired (Bellugi et al., 1988; Pezzini et al., 1999; Farran et al., 2001; Nakamura et al., 2001; Atkinson et al., 2003, 2006; Hoffman et al., 2003; Farran and

Jarrold, 2004, 2005). The observed deficits in visuospatial processing in people with WS may be due to atypical processing in the construction, but not the modality, of perception (Farran and Jarrold, 2003; Hoffman et al., 2003). Some evidence has also suggested that such visuospatial deficits extend to the memory domain (Vicari et al., 2003, 2005), and may, for instance, include the abnormal representation of reference frames (Nardini et al., 2008).

Neuroimaging research has indicated that visuospatial deficits in individuals with WS may be caused by a dysfunctional dorsal stream (Atkinson et al., 1997). Several atypical cortical structures have been observed in this population, such as (1) a low density of gray matter in the superior parietal regions (Reiss et al., 2004; Eckert et al., 2005), including the intraparietal sulcus (Meyer-Lindenberg et al., 2004), (2) bilateral reductions in the depth of the intraparietal/occipitotemporal sulci (Kippenhan et al., 2005) compared with controls, and (3) prominent folding abnormalities

in the dorsal parietal cortex (Van Essen et al., 2006). Atypical fractional anisotropy in the right superior longitudinal fasciculus, which is associated with deficits in visuospatial construction, has also been reported in individuals with WS (Hoeft et al., 2007).

One prominent social phenotype of people with WS is that they display an empathetic nature and an extreme interest in both familiar and unfamiliar people. This particular trait has been termed “hypersociability” (Jones et al., 2000). Individuals with WS are often able to retrieve explicit emotional information from facial expressions (Gagliardi et al., 2003; Plesa-Skwerer et al., 2006; Skwerer et al., 2006) and perceive human actions from point-light motion (Jordan et al., 2002; Reiss et al., 2005; Hirai et al., 2009). However, the ability of this population to interpret emotional states seems to be atypical, such that they may have difficulty understanding unfamiliar facial expressions (Frigerio et al., 2006; Porter et al., 2007) or retrieving information about intent from motion (Van Der Fluitt et al., 2012).

People with WS may have difficulty inferring the thoughts or emotions of others, although research on this issue has produced unclear results. An early WS study reported that individuals with this disorder perform well in a location change task (Karmiloff-Smith et al., 1995). Another study found that half of a group of people with WS performed similarly to normal adults on a task where participants were asked to identify complex emotional states from photographs of eyes (Tager-Flusberg et al., 1998). However, later studies of children with WS found impaired mentalizing ability (Tager-Flusberg et al., 1997; Sullivan and Tager-Flusberg, 1999; Tager-Flusberg and Sullivan, 2000). Porter et al. (2008) reported a specific deficit in social understanding in one of two WS subgroups, indicated by poor performance on a non-verbal version of the theory of mind (ToM) task. This effect persisted even when the effects of mental or chronological age were removed. This finding suggests cognitive heterogeneity in the social cognition of individuals with WS (Porter and Coltheart, 2005).

Accumulating evidence shows that individuals with WS have atypical cognitive abilities, such as hypersociability and impaired visuospatial cognition. However, the mechanisms underlying these deficits are unclear, as is the relationship between impaired social cognition and impaired visuospatial cognition.

Visual perspective taking tasks can be used to assess connections between visuospatial and social cognitive processes. Visual perspective taking has two levels: Level 1 visual perspective taking (VPT1) refers to knowledge about which objects in one's frame of view are visible to another observer, while Level 2 visual perspective taking (VPT2) refers to the knowledge that two different observers can have unique visual experiences of the same scene or object (Flavell et al., 1984). Developmental psychological studies have shown that both levels are not acquired simultaneously. Infants are first able to understand VPT1 at approximately 24 months (Moll and Tomasello, 2006). It is not until later, in the preschool period, that individuals are able to understand VPT2 (Flavell, 1999). For instance, a recent study reported that 3-year-old children are able to successfully complete a VPT2 task (Moll and Meltzoff, 2011).

Several studies have investigated the connection between different characteristics of cognitive tasks. For instance, one

behavioral study reported a clear relationship between the performance of children aged 4–8 years on a ToM and a VPT2 task, but not between a ToM and a mental rotation (MR) task (Hamilton et al., 2009). This suggests that ToM and VPT2 tasks may have common cognitive processes that may not be required for MR tasks. Therefore, the VPT2 task may be useful in assessing mentalizing ability in individuals with WS. The notion that ToM and VPT2 tasks may have common cognitive processes has been supported by several neuroimaging findings. For instance, in adults, the temporoparietal junction (TPJ) is activated by VPT2 tasks (Zacks et al., 2003b; Aichhorn et al., 2006) and false-belief tasks (Saxe and Kanwisher, 2003). The importance of the TPJ for performance on the above-mentioned tasks has been demonstrated by lesion studies (Apperly et al., 2004) and transcranial direct current stimulation studies (Santesteban et al., 2012). However, these studies reported no overlap in terms of the neural activities underlying the VPT2 and MR tasks, indicating that differential brain networks are involved.

The current study comprised two experiments. The first focused on developmental changes in MR and VPT2 task performance in individuals with WS, and employed tasks developed by Hamilton et al. (2009). In Experiment 1, we hypothesized that, (1) in light of previous findings regarding deficient visuospatial skills in individuals with WS, this population would have impaired MR ability compared with normal controls, and (2) if individuals with WS exhibited impaired mentalizing ability (Tager-Flusberg et al., 1997; Tager-Flusberg and Sullivan, 2000; Porter and Coltheart, 2005; Porter et al., 2008), then VPT2 task performance would be poor compared with normal controls.

In our preliminary experiment, we found that members of the WS group consistently had difficulties completing the VPT2 task. Therefore, our second experiment was designed to explore the nature of these difficulties. Although a recent neuroimaging study has demonstrated that different brain regions are involved in the spatial transformation of oneself vs. another person (Mazzarella et al., 2013), behavioral evidence suggests that spatial perspective taking is an embodied cognitive process, in the sense that the participant's own body posture can interfere with performance on a VPT2 task. This implies that cognitive processes underlying spatial transformation of oneself and of others may overlap (e.g., Kessler and Thomson, 2010). Thus, differential performance on VPT2 and spatial transformation tasks could help to explain the difficulty observed in the VPT2 task in Experiment 1.

In Experiment 2, we manipulated the location of the participants with respect to an object (first-person location). We asked the participants to either move to a new position or to imagine that they had moved. Both manipulations were designed to match the difficulty of the procedure in Experiment 1. If the expected difficulties in VPT2 task completion in Experiment 1 were due to defective mental body motion simulation in people with WS, then this would reflect performance on the self-motion imagery task.

MATERIALS AND METHODS (EXPERIMENT 1)

PARTICIPANTS

Twenty-six people with WS (13 males and 13 females) participated in the experiments (Table 1). Twenty participants were recruited from our institute, and six were recruited through the

Table 1 | Participants.

Group	N (F/M)	Chronological age mean (year) range (years; months)	Verbal mental age mean (year) range (years; months)
WS	26 (13/13)	16.2 ± 7.2 (6;0–33;4)	7.46 ± 2.46 (3;3–11;1)
VMA	26 (13/13)	7.67 ± 2.7 (3;9–13;2)	7.62 ± 2.5 (3;5–12;3)
CA	26 (13/13)	16.3 ± 8.4 (6;5–40;6)	N/A

(Mean ± SD).

Williams Syndrome Association in Aichi prefecture (Elfin Chubu, Nagoya). All participants had been phenotypically diagnosed by clinicians, with their diagnoses confirmed through positive fluorescence *in situ* hybridization testing. The ages of the participants ranged from 6 years 0 months to 33 years 5 months (mean age = 16 years and 2 months). Verbal intelligence was measured with the Japanese version of the Picture Vocabulary Scale (JPVS) (Ueno et al., 2008).

Fifty-two typically developed children, adolescents, and adults were recruited from elementary schools, junior high schools, and universities near the institute as control groups (Table 1). For the verbal mental age-matched (VMA) group, 26 children (13 males) were selected to match individual JPVS scores obtained from participants with WS. For the chronological age-matched (CA) group, the ages of the control participants were individually matched to the ages of the participants with WS.

ETHICAL CONSIDERATIONS

All children, their parents, and adult participants provided informed consent. The study protocol was approved by the Ethics Committee at the Institute for Developmental Research in the Aichi Human Service Center.

THEORY OF MIND TESTING

As in previous studies (Tager-Flusberg and Sullivan, 2000; Hamilton et al., 2009), we conducted the location change task (Wimmer and Perner, 1983; Baron-Cohen et al., 1985) and the unexpected contents task (Hogrefe et al., 1986) prior to conducting the MR and VPT2 tasks in the WS and VMA groups. Both tasks were scored such that one point was given when a participant successfully completed a ToM task; otherwise the score remained at 0. Because all of the participants in the CA group were above 6 years of age, they easily passed the ToM tasks. Thus, we did not include their performance on these tasks in the analysis.

MENTAL ROTATION TASK AND LEVEL 2 VISUAL PERSPECTIVE TAKING TASK

As in a previous study (Hamilton et al., 2009), we conducted two experimental tasks (MR and VPT2) in same session, with a short (a few minutes) break between them. We performed three familiarization trials to familiarize the participant with the experimental settings prior to the first session. At the beginning of each familiarization trial, a small toy (a dog) was placed on a square turntable, which had distinctly colored sides. The participant was shown a piece of paper in a transparent folder (to prevent any

damage to the paper) with four pictures of the toy, taken from four perspectives (front, back, left, and right). The participant was then asked: “Which dog are you looking at?” The participant was instructed to point to the picture that matched the perspective of the toy as it appeared on the turntable. After the participant pointed to one of the four pictures, the toy was covered with a transparent bucket, and the participant was asked: “When I lift the bucket, which dog will you see?” If the participant made errors during the trials, the experimenter corrected them. We initially found that the familiarization task was difficult for young children with WS, so we decided to use a transparent bucket.

Following the familiarization session, we conducted six trials for each task (MR and VPT2). The task order was counterbalanced across participants. For each task, we put a toy in either a front or back position for three trials, and then in a profile position for three trials. We used six different toys (one for each task; car, dump truck, loading shovel, reindeer, panda, and owl) to prevent the participant from remembering the position of each toy, and to draw their attention to the toy during the experiment. The response sheet contained four pictures of each toy, taken from four perspectives. These were placed in a random order to exclude any response bias effects.

For the MR task, the experimenter told each participant to “watch carefully” and then placed a new toy on the table. The experimenter then showed the participant the response sheet and asked them to point to the picture that matched the position of the toy. This ensured that the participant was paying attention to the toy. The experimenter covered the toy with an opaque bucket and turned the table 90° clockwise, 180°, or 90° counter-clockwise. After turning the table, the experimenter asked the child: “If I lift the bucket, which “toy name” (i.e., “Panda” in Figure 1A) will you see?” The participant was instructed to point to the picture that they thought matched the position of the toy (Figure 1A).

For the VPT2 task, the experimenter placed a toy on the table and told the participant to “watch carefully.” The experimenter then gave the participant the response sheet and asked them to point to the picture that matched that position of the toy. The experimenter covered the toy with the opaque bucket, took out a doll from behind their back, and placed it on the left, right, or far side of the table, away from the participant. The experimenter then shook the doll side to side to draw the participant’s attention, and asked: “This is Ai-chan; when I lift the bucket, which “toy name” (i.e., “Panda” in Figure 1B) will Ai-chan see?” Emphasis was put on the word “Ai-chan” when asking the question. The experimenter asked the participant to point to the picture that matched the perspective of the toy that the doll would see (Figure 1B).

The experiment was performed in a quiet playroom at our institute. During the sessions, the experimenter provided motivational feedback to the participant (e.g., “You are doing well!”) to keep their attention focused on the task, irrespective of their responses. We did not give any feedback regarding accuracy to the participants, and the experimenter told the participants that there was no time limit within which they had to respond.

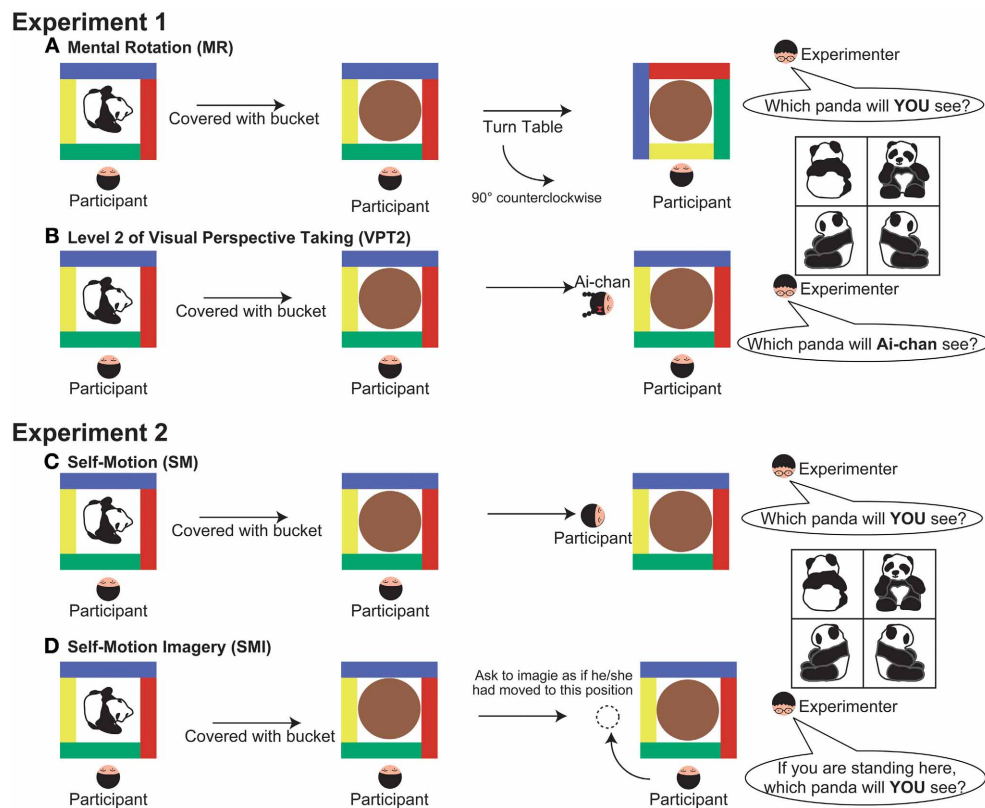


FIGURE 1 | Detail of the experimental procedure (a top-down view).

(A) mental rotation (MR), (B) level 2 of visual perspective taking (VPT2), (C) self-motion (SM), and (D) self-motion-imagery (SMI). In all experimental conditions, the experimenter began by confirming that the participants were

attending to the orientation of the toy. After covering the bucket and completing the experimental manipulation, the experimenter asked the participants to point to one of the pictures on the response sheet, shown at the right side of the figure.

DATA ANALYSIS

We counted the number of participants who successfully completed each ToM task. Chi-square analysis was used to assess performance across groups.

As per previous studies, we focused on correct answer responses (Hamilton et al., 2009) and error responses (e.g., Samson et al., 2007) when analyzing the data from Experiment 1. Our preliminary observations suggested that younger children tend to show an egocentric bias (i.e., even when the doll was placed in a different position, their response was identical to the response they gave before the toy was covered with the bucket) during the VPT2 task, as previously depicted in the three-mountain paradigm (Piaget and Inhelder, 1956). We defined this type of error as “egocentric-bias error”; and any other error was defined as a “non-egocentric-bias error.” In our analysis, we calculated the proportion of egocentric errors (the proportion of egocentric errors made in relation to the overall number of errors).

For statistical analysis, we applied a two-way mixed-design repeated measures analysis of variance (ANOVA) to the correct answers and the proportion of egocentric-bias errors. Group (WS, VMA, and CA groups) was used as a between-subject factor, and Task (MR and VPT2) was used as a within-subject factor.

We also analyzed correct responses based on performance in the two ToM tasks. In this analysis, we focused on the data from the WS and VMA groups, because the participants in the CA group were all older than 6 years, as mentioned above. We defined participants who passed both ToM tasks (i.e., the score was 2 points) as members of the ToM pass group. The two ToM tasks had similar levels of difficulty, and so a participant who passed one but not the other may just have been guessing. We applied a Three-Way mixed-design repeated measures ANOVA to the correct responses. Group (WS and VMA) and ToM performance (Pass group and Fail group) were used as between-subject factors, and Task (MR and VPT2) was used as a within-subject factor.

If the sphericity assumption was violated, as indicated by Mauchly’s sphericity test, then the Greenhouse–Geisser epsilon coefficient was used to correct the degrees of freedom. Tukey’s honestly significant difference test was applied for multiple comparisons. The *F* and *P*-values were then recalculated. A *P*-value of < 0.05 was considered statistically significant.

In addition to these analyses, we adopted a developmental trajectory approach (Thomas et al., 2009) to assess developmental changes in task performance in both the WS and VMA groups. We did not include the CA group in this analysis because their

performance scores reached a ceiling level, and therefore, further developmental changes could not be observed. For this analysis, we calculated coefficients and evaluated improvements in performance based on developmental changes in verbal mental age.

RESULTS (EXPERIMENT 1)

THEORY OF MIND TESTING

A comparison of the location change task scores from the three groups revealed a significant difference in performance [$\chi^2_{(1)} = 4.16, p < 0.05$]. Further binomial testing revealed that significantly more than half of the participants in the VMA group passed the test ($p < 0.01$), while this was not the case in the WS group ($p = 0.17$). A comparison of the unexpected contents task scores also revealed a significant difference in performance [$\chi^2_{(1)} = 11.5, p < 0.01$]. Further binomial testing revealed that significantly more than half of the participants in the VMA group passed the test ($p < 0.01$), while this was not the case in the WS group ($p = 1.0$). The results indicate that significantly more participants in the VMA group passed the ToM tasks compared with the WS group. Conversely, significantly more participants in the WS group failed the ToM tasks compared with the VMA group (Table 2).

MENTAL ROTATION TASK AND LEVEL 2 VISUAL PERSPECTIVE TAKING TASK

To examine performance on the MR and VPT2 tasks, we first compared the number of correct responses in each group (Figure 2A). We observed significant effects of Group [$F_{(2, 75)} = 39.8, p < 0.01$] and Task [$F_{(1, 75)} = 50.7, p < 0.01$], and a significant two-way interaction between Group \times Task [$F_{(2, 75)} = 5.8, p < 0.01$]. Subsequent follow-up analyses revealed that performance on the MR task was significantly greater than performance on the VPT2 task for participants in the WS [$F_{(1, 75)} = 35.6, p < 0.01$] and VMA [$F_{(1, 75)} = 24.1, p < 0.01$] groups. No significant differences were observed in the CA group [$F_{(1, 75)} = 1.9, p = 0.17$].

In terms of group differences, we observed that the performance of the WS group was worse than the performance of the VMA ($p < 0.01$) and CA groups ($p < 0.01$) on the MR task, although we found no difference between the VMA and CA groups ($p = 0.07$). For the VPT2 task, performance scores from the CA group were significantly better than scores from the VMA ($p < 0.01$) and WS groups ($p < 0.01$). Performance scores from the VMA group were significantly better than performance scores from the WS group ($p < 0.01$).

In all groups, MR task scores were significantly above chance [CA: $t_{(25)} = 60.2, p < 0.01$; VMA: $t_{(25)} = 12.9, p < 0.01$; WS: $t_{(25)} = 4.7, p < 0.01$]. In contrast, the scores from the WS group

on the VPT2 task were not significantly better than chance [$t_{(25)} = 1.6, p = 0.13$]. The scores from the VMA group on the VPT2 task [$t_{(25)} = 3.2, p < 0.01$] and CA [$t_{(25)} = 12.8, p < 0.01$] were significantly above chance.

We also examined the proportion of egocentric-bias errors (Figure 2B). The effects of Group [$F_{(2, 75)} = 7.06, p < 0.01$] and Task [$F_{(1, 75)} = 59.2, p < 0.01$] were significant, but the two-way interaction between Group \times Task [$F_{(2, 75)} = 1.10, p = 0.34$] was not. This suggests that the proportion of egocentric-bias errors in the VPT2 task was significantly higher than that in the MR task, for all groups. In terms of group differences, the proportion of egocentric-bias errors in both the WS and VMA groups ($p < 0.01$) was significantly higher than that in the CA group, for both tasks. However, no significant differences were observed between the WS and VMA group.

Regarding ToM task performance (Figure 3), we found that the main effects of Group [$F_{(1, 48)} = 4.31, p < 0.05$], ToM

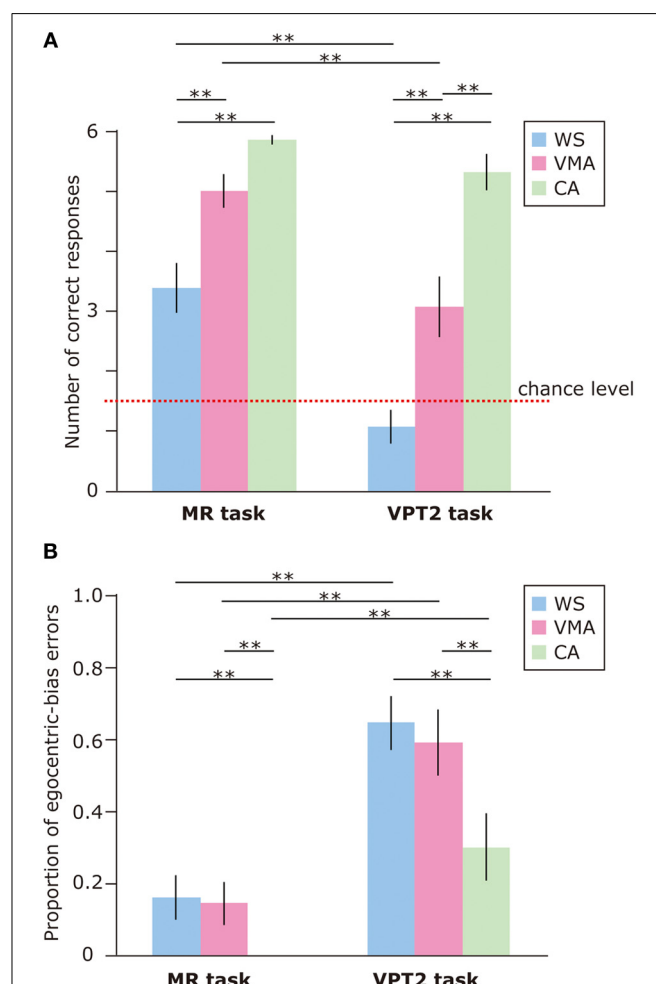


FIGURE 2 | (A) Mean number of correct trials (max: 6) in the MR and VPT2 tasks for three groups [blue: Williams syndrome (WS) group; pink: verbal mental age-matched (VMA) group; green: chronological age-matched (CA) group]. **(B)** Mean proportion of egocentric errors (the proportion of egocentric errors made in relation to the overall errors) for both tasks. Error bars indicate standard error. ** $p < 0.01$.

Table 2 | Performance on two theory of mind tasks.

Group	Unexpected contents task	Location change task
WS	9/26	13/26
VMA	22/26	21/26

(Number of participants who passed the task/ Number of participants).

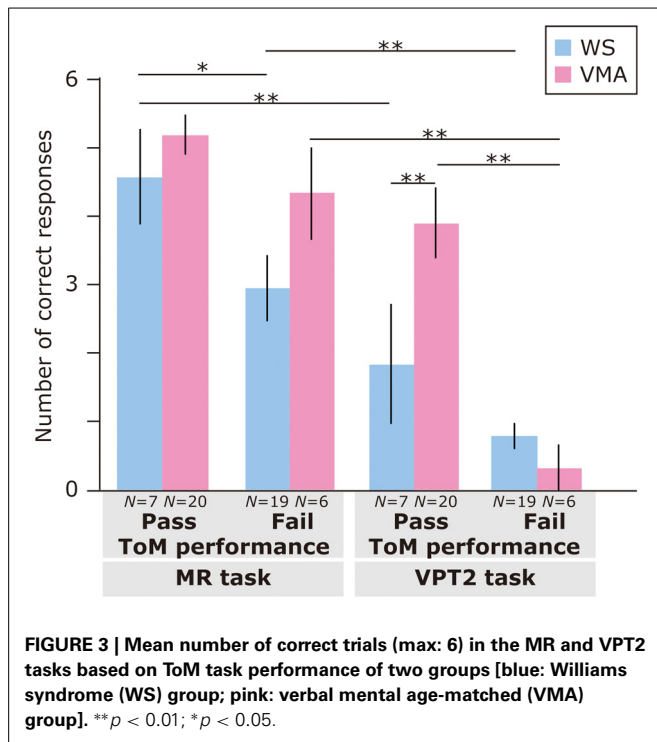


FIGURE 3 | Mean number of correct trials (max: 6) in the MR and VPT2 tasks based on ToM task performance of two groups [blue: Williams syndrome (WS) group; pink: verbal mental age-matched (VMA) group]. ** $p < 0.01$; * $p < 0.05$.

performance [$F_{(1, 48)} = 16.9$, $p < 0.01$], and Task [$F_{(1, 48)} = 58.5$, $p < 0.01$] were significant. Moreover, a three-way interaction of Group \times ToM performance \times Task [$F_{(1, 48)} = 6.0$, $p < 0.01$] was significant.

A follow-up analysis revealed that, for the VMA group, there were significantly more correct responses on the VPT2 task in the ToM pass group than in the ToM fail group [$F_{(1, 96)} = 21.3$, $p < 0.01$]. This was not the case for the MR task [$F_{(1, 96)} = 1.26$, $p = 0.27$]. For the WS group, there were significantly more correct responses on the MR task in the ToM pass group than in the ToM fail group [$F_{(1, 96)} = 4.41$, $p < 0.05$]. This effect was not observed for the VPT2 task [$F_{(1, 96)} = 1.90$, $p = 0.17$].

Regarding group differences, the VMA children who passed both ToM tasks had a significantly higher rate of correct VPT2 task performance than the individuals with WS who passed both ToM tasks [$F_{(1, 96)} = 7.0$, $p < 0.01$]. All other effects were not significant (all F s < 3.2 , p s > 0.08).

Regarding differences in performance across tasks, the WS group obtained significantly more correct answers in the MR task than in the VPT2 task, regardless of ToM task performance [ToM pass group: $F_{(1, 48)} = 16.6$, $p < 0.01$; ToM fail group: $F_{(1, 48)} = 10.5$, $p < 0.01$]. For the VMA participants, the above was true for the ToM fail group [$F_{(1, 48)} = 36.2$, $p < 0.01$], but not the ToM pass group [$F_{(1, 48)} = 3.8$, $p = 0.06$] in the VMA group.

We used a developmental trajectory approach to explore developmental changes in the WS and VMA groups in terms of correct and egocentric-bias error responses for both tasks (Figure 4). For the WS group, we observed a significant positive correlation between verbal mental age and performance on the MR ($r = 0.47$, $p = 0.01$) but not the VPT2 task ($r = 0.02$, $p = 0.91$). For the VMA group, we observed significant positive correlations between verbal mental age and performance for both the MR and

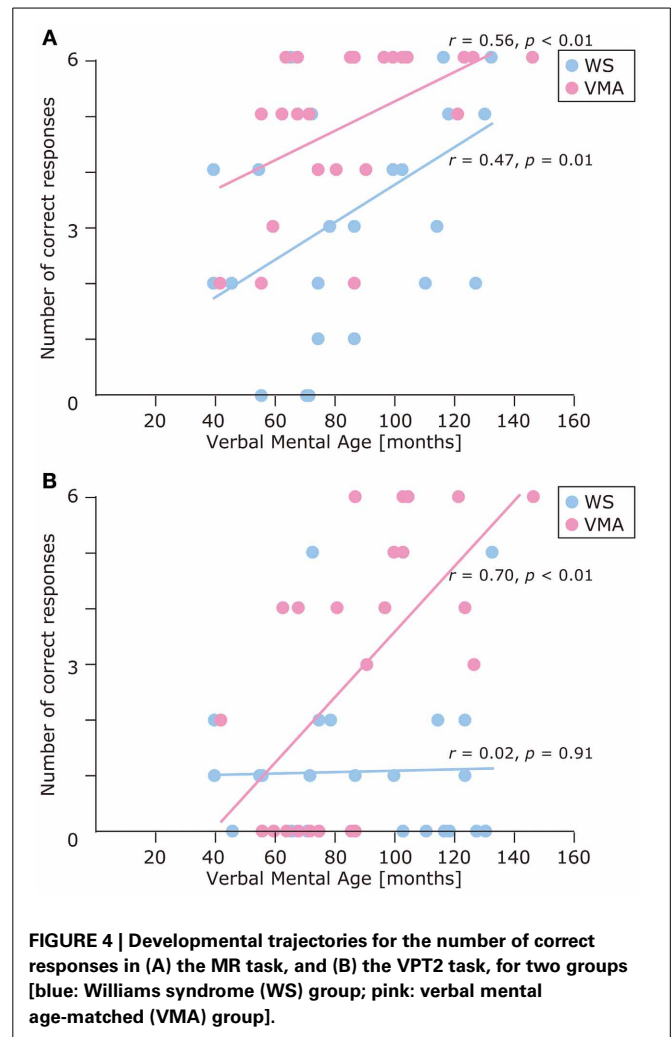


FIGURE 4 | Developmental trajectories for the number of correct responses in (A) the MR task, and (B) the VPT2 task, for two groups [blue: Williams syndrome (WS) group; pink: verbal mental age-matched (VMA) group].

VPT2 tasks (MR task; $r = 0.56$, $p < 0.01$; VPT2 task; $r = 0.70$, $p < 0.01$). In terms of egocentric-bias errors in the WS group, we did not observe any significant correlations (MR: $r = -0.25$, $p = 0.22$; VPT2: $r = 0.01$, $p = 0.97$). For the VMA group, we observed a significant negative correlation for the VPT2 ($r = -0.57$, $p < 0.01$) but not the MR task ($r = -0.24$, $p = 0.24$) (Figure 5).

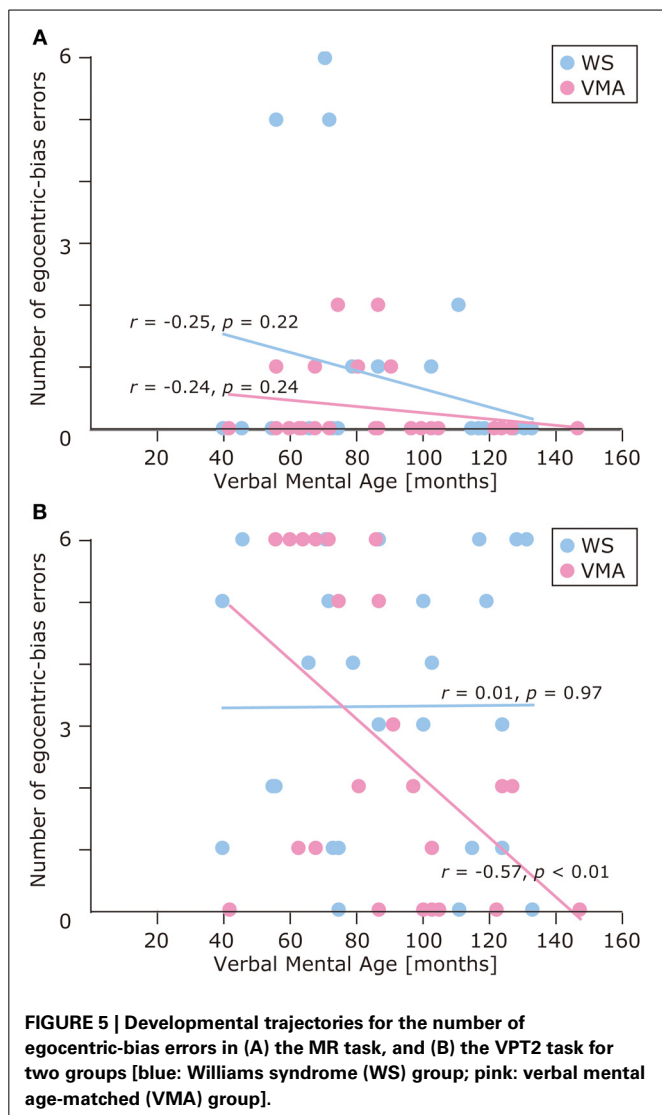
MATERIALS AND METHODS (EXPERIMENT 2)

PARTICIPANTS

The participants that took part in Experiment 1 also took part in Experiment 2 (Table 1).

SELF-MOTION TASK AND SELF-MOTION-IMAGERY TASK

In Experiment 1, we found that the VPT2 task was more difficult for individuals with WS than the MR task. The performance of the WS group on the VPT2 task did not improve across development, in contrast with performance on the MR task. This motivated us to conduct a further experiment to explore alternative explanations for the observed difficulty, such as impaired mental simulation of one's own body motion. Behavioral evidence suggests that spatial perspective taking is an embodied cognitive



process (Kessler and Thomson, 2010). Imagining one's own bodily motion can induce activation in distinct cortical regions, such as the left posterior parietal cortex (Creem et al., 2001), or supplementary motor areas (Wraga et al., 2005). Although these findings suggest that the demands of the VPT2 task include embodiment processes, it is likely that the neural activities involved in imagining one's own bodily motion are distinct from those activated by the VPT2. Thus, if we observed differential performance between VPT2 tasks and tasks requiring one to imagine the motion of their body, this might help to explain the difficulty observed in completing the VPT2 task in Experiment 1. To verify this possibility, we designed an experiment in which we manipulated the position (perspective) of the participant, instead of asking the participant to imagine a third-person perspective, as in Experiment 1. In Experiment 2, therefore, we introduced two experimental tasks, self-motion (SM) and self-motion imagery (SMI), in an attempt to match the task difficulty to that of Experiment 1.

For the SM condition, the experimenter placed a toy on a table and asked the participant to point to the picture on the response

sheet (described in the methods for Experiment 1) that matched the position of the toy. This was done to make sure that the participant was paying attention to the toy. The experimenter then covered the toy with the opaque bucket and another experimenter gently took the participant's arms or shoulders to guide them in changing his or her location (to the left, right, or far side of the table with respect to the original position). After guiding the participant to the new position, the experimenter asked: "If I lift the bucket, which 'toy name' (i.e., "Panda" in Figure 1C) will you see?" The participant was instructed to point to the picture that matched the perspective of the toy that they would see from their new position (Figure 1C).

For the SMI condition, the procedure was the same as in the SM condition, except that the experimenter pointed to a location (left, right, or far side of the table) instead of guiding the participant to that position. Before pointing to the location, the experimenter made sure that the participant understood the concept of imagining self-movement. The experimenter then asked the participant: "If you moved to this position and I lifted the bucket, which 'toy name' (i.e., "Panda" in Figure 1D) would you see?" The participant was asked to point to the picture that matched the perspective of the toy that they would see from their new imagined position (Figure 1D).

Other than those detailed above, the experimental procedures were identical to those in Experiment 1. Six trials were performed for each task and the task order was counterbalanced across participants. The experiment was conducted in the same room as Experiment 1.

DATA ANALYSIS

As in Experiment 1, a Two-Way ANOVA was applied to the correct responses and the proportion of egocentric-bias errors. In the analysis, Group (WS, VMA, and CA groups) was used as a between-subject factor, and Task (SM and SMI) was used as a within-subject factor.

In addition to the ANOVA, we used the same methods as in Experiment 1 to analyze correct and incorrect ToM task responses for the WS and VMA groups. For each ToM task, a three-way mixed-design repeated measures ANOVA was applied to the correct responses. Group (WS and VMA) and ToM performance (Pass group and Fail group of participants) were used as between-subject factors, and Task (SM and SMI) was used as a within-subject factor. If the sphericity assumption was violated as per Mauchly's sphericity test, then the Greenhouse–Geisser epsilon coefficient was used to correct the degrees of freedom. Both the *F* and *P*-values were then recalculated. A *P*-value of < 0.05 was considered statistically significant.

In addition to these analyses, we adopted a developmental trajectory approach to assess developmental changes in task performance for both the WS and VMA groups (Thomas et al., 2009). As in Experiment 1, we did not apply this analysis to the CA group because their performance scores reached a ceiling level, thus, preventing further developmental changes from being observed. For this analysis, we calculated coefficients and evaluated improvements in performance based on developmental changes in verbal mental age.

RESULTS (EXPERIMENT 2)

To examine performance on the SM and SMI tasks, we applied an ANOVA to the number of correct responses (**Figure 6A**). We found that the effects of Group [$F_{(2, 75)} = 59.8, p < 0.01$] and Task [$F_{(1, 75)} = 6.7, p < 0.05$] were significant. A two-way interaction between Group \times Task was marginally significant [$F_{(2, 75)} = 2.5, p = 0.09$]. This suggests that performance on the SM task was significantly better than performance on the SMI task, for all groups. With respect to group differences, the CA group performed significantly better than the VMA ($p < 0.01$) and WS ($p < 0.01$) groups. Performance in the VMA group was better than performance in the WS group ($p < 0.01$).

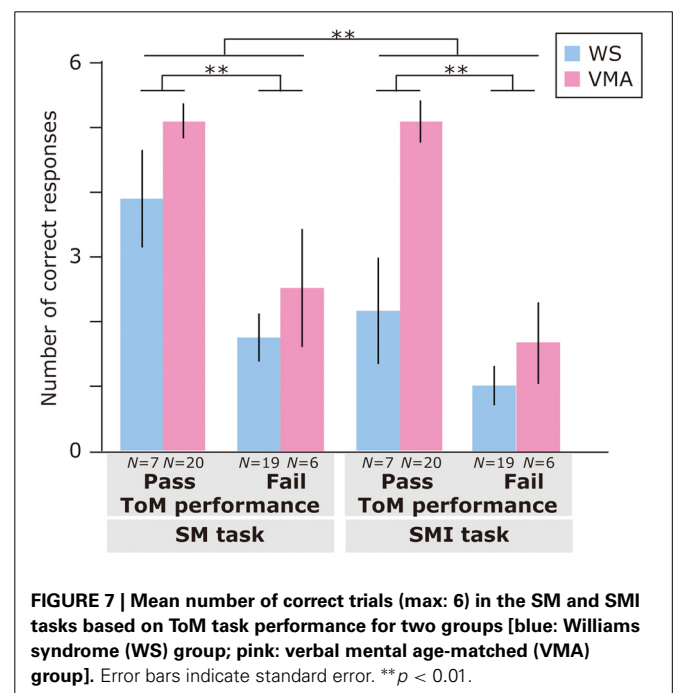
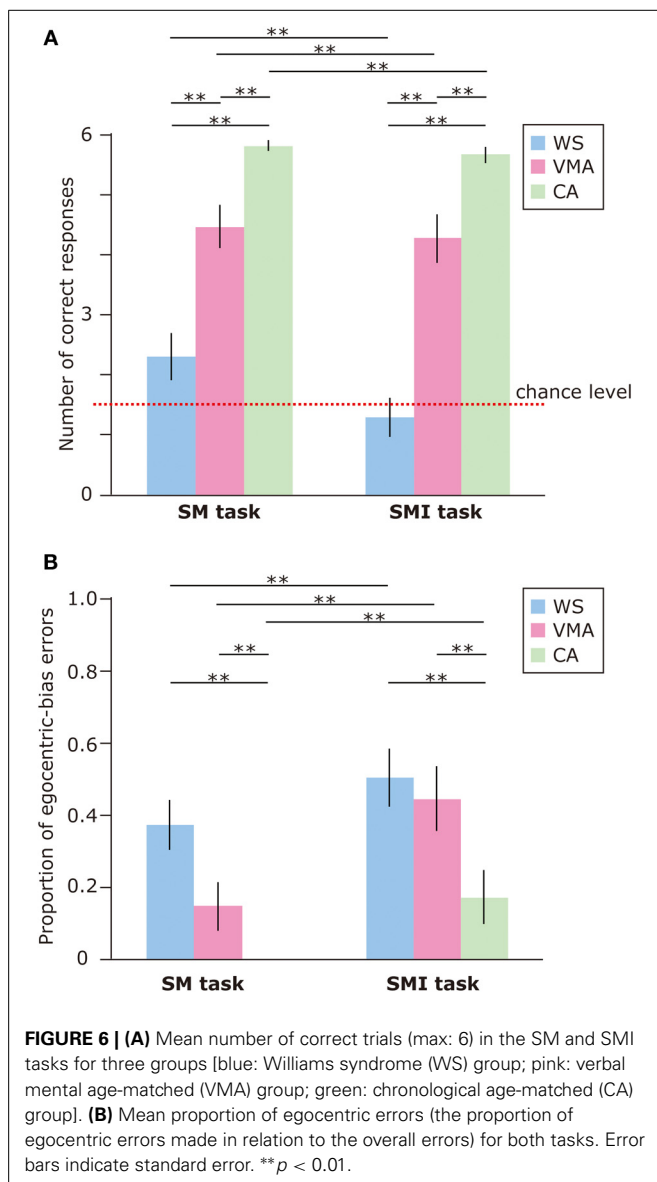
We also examined the proportion of egocentric-bias errors (**Figure 6B**). The effects of Group [$F_{(2, 75)} = 10.4, p < 0.01$] and Task [$F_{(1, 75)} = 18.7, p < 0.01$] were significant, but the two-way interaction between Group \times Task [$F_{(2, 75)} = 1.18, p = 0.31$] was

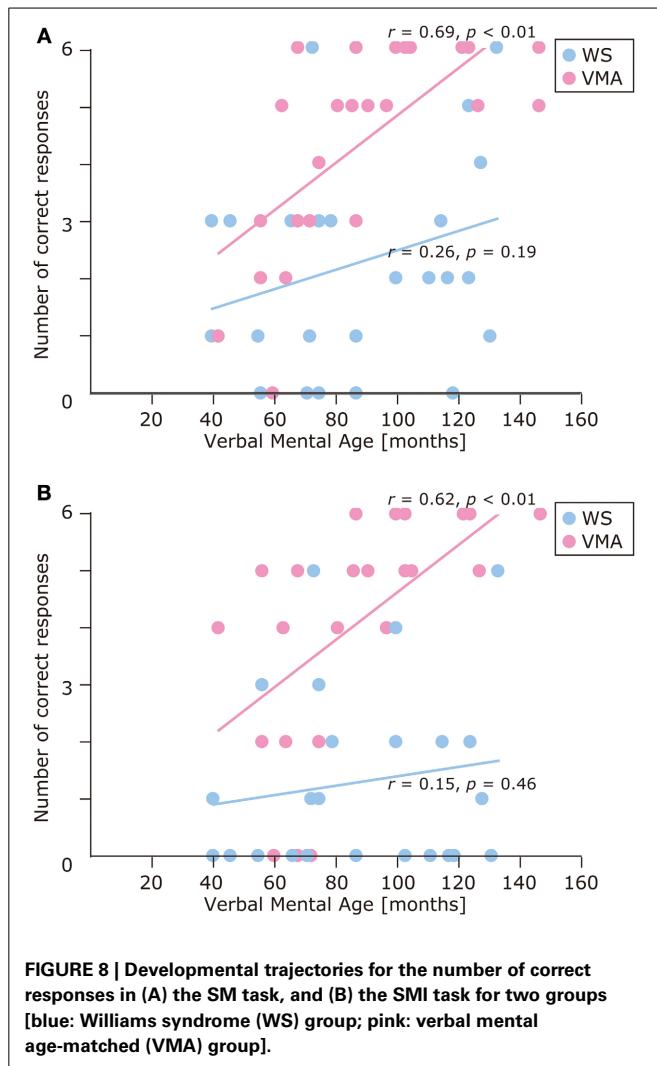
not. This indicates that the proportion of egocentric-bias errors was significantly higher in the SMI task than the SM task, for all groups. With respect to group differences, the proportion of egocentric-bias errors in both the WS ($p < 0.01$) and VMA ($p < 0.01$) groups was significantly higher than that in the CA group. However, no significant differences were observed between the WS and VMA groups.

In all groups, SM task performance was significantly above chance [CA: $t_{(25)} = 54.7, p < 0.01$; VMA: $t_{(25)} = 8.52, p < 0.01$; WS: $t_{(25)} = 2.18, p < 0.05$]. In contrast, the performance of the WS group on the SMI task was not significantly better than chance [$t_{(25)} = 0.61, p = 0.54$]. Performance in the VMA [$t_{(25)} = 7.02, p < 0.01$] and CA [$t_{(25)} = 33.7, p < 0.01$] groups on the SMI task was significantly better than chance.

Regarding the relationship between ToM task performance and the number of correct responses (**Figure 7**), we found a significant main effects of Group [$F_{(1, 48)} = 12.5, p < 0.01$], ToM performance [$F_{(1, 48)} = 34.6, p < 0.01$], and Task [$F_{(1, 48)} = 8.06, p < 0.01$]. No other interactions were significant [all $F_s < 2.9, p_s > 0.09$]. This suggests that the VMA group performed significantly better than the WS group, and that the members of the ToM pass group performed significantly better than the members of the ToM fail group. Moreover, SM task performance was significantly greater than SMI task performance.

The results of the developmental trajectory analysis indicated significant positive correlations between verbal mental age and correct performance in the VMA group for both the SM ($r = 0.69, p < 0.01$) and SMI ($r = 0.62, p < 0.01$) tasks. No significant effects were observed in individuals with WS (SM task: $r = 0.26, p = 0.19$; SMI task: $r = 0.15, p = 0.46$) (**Figure 8**). With respect to egocentric-bias errors, we observed a significant negative correlation with the SMI ($r = -0.44, p < 0.01$), but not





the SM task in the VMA group ($r = -0.31, p = 0.13$). We did not observe any significant correlations in the WS group (SM task: $r = -0.26, p = 0.19$; SMI task: $r = -0.27, p = 0.19$) (Figure 9).

We found similar correct responses and egocentric-bias error patterns between the MR and VPT2 tasks in Experiment 1, and between the SM and SMI tasks in Experiment 2. Thus, it is possible that MR and SM tasks engage similar mental processes. However, the results of the developmental trajectory analysis of the WS group indicated that, while MR performance significantly improved, SM performance did not. Therefore, we directly compared MR and SM task performance and found that the SM task performance was significantly worse than MR task performance in the WS group ($p < 0.001$), but not in the VMA ($p = 0.06$) and CA ($p = 0.89$) groups. We also directly compared VPT2 and SMI task performance and found that performance on the SMI task was significantly better than that on the VPT2 task in the VMA group ($p < 0.01$). This was not the case for the WS ($p = 0.20$) and the CA groups ($p = 0.17$). This data was affected by the fact that performance in the CA group for both tasks reached a ceiling level while performance in the WS group for both tasks was at chance level.

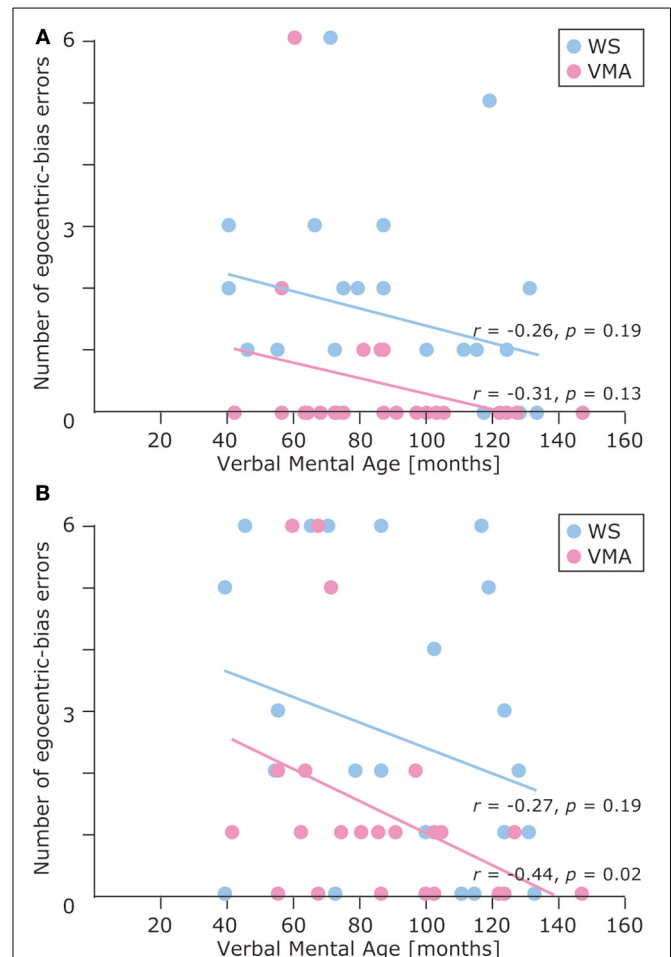


FIGURE 9 | Developmental trajectories for the number of egocentric-bias errors in (A) the SM task and (B) the SMI task for two groups [blue: Williams syndrome (WS) group; pink: verbal mental age-matched (VMA) group].

DISCUSSION

To the best of our knowledge, the current study is the first to investigate both MR and VPT2 task performance in individuals with WS, while considering developmental changes and the potential mechanisms that lead individuals with WS to exhibit impaired performance on the VPT2 task.

In Experiment 1, we found that people with WS performed poorly on MR and VPT2 tasks compared with normal controls. In terms of developmental trajectory, we found that in people with WS, MR task performance improved significantly with development, while VPT2 task performance did not. In Experiment 2, we manipulated the physical location of participants to investigate the source of difficulties that people with WS experience when completing VPT2 tasks. We introduced two experimental conditions: a self-motion task and a self-motion-imagery task. We found that both SM and SMI task performance was lower in the WS group than in control individuals. Moreover, task performance in the WS group did not improve with development, in contrast with the results of the control group.

Our findings can be summarized in three main points. First, the mental processes involved in the MR and VPT2 tasks were distinct, while the requirements of the VPT2 task were related to performance on the ToM tasks, as previously reported. Second, while processes related to MR tasks tend to develop slowly, processes related to VPT2 tasks seem to be impaired in individuals with WS (Experiment 1). Third, the poor VPT2 task performance previously observed in people with WS appears to be due to difficulty transitioning between the participant's perspective and a third-person perspective, and may also involve defective mental simulation of one's own body motion (Experiment 2).

Concordant with previous studies that investigated MR task performance in individuals with WS (Farran et al., 2001; Stinton et al., 2008), MR task performance was poor in people with WS compared with control individuals. As in a previous study that used a geometric figure with various orientations (Stinton et al., 2008), we found that performance in the VMA group was better than that in the WS group. However, contrary to the findings of Stinton et al. (2008), our results indicated that MR task performance in individuals with WS was significantly above chance. This discrepancy may be due to the fact that Stinton et al. (2008) used geometric shapes, which may have been less familiar to participants, while we used more familiar objects, such as toy animals, dolls, and cars. This discrepancy in familiarity may be related to differences in the amount of attention that the participants gave to the objects. As Hamilton et al. (2009) pointed out, the current task was relatively easy; that is, it consisted simply of pointing to one of four pictures. This minimized the need for verbal ability (Huttenlocher and Presson, 1973). Thus, the current task might have required a cognitive load that was lower than that of the task used by Stinton et al. (2008). This may have resulted in more attention being directed at the target objects, leading to better performance compared with the previous findings.

The correct responses in the VPT2 task were not significantly better than chance for the WS group, but were significantly better than chance in the VMA group. We adopted an experimental paradigm used by Hamilton et al. (2009), and so it is not surprising that performance in the VPT2 task in the VMA group was similar to their findings from children aged 6–10 years, whose performance was significantly above chance (Hamilton et al., 2009). Studies that have used a more complex VPT2 task or an appearance-reality task have reported that children do not reliably perform well until approximately 5–6 years of age (Flavell et al., 1986). Thus, VPT2 task performance seems to be task-dependent.

As Hamilton et al. (2009) noted, “the relationship between VPT2 and mentalizing supports the idea that the VPT2 should be considered a mentalizing task.” Further analysis in our study revealed that VPT2 task performance reflected ToM task performance in the VMA group, but not in the WS group. Additionally, VPT2 task performance in VMA children who passed the ToM tasks was significantly better than that in VMA children who failed the tasks. However, this difference was not observed for the MR task. Contrary to the results from the VMA group, we found a significant difference on the MR task, but not on the VPT2 task, in the WS group. This may be due to the overall low performance of WS participants on the VPT2 task. As a result, no significant effects were observed, in contrast with the findings

from the MR task. Moreover, as we did not find a clear interaction between ToM task performance and SM/SMI task performance in Experiment 2, it appears that neither task is sensitive to ToM task performance.

Therefore, our findings indicate that mentalizing ability might be impaired among some individuals with WS. This interpretation supports the view that socio-cognitive impairments are a component of WS (Tager-Flusberg and Sullivan, 2000). It should be noted that we found only two participants in the WS group who received a nearly perfect score (5 points) in the VPT2 task (**Figure 3**) and successfully completed both the location change and unexpected contents tasks. Concordant with this view, Porter et al. (2008) reported a specific deficit in social understanding within one of two WS subgroups using a non-verbal version of the ToM task. This deficit was observed even when the effects of mental or chronological age were controlled.

The developmental trajectory approach (Thomas et al., 2009) revealed differential developmental differences between the MR and VPT2 tasks in the VMA and WS groups. Whereas task success in the VMA group significantly improved with development in both tasks, in the WS group, development only improved MR task performance. Because both tasks were closely matched in terms of task difficulty (Hamilton et al., 2009), these findings suggest distinct mental processes. In the WS group, the processes related to the MR task appear to develop slowly while those related to the VPT2 task remain impaired regardless of development.

Recent neuroimaging reports suggest that differential brain regions are activated during MR and VPT2 tasks. For instance, the right inferior parietal sulcus is involved in a MR task (Harris et al., 2000; Podzebenko et al., 2002; Harris and Miniussi, 2003; Zacks et al., 2003a; Zacks, 2008) and the TPJ region plays an important role in completing a VPT2 task (Zacks et al., 2003b; Samson et al., 2004, 2005; Aichhorn et al., 2006; Santiesteban et al., 2012).

Considering the possibility of an abnormal dorsal stream in individuals with WS (Atkinson et al., 1997) in addition to the neuroimaging findings outlined above, it is plausible that the delayed development in MR task performance observed in our WS group may be associated with an atypical brain structure or atypical activation in dorsal brain regions. In line with this possibility, several studies have shown the existence of several atypical cortical structures in people with WS, such as reduced gray matter density in the superior parietal regions (Reiss et al., 2004; Eckert et al., 2005), including the intraparietal sulcus (Meyer-Lindenberg et al., 2004), bilateral reductions in sulcus depth in the intraparietal/occipitotemporal sulcus (Kippenhan et al., 2005), and prominent folding abnormalities in the dorsal parietal cortex (Van Essen et al., 2006). Atypical fractional anisotropy in the right superior longitudinal fasciculus, which is associated with deficits in visuospatial construction, has also been reported in WS individuals (Hoeft et al., 2007).

Although reduced activation has been reported in the inferior parietal cortices (Mobbs et al., 2007), there is little evidence of cortical abnormalities in the TPJ region in individuals with WS (Eckert et al., 2005). Therefore, the observed impaired VPT2 task performance of individuals with WS may be due to cortical abnormalities in other regions. A recent study showed that differential cortical regions, such as the right inferior frontal

gyrus and the dorsomedial prefrontal cortex, are involved in spatial tasks concerning the location of the self (Mazzarella et al., 2013). Furthermore, as several studies suggest that spatial perspective taking is an embodied cognitive process (May, 2004; Zacks and Michelon, 2005; Keehner et al., 2006; Kessler and Thomson, 2010), it is possible that impaired VPT2 task performance is related to the defective mental simulation of one's own body motion. Supporting this view, the results of Experiment 2 clearly indicate that SMI task performance in people with WS is significantly worse than that of normal controls. This suggests that people with WS experience difficulty updating the mental representation of their own perspective as it relates to the imagery of their bodily motion. Furthermore, our direct comparison between performance in the MR and SM tasks revealed that individuals with WS also have difficulty updating the mental representation of their own perspective as it relates to their physical bodily motion. Concordant with our findings, Nardini et al. (2008) investigated developmental changes for both body- and environmental-based reference frames in individuals with WS. They found no developmental improvement in the participant-move (body-based frame of reference) condition, but did find developmental changes in the array-move (environment-based frame of reference) condition. Considering these findings, the difficulty in VPT2 task performance observed in people with WS might be due to impaired simulation of the motion of one's own body. As outlined above, neuroimaging literature has indicated that the left posterior parietal cortex (Creem et al., 2001) or supplementary motor areas (Wraga et al., 2005), insula, and hippocampus (Lambrey et al., 2012) are involved in imagined rotations of one's self. Further studies are required to address these points and explore the cognitive and neural mechanisms underlying the task of adopting the viewpoint of another person, as well as the simulation of movement of one's own body.

In addition to correct responses (Hamilton et al., 2009), we analyzed patterns of error responses and found that egocentric-bias errors were significant in both the VPT2 and SMI tasks compared with the MR and SM. We observed significant reductions in egocentric-bias errors with subsequent development in the VMA group, but not in the WS group. This finding seems to be concordant with initial observations in the literature, which suggest that children aged 4–6 years typically report their own perspective (Piaget and Inhelder, 1956). We speculate that the consistent egocentric-bias error found in the WS group might reflect executive dysfunction (Jawaid et al., 2012) because previous behavioral studies have reported a close relationship between executive function ability and the theory of mind (Frye et al., 1995; Hughes, 1998; Perner and Lang, 2000; Carlson and Moses, 2001; Perner et al., 2002; Kloo and Perner, 2003; Carlson et al., 2004; Sabbagh et al., 2006).

In conclusion, our findings can be summarized in three points. First, we found that VPT2 task performance was lower than MR task performance in individuals with WS, and both performance scores were lower than those of the control groups. Second, we observed delayed developmental improvement in MR task performance and consistently impaired VPT2 task performance, irrespective of development, in individuals with WS.

Third, the findings of our second experiment indicate that difficulties faced by people with WS in terms of VPT2 task performance (Experiment 1) may be due to defective mental simulation of the motion of one's own body.

ACKNOWLEDGMENTS

We are grateful to the Williams Syndrome Association (Elfin Chubu, Nagoya) for their support of this research, and we thank all the young and adult participants, as well as their caregivers, for their participation. This work was supported by JSPS KAKENHI Grant Number 23830127, a Grant from the Daiko Foundation to Masahiro Hirai, and by JSPS KAKENHI Grant Number 23119733 to Miho Nakamura.

REFERENCES

- Aichhorn, M., Perner, J., Kronbichler, M., Staffen, W., and Ladurner, G. (2006). Do visual perspective tasks need theory of mind? *Neuroimage* 30, 1059–1068. doi: 10.1016/j.neuroimage.2005.10.026
- Apperly, I. A., Samson, D., Chiavarino, C., and Humphreys, G. W. (2004). Frontal and temporo-parietal lobe contributions to theory of mind: neuropsychological evidence from a false-belief task with reduced language and executive demands. *J. Cogn. Neurosci.* 16, 1773–1784. doi: 10.1162/0898929042947928
- Atkinson, J., Braddick, O., Anker, S., Curran, W., Andrew, R., Wattam-Bell, J., et al. (2003). Neurobiological models of visuospatial cognition in children with Williams syndrome: measures of dorsal-stream and frontal function. *Dev. Neuropsychol.* 23, 139–172. doi: 10.1080/87565641.2003.9651890
- Atkinson, J., Braddick, O., Rose, F. E., Searcy, Y. M., Wattam-Bell, J., and Bellugi, U. (2006). Dorsal-stream motion processing deficits persist into adulthood in Williams syndrome. *Neuropsychologia* 44, 828–833. doi: 10.1016/j.neuropsychologia.2005.08.002
- Atkinson, J., King, J., Braddick, O., Nokes, L., Anker, S., and Braddick, F. (1997). A specific deficit of dorsal stream function in Williams' syndrome. *Neuroreport* 8, 1919–1922. doi: 10.1097/00001756-199705260-00025
- Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition* 21, 37–46. doi: 10.1016/0010-0277(85)90022-8
- Bellugi, U., Bihle, A., Jernigan, T., Trauner, D., and Doherty, S. (1990). Neuropsychological, neurological, and neuroanatomical profile of Williams syndrome. *Am. J. Med. Genet. Suppl.* 6, 115–125.
- Bellugi, U., Lichtenberger, L., Jones, W., Lai, Z., and St George, M. (2000). I. The neurocognitive profile of Williams Syndrome: a complex pattern of strengths and weaknesses. *J. Cogn. Neurosci.* 12(Suppl. 1), 7–29. doi: 10.1162/089892900561959
- Bellugi, U., Sabo, H., and Vaid, J. (1988). "Spatial deficits in children with Williams syndrome," in *Spatial Cognition: Brain Bases and Development*, eds J. Stiles-Davis, U. Kritchavsky and U. Bellugi (Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.), 273–297.
- Brook, J. (2007). Language abilities in Williams syndrome: a critical review. *Dev. Psychopathol.* 19, 97–127. doi: 10.1017/S095457940707006X
- Carlson, S. M., Mandell, D. J., and Williams, L. (2004). Executive function and theory of mind: stability and prediction from ages 2–3. *Dev. Psychol.* 40, 1105–1122. doi: 10.1037/0012-1649.40.6.1105
- Carlson, S. M., and Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Dev.* 72, 1032–1053. doi: 10.1111/1467-8624.00333
- Creem, S. H., Downs, T. H., Wraga, M., Harrington, G. S., Proffitt, D. R., and Downs, J. H. 3rd. (2001). An fMRI study of imagined self-rotation. *Cogn. Affect. Behav. Neurosci.* 1, 239–249. doi: 10.3758/CABN.1.3.239
- Eckert, M. A., Hu, D., Eliez, S., Bellugi, U., Galaburda, A., Korenberg, J., et al. (2005). Evidence for superior parietal impairment in Williams syndrome. *Neurology* 64, 152–153. doi: 10.1212/01.WNL.0000148598.63153.8A
- Farran, E. K., and Jarrold, C. (2003). Visuospatial cognition in Williams syndrome: reviewing and accounting for the strengths and weaknesses in performance. *Dev. Neuropsychol.* 23, 173–200. doi: 10.1080/87565641.2003.9651891
- Farran, E. K., and Jarrold, C. (2004). Exploring block construction and mental imagery: evidence of atypical orientation discrimination in Williams syndrome. *Vis. Cogn.* 11, 1019–1039. doi: 10.1080/13506280444000058b

- Farran, E. K., and Jarrold, C. (2005). Evidence for unusual spatial location coding in Williams syndrome: an explanation for the local bias in visuo-spatial construction tasks? *Brain Cogn.* 59, 159–172. doi: 10.1016/j.bandc.2005.05.011
- Farran, E. K., Jarrold, C., and Gathercole, S. E. (2001). Block design performance in the Williams syndrome phenotype: a problem with mental imagery? *J. Child Psychol. Psychiatry* 42, 719–728. doi: 10.1111/1469-7610.00768
- Flavell, J. H. (1999). Cognitive development: children's knowledge about the mind. *Annu. Rev. Psychol.* 50, 21–45. doi: 10.1146/annurev.psych.50.1.21
- Flavell, J. H., Everett, B. A., Croft, K., and Flavell, E. R. (1984). Young children's knowledge about visual perception: Further evidence for the Level 1–Level 2 distinction. *Dev. Psychol.* 17, 99–103. doi: 10.1037/0012-1649.17.1.99
- Flavell, J. H., Green, F. L., and Flavell, E. R. (1986). Development of knowledge about the appearance-reality distinction. *Monogr. Soc. Res. Child Dev.* 51, i–v, 1–87. doi: 10.2307/1165866
- Frigerio, E., Burt, D. M., Gagliardi, C., Cioffi, G., Martelli, S., Perrett, D. I., et al. (2006). Is everybody always my friend? Perception of approachability in Williams syndrome. *Neuropsychologia* 44, 254–259. doi: 10.1016/j.neuropsychologia.2005.05.008
- Frye, D., Zalazo, P. D., and Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cogn. Dev.* 10, 483–527. doi: 10.1016/0885-2014(95)90024-1
- Gagliardi, C., Frigerio, E., Burt, D. M., Cazzaniga, I., Perrett, D. I., and Borgatti, R. (2003). Facial expression recognition in Williams syndrome. *Neuropsychologia* 41, 733–738. doi: 10.1016/S0028-3932(02)00178-1
- Hamilton, A. F., Brindley, R., and Frith, U. (2009). Visual perspective taking impairment in children with autistic spectrum disorder. *Cognition* 113, 37–44. doi: 10.1016/j.cognition.2009.07.007
- Harris, I. M., Egan, G. F., Sonkila, C., Tochon-Danguy, H. J., Paxinos, G., and Watson, J. D. (2000). Selective right parietal lobe activation during mental rotation: a parametric PET study. *Brain* 123(Pt 1), 65–73. doi: 10.1093/brain/123.1.65
- Harris, I. M., and Miniussi, C. (2003). Parietal lobe contribution to mental rotation demonstrated with rTMS. *J. Cogn. Neurosci.* 15, 315–323. doi: 10.1162/089892903321593054
- Hirai, M., Nakamura, M., Kaneoke, Y., and Kakigi, R. (2009). Intact point-light walker processing in Williams syndrome: a magnetoencephalography study. *Neuroreport* 20, 267–272. doi: 10.1097/WNR.0b013e3283283000de
- Hoef, F., Barnea-Goraly, N., Haas, B. W., Golarai, G., Ng, D., Mills, D., et al. (2007). More is not always better: increased fractional anisotropy of superior longitudinal fasciculus associated with poor visuospatial abilities in Williams syndrome. *J. Neurosci.* 27, 11960–11965. doi: 10.1523/JNEUROSCI.3591-07.2007
- Hoffman, J. E., Landau, B., and Pagani, B. (2003). Spatial breakdown in spatial construction: evidence from eye fixations in children with Williams syndrome. *Cogn. Psychol.* 46, 260–301. doi: 10.1016/S0010-0285(02)00518-2
- Hogrefe, G., Wimmer, H., and Perner, J. (1986). Ignorance vs. false belief: a developmental lag in attribution of epistemic states. *Child Dev.* 57, 567–582. doi: 10.2307/1130337
- Hughes, C. (1998). Executive function in preschoolers: links with theory of mind and verbal ability. *Br. J. Dev. Psychol.* 16, 233–253. doi: 10.1111/j.2044-835X.1998.tb00921.x
- Huttenlocher, J., and Presson, C. C. (1973). Mental rotation and the perspective problem. *Cogn. Psychol.* 4, 277–299. doi: 10.1016/0010-0285(73)90015-7
- Jarvinen, A., Korenberg, J. R., and Bellugi, U. (2013). The social phenotype of Williams syndrome. *Curr. Opin. Neurobiol.* 23, 414–422. doi: 10.1016/j.conb.2012.12.006
- Jawaid, A., Riby, D. M., Owens, J., White, S. W., Tarar, T., and Schulz, P. E. (2012). Too withdrawn or too friendly: considering social vulnerability in two neuro-developmental disorders. *J. Intellect. Disabil. Res.* 56, 335–350. doi: 10.1111/j.1365-2788.2011.01452.x
- Jones, W., Bellugi, U., Lai, Z., Chiles, M., Reilly, J., Lincoln, A., et al. (2000). “Hypersociability: the social and affective phenotype of Williams syndrome,” in *Journey from Cognition to Brain to Gene*, ed M. S. George (London: The MIT Press.), 43–71.
- Jordan, H., Reiss, J. E., Hoffman, J. E., and Landau, B. (2002). Intact perception of biological motion in the face of profound spatial deficits: Williams syndrome. *Psychol. Sci.* 13, 162–167. doi: 10.1111/1467-9280.00429
- Karmiloff-Smith, A., Grant, J., Berthoud, I., Davies, M., Howlin, P., and Udwin, O. (1997). Language and Williams syndrome: how intact is “intact”? *Child Dev.* 68, 246–262. doi: 10.2307/1131848
- Karmiloff-Smith, A., Klima, E., Bellugi, U., Grant, J., and Baron-Cohen, S. (1995). Is There a social module? language, face processing, and theory of mind in individuals with Williams syndrome. *J. Cogn. Neurosci.* 7, 196–208. doi: 10.1162/jocn.1995.7.2.196
- Keehner, M., Guerin, S. A., Miller, M. B., Turk, D. J., and Hegarty, M. (2006). Modulation of neural activity by angle of rotation during imagined spatial transformations. *Neuroimage* 33, 391–398. doi: 10.1016/j.neuroimage.2006.06.043
- Kessler, K., and Thomson, L. A. (2010). The embodied nature of spatial perspective taking: embodied transformation vs. sensorimotor interference. *Cognition* 114, 72–88. doi: 10.1016/j.cognition.2009.08.015
- Kippenhan, J. S., Olsen, R. K., Mervis, C. B., Morris, C. A., Kohn, P., Meyer-Lindenberg, A., et al. (2005). Genetic contributions to human gyrification: sulcal morphometry in Williams syndrome. *J. Neurosci.* 25, 7840–7846. doi: 10.1523/JNEUROSCI.1722-05.2005
- Kloo, D., and Perner, J. (2003). Training transfer between card sorting and false belief understanding: helping children apply conflicting descriptions. *Child Dev.* 74, 1823–1839. doi: 10.1046/j.1467-8624.2003.00640.x
- Lambrey, S., Doeller, C., Berthoz, A., and Burgess, N. (2012). Imagining being somewhere else: neural basis of changing perspective in space. *Cereb. Cortex* 22, 166–174. doi: 10.1093/cercor/bhr101
- Martens, M. A., Wilson, S. J., and Reutens, D. C. (2008). Research Review: Williams syndrome: a critical review of the cognitive, behavioral, and neuroanatomical phenotype. *J. Child Psychol. Psychiatry* 49, 576–608. doi: 10.1111/j.1469-7610.2008.01887.x
- May, M. (2004). Imaginal perspective switches in remembered environments: transformation vs. interference accounts. *Cogn. Psychol.* 48, 163–206. doi: 10.1016/S0010-0285(03)00127-0
- Mazzarella, E., Ramsey, R., Conson, M., and Hamilton, A. (2013). Brain systems for visual perspective taking and action perception. *Soc. Neurosci.* 8, 248–267. doi: 10.1080/17470919.2012.761160
- Meyer-Lindenberg, A., Kohn, P., Mervis, C. B., Kippenhan, J. S., Olsen, R. K., Morris, C. A., et al. (2004). Neural basis of genetically determined visuospatial construction deficit in Williams syndrome. *Neuron* 43, 623–631. doi: 10.1016/j.neuron.2004.08.014
- Meyer-Lindenberg, A., Mervis, C. B., and Berman, K. F. (2006). Neural mechanisms in Williams syndrome: a unique window to genetic influences on cognition and behavior. *Nat. Rev. Neurosci.* 7, 380–393. doi: 10.1038/nrn1906
- Mobbs, D., Eckert, M. A., Menon, V., Mills, D., Korenberg, J., Galaburda, A. M., et al. (2007). Reduced parietal and visual cortical activation during global processing in Williams syndrome. *Dev. Med. Child Neurol.* 49, 433–438. doi: 10.1111/j.1469-8749.2007.00433.x
- Moll, H., and Meltzoff, A. N. (2011). How does it look? Level 2 perspective-taking at 36 months of age. *Child Dev.* 82, 661–673. doi: 10.1111/j.1467-8624.2010.01571.x
- Moll, H., and Tomasello, M. (2006). Level 1 perspective-taking at 24 months of age. *Br. J. Dev. Psychol.* 24, 603–613. doi: 10.1348/026151005X55370
- Nakamura, M., Watanabe, K., Matsumoto, A., Yamanaka, T., Kumagai, T., Miyazaki, S., et al. (2001). Williams syndrome and deficiency in visuospatial recognition. *Dev. Med. Child Neurol.* 43, 617–621. doi: 10.1017/S0012162201001128
- Nardini, M., Atkinson, J., Braddick, O., and Burgess, N. (2008). Developmental trajectories for spatial frames of reference in Williams syndrome. *Dev. Sci.* 11, 583–595. doi: 10.1111/j.1467-7687.2007.00662.x
- Perner, J., and Lang, B. (2000). “Theory of mind and executive function: is there a developmental relationship?” in *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience*, eds S. Baron-Cohen, H. Tager-Flusberg and D. Cohen (Oxford: Oxford University Press), 150–181.
- Perner, J., Lang, B., and Kloo, D. (2002). Theory of mind and self-control: more than a common problem of inhibition. *Child Dev.* 73, 752–767. doi: 10.1111/1467-8624.00436
- Pezzini, G., Vicari, S., Volterra, V., Milani, L., and Ossella, M. T. (1999). Children with Williams syndrome: Is there a single neuropsychological profile? *Dev. Neuropsychol.* 15, 141–155. doi: 10.1080/87565649909540742
- Piaget, J., and Inhelder, B. (1956). *The Child's Conception of Space*. London: Routledge & Kegan Paul.

- Plesa-Skwerer, D., Faja, S., Schofield, C., Verbalis, A., and Tager-Flusberg, H. (2006). Perceiving facial and vocal expressions of emotion in individuals with Williams syndrome. *Am. J. Ment. Retard.* 111, 15–26. doi: 10.1352/0895-8017(2006)111[15:PFAVEO]2.0.CO;2
- Podzebenko, K., Egan, G. F., and Watson, J. D. (2002). Widespread dorsal stream activation during a parametric mental rotation task, revealed with functional magnetic resonance imaging. *Neuroimage* 15, 547–558. doi: 10.1006/nimg.2001.0999
- Porter, M. A., and Coltheart, M. (2005). Cognitive heterogeneity in Williams syndrome. *Dev. Neuropsychol.* 27, 275–306. doi: 10.1207/s15326942dn2702_5
- Porter, M. A., Coltheart, M., and Langdon, R. (2007). The neuropsychological basis of hypersociability in Williams and Down syndrome. *Neuropsychologia* 45, 2839–2849. doi: 10.1016/j.neuropsychologia.2007.05.006
- Porter, M. A., Coltheart, M., and Langdon, R. (2008). Theory of mind in Williams syndrome assessed using a non-verbal task. *J. Autism Dev. Disord.* 38, 806–814. doi: 10.1007/s10803-007-0447-4
- Reiss, A. L., Eckert, M. A., Rose, F. E., Karchemskiy, A., Kesler, S., Chang, M., et al. (2004). An experiment of nature: brain anatomy parallels cognition and behavior in Williams syndrome. *J. Neurosci.* 24, 5009–5015. doi: 10.1523/JNEUROSCI.5272-03.2004
- Reiss, J. E., Hoffman, J. E., and Landau, B. (2005). Motion processing specialization in Williams syndrome. *Vision Res.* 45, 3379–3390. doi: 10.1016/j.visres.2005.05.011
- Riby, D. M., and Porter, M. A. (2010). Williams syndrome. *Adv. Child Dev. Behav.* 39, 163–209. doi: 10.1016/B978-0-12-374748-8.00005-6
- Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., and Lee, K. (2006). The development of executive functioning and theory of mind. A comparison of Chinese and U.S. preschoolers. *Psychol. Sci.* 17, 74–81. doi: 10.1111/j.1467-9280.2005.01667.x
- Samson, D., Apperly, I. A., Chiavarino, C., and Humphreys, G. W. (2004). Left temporoparietal junction is necessary for representing someone else's belief. *Nat. Neurosci.* 7, 499–500. doi: 10.1038/nn1223
- Samson, D., Apperly, I. A., and Humphreys, G. W. (2007). Error analyses reveal contrasting deficits in “theory of mind”: neuropsychological evidence from a 3-option false belief task. *Neuropsychologia* 45, 2561–2569. doi: 10.1016/j.neuropsychologia.2007.03.013
- Samson, D., Apperly, I. A., Kathirgamanathan, U., and Humphreys, G. W. (2005). Seeing it my way: a case of a selective deficit in inhibiting self-perspective. *Brain* 128, 1102–1111. doi: 10.1093/brain/awh464
- Santiesteban, I., Baniassy, M. J., Catmur, C., and Bird, G. (2012). Enhancing Social Ability by Stimulating Right Temporoparietal Junction. *Curr. Biol.* 23, 2274–2277. doi: 10.1016/j.cub.2012.10.018
- Saxe, R., and Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in “theory of mind.” *Neuroimage* 19, 1835–1842. doi: 10.1016/S1053-8119(03)00230-1
- Skwerer, D. P., Verbalis, A., Schofield, C., Faja, S., and Tager-Flusberg, H. (2006). Social-perceptual abilities in adolescents and adults with Williams syndrome. *Cogn. Neuropsychol.* 23, 338–349. doi: 10.1080/02643290542000076
- Stinton, C., Farran, E. K., and Courbois, Y. (2008). Mental rotation in Williams syndrome: an impaired ability. *Dev. Neuropsychol.* 33, 565–583. doi: 10.1080/87565640802254323
- Stromme, P., Bjornstad, P. G., and Ramstad, K. (2002). Prevalence estimation of Williams syndrome. *J. Child Neurol.* 17, 269–271. doi: 10.1177/088307380201700406
- Sullivan, K., and Tager-Flusberg, H. (1999). Second-Order Belief Attribution in Williams Syndrome: Intact or Impaired? *Am. J. Ment. Retard.* 104, 523–532. doi: 10.1352/0895-8017(1999)104<0523:SBATWS>2.0.CO;2
- Tager-Flusberg, H., Boshart, J., and Baron-Cohen, S. (1998). Reading the windows to the soul: evidence of domain-specific sparing in Williams syndrome. *J. Cogn. Neurosci.* 10, 631–639. doi: 10.1162/089892998563031
- Tager-Flusberg, H., and Sullivan, K. (2000). A componential view of theory of mind: evidence from Williams syndrome. *Cognition* 76, 59–90. doi: 10.1016/S0010-0277(00)00069-X
- Tager-Flusberg, H., Sullivan, K., and Boshart, J. (1997). Executive functions and performance on false belief tasks. *Dev. Neuropsychol.* 13, 487–493. doi: 10.1080/87565649709540689
- Thomas, M. S., Annaz, D., Ansari, D., Scerif, G., Jarrold, C., and Karmiloff-Smith, A. (2009). Using developmental trajectories to understand developmental disorders. *J. Speech Lang. Hear. Res.* 52, 336–358. doi: 10.1044/1092-4388(2009/07-0144)
- Ueno, K., Nagoshi, S., and Iinaga, K. (2008). *PVT-R Kaiga goi Hattatsu Kensa (Picture Vocabulary Test-Revised)*. Tokyo: Nihon Bunka Kagakusha.
- Van Der Fluit, F., Gaffrey, M. S., and Klein-Tasman, B. P. (2012). Social cognition in Williams syndrome: relations between performance on the social attribution task and cognitive and behavioral characteristics. *Front. Psychol.* 3:197. doi: 10.3389/fpsyg.2012.00197
- Van Essen, D. C., Dierker, D., Snyder, A. Z., Raichle, M. E., Reiss, A. L., and Korenberg, J. (2006). Symmetry of cortical folding abnormalities in Williams syndrome revealed by surface-based analyses. *J. Neurosci.* 26, 5470–5483. doi: 10.1523/JNEUROSCI.4154-05.2006
- Vicari, S., Bellucci, S., and Carlesimo, G. A. (2003). Visual and spatial working memory dissociation: evidence from Williams syndrome. *Dev. Med. Child Neurol.* 45, 269–273. doi: 10.1111/j.1469-8749.2003.tb00342.x
- Vicari, S., Bellucci, S., and Carlesimo, G. A. (2005). Visual and spatial long-term memory: differential pattern of impairments in Williams and Down syndromes. *Dev. Med. Child Neurol.* 47, 305–311. doi: 10.1017/S0012162205000599
- Wimmer, H., and Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13, 103–128. doi: 10.1016/0010-0277(83)90004-5
- Wraga, M., Shephard, J. M., Church, J. A., Inati, S., and Kosslyn, S. M. (2005). Imagined rotations of self vs. objects: an fMRI study. *Neuropsychologia* 43, 1351–1361. doi: 10.1016/j.neuropsychologia.2004.11.028
- Zacks, J. M. (2008). Neuroimaging studies of mental rotation: a meta-analysis and review. *J. Cogn. Neurosci.* 20, 1–19. doi: 10.1162/jocn.2008.20013
- Zacks, J. M., Gilliam, F., and Ojemann, J. G. (2003a). Selective disturbance of mental rotation by cortical stimulation. *Neuropsychologia* 41, 1659–1667. doi: 10.1016/S0028-3932(03)00099-X
- Zacks, J. M., Vettel, J. M., and Michelon, P. (2003b). Imagined viewer and object rotations dissociated with event-related fMRI. *J. Cogn. Neurosci.* 15, 1002–1018. doi: 10.1162/089892903770007399
- Zacks, J. M., and Michelon, P. (2005). Transformations of visuospatial images. *Behav. Cogn. Neurosci. Rev.* 4, 96–118. doi: 10.1177/1534582305281085

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 16 July 2013; accepted: 22 November 2013; published online: 11 December 2013.

Citation: Hirai M, Muramatsu Y, Mizuno S, Kurahashi N, Kurahashi H and Nakamura M (2013) Developmental changes in mental rotation ability and visual perspective-taking in children and adults with Williams syndrome. *Front. Hum. Neurosci.* 7:856. doi: 10.3389/fnhum.2013.00856

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2013 Hirai, Muramatsu, Mizuno, Kurahashi, Kurahashi and Nakamura. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A review of visual perspective taking in autism spectrum disorder

Amy Pearson^{*†}, Danielle Ropar and Antonia F. de C. Hamilton[†]

School of Psychology, University of Nottingham, Nottingham, UK

Edited by:

Klaus Kessler, University of Glasgow, UK

Reviewed by:

Tad Brunye, US Army NSRDEC and Tufts University, USA

Bert Timmermans, University of Aberdeen, UK

*Correspondence:

Amy Pearson, Science Laboratories, Psychology Department, Durham University, South Road, Durham, DH1 3LE, UK

e-mail: amy.pearson@durham.ac.uk

†Present address:

Amy Pearson, Science Laboratories, Psychology Department, Durham University, Durham, UK;
Antonia F. de C. Hamilton, University College London, London, UK

Impairments in social cognition are a key symptom of autism spectrum disorder (ASD). People with autism have great difficulty with understanding the beliefs and desires of other people. In recent years literature has begun to examine the link between impairments in social cognition and abilities which demand the use of spatial and social skills, such as visual perspective taking (VPT). Flavell (1977) defined two levels of perspective taking: VPT level 1 is the ability to understand that other people have a different line of sight to ourselves, whereas VPT level 2 is the understanding that two people viewing the same item from different points in space may see different things. So far, literature on whether either level of VPT is impaired or intact in autism is inconsistent. Here we review studies which have examined VPT levels 1 and 2 in people with autism with a focus on their methods. We conclude the review with an evaluation of the findings into VPT in autism and give recommendations for future research which may give a clearer insight into whether perspective taking is truly impaired in autism.

Keywords: visual perspective taking, autism spectrum disorder, spatial transformations, social cognition, spatial cognition, theory of mind

Visual perspective taking (VPT) is the ability to see the world from another person's perspective, taking into account what they see and how they see it (Flavell, 1977). In order to perform VPT successfully a person must draw upon both spatial and social information. The spatial information used in VPT includes the current position of both the viewer and the target and the position of objects in the environment in relation to the self and others (Zacks et al., 2003; Kessler and Thomson, 2009; Kessler and Wang, 2012). For instance, you are sitting at a table with a friend drinking tea, the sugar pot is on their left hand side and the teapot is oriented with the handle toward your friend. The social information used in VPT involves the simultaneous representation of two differing points of view, taking into account whether someone else can see an object, or how they see that object (Aichhorn et al., 2006). For example, your friend can see the handle of the teapot while you see the spout. By interpreting the spatial relationships between objects in a social framework it becomes possible to form a rich representation of differing viewpoints which are useful in a variety of social tasks.

Impairments in social skills are a key symptom of autism spectrum disorder (ASD) (Baron-Cohen, 1995; Happe, 1995; Frith and Frith, 2007; Frith, 2012; Senju, 2012). Research has shown that people with autism have particular difficulty with theory of mind (ToM) and representing differing beliefs (Baron-Cohen et al., 1985; Baron-Cohen, 1995; Happe, 1995; Baron-Cohen et al., 1997; Frith, 2001; Senju et al., 2009; Senju, 2012). Some theorists believe that ToM and VPT share common cognitive processes (Hamilton et al., 2009) as they both involve the simultaneous representation of two differing points of view (Aichhorn et al., 2006). If this is the case then we may expect that people with autism would be impaired at VPT as well as ToM. However, others have

suggested that VPT and ToM are completely separate constructs and that it is entirely possible to be impaired at one and not the other (Leslie, 1987). Studies of whether VPT is intact in autism have been inconsistent (Reed and Peterson, 1990; Tan and Harris, 1991; Yirmiya et al., 1994; Hamilton et al., 2009). The focus of this review will be to examine studies of VPT in autism, assessing evidence for the existence of impairment. It will also consider the relationship between VPT and ToM, as well as the contribution of spatial abilities in VPT. We hope to set out a clear distinction for testing different types of VPT in autism as well as recommendations for experimental paradigms which may help to answer the question of whether these abilities may be impaired.

VISUAL PERSPECTIVE TAKING

There are two different levels of VPT outlined in the literature (Flavell, 1977). VPT level one (VPT1) is the basic ability to judge what a person can and cannot see (i.e., whether an item is occluded from their line of sight). The development of VPT1 marks the period at which children begin to understand that other people may be able to see different things, for example, knowing that if a toy is behind a parent that they will not see it until they turn around. VPT1 has been measured using a variety of tasks which require children to identify whether an adult can see an item which may/may not be occluded (Masangkay et al., 1974; Flavell et al., 1981). VPT level two (VPT2) is the ability to understand that two different people viewing a scene or object simultaneously do not necessarily see objects in the same way (Flavell, 1977). Tasks measuring VPT2 require a participant to be able to say *how* someone else sees an object or scene, for example, if you are standing opposite another person looking at a car, they may see the back of the car and you may see the front.

The development of VPT skills occur in succession, with VPT1 developing first followed by VPT2 (Flavell, 1977). Currently, it is thought that VPT1 develops between the ages of 18–24 months in typical children (Flavell et al., 1981; Moll and Tomasello, 2004, 2006; Moll et al., 2007) and VPT2 later at around 4–5 years old (Gzesh and Surber, 1985). Recent advances in the field of ToM research have shown that by using more implicit measures which are less reliant on language (such as eye tracking) we can find evidence of ToM skills earlier in infancy (Southgate et al., 2007). It has also been suggested that VPT1 may be able to operate in a spontaneous and implicit fashion (Samson et al., 2005; Surtees et al., 2012). Studies of VPT to date have used only explicit measures in their methodology (i.e., asking a child to point to an item or verbally report where someone is looking). Thus, it is possible that if implicit measures similar to those of Southgate et al. (2007) were used to examine VPT we may find that it develops earlier than previously thought.

Recently, efforts have been made to provide a clear distinction between VPT levels 1 and 2, and there are several ways in which this division can be drawn. This includes reference to embodiment, implicit/explicit processing, and dyadic/triadic representations. Surtees et al. (2013) makes a distinction based on embodiment. He suggests that VPT1 tasks require only visual (line of sight) information and not an egocentric embodied transformation, while VPT2 tasks require greater spatial information processing including the full transformation of the participant's viewpoint to that of the target. A different distinction is based on implicit/explicit processing. Samson et al. (2005) suggest VPT1 can occur implicitly and spontaneously. She presented participants with images of a room in which there was a human avatar and colored disks on the walls. Participants were asked to judge how many disks they could see or how many the avatar could see. The number of disks visible to the participants and the avatar were not always the same (for example, sometimes the avatar could not see all of the disks), creating perspective congruent and perspective incongruent conditions. The authors found that typical adults' responses were slower and less accurate when the avatar's view was incongruent with their own, suggesting that they implicitly coded the avatar's visual perspective (implicit VPT1) even when not required to by the task. A third way to distinguish VPT1 and 2 focuses on the number of relationships that a participant must encode in order to perform. Warreyn et al. (2005) argue VPT1 is based upon the use of dyadic representations whereas VPT2 is reliant on triadic representations. Dyads involve a representation of the relationship between a person and an object independent of the self (i.e., Jim can see the cat). Dyadic representations appear to be based upon the use of eye gaze following and line of sight (Warreyn et al., 2005). Triadic representations, involve coding the relationship between the self, another and an object (i.e., I can see the cat's tail whereas Jim can see the cat's nose). It remains to be seen which of these three types of division between level 1 and level 2 VPT is more valuable in understanding the overall phenomenon of perspective taking.

The present review focuses on studies of VPT in autism, where these distinctions have seldom been made clear. Previous studies suggest that embodiment may be reduced in autism (Brunye et al., 2012; Kessler and Wang, 2012; Eigsti, 2013) which would imply

that VPT1 should be intact but VPT2 impaired. In contrast, studies pointing to abnormal implicit ToM (in the presence of normal explicit ToM) (Senju, 2012) would predict that VPT1 should be harder in autism than VPT2. However, this might only be the case when VPT1 and 2 are tested with appropriately implicit methods, which has rarely been the case. Finally, it has been suggested that dyadic representation is intact in autism while triadic representation is impaired (Leekam et al., 1997). This implies that VPT1 should be normal in autism while VPT2 might not be. We revisit the issue of how VPT performance in autism relates to the key cognitive differences between VPT1 and VPT2 in the discussion.

One of the issues in assessing VPT in autism is the variety of methodologies that have been used. It has been suggested that people with autism may find some tasks easier to perform than others (Langdon and Coltheart, 2001) making it difficult to assert whether a lack of impairment is a result of intact VPT skills or the task used. Studies of VPT can be categorized by the types of questions they use (Figure 1). Most often studies focus on questions about item appearance (*"turn it so I can see the ____"*) or location (*"which side of the person is the counter?"*), as well as viewer or object rotations (*"imagine yourself at the blue side of the table"* vs. *"turn it so that you can see the apple"*). Studies which examine VPT1 are most likely to ask questions about line of sight (*"can this person see an object?"*) rather than questions about the items appearance from different viewpoints, which is a level 2 VPT skill (Figure 1).

Evidence for intact/impaired VPT1 and VPT2 in autism has so far been inconsistent, with studies showing evidence for both (Hobson, 1984; Leslie and Frith, 1988; Tan and Harris, 1991; Yirmiya et al., 1994; Leekam et al., 1997; Warreyn et al., 2005; Hamilton et al., 2009). Here we will examine studies and the methods they have used, taking into account what they add to the study of VPT in autism.

INCLUSION CRITERIA

An exhaustive search of the literature on VPT in autism was conducted using PubMed, web of science and Google Scholar. The search terms entered were "autism"/"ASD" and "visual perspective taking"/"VPT." Thirteen papers were identified which appeared to fit these criteria. All 13 papers examining VPT in autism have been included in this review.

Though studies aim to examine either VPT1 or VPT2, many of the tasks that have been used to test VPT could be completed using either, i.e., some VPT2 tasks could be completed using a simple line of sight VPT1 strategy. Here we discuss all studies which have examined VPT in autism and evaluate whether they fall into the category of VPT1 or VPT2.

VPT IN AUTISM

VPT has often been examined using tasks which ask questions about item visibility (Moll and Tomasello, 2004). In these studies, the child is presented with an item which is either in view or occluded from an adult. The child has to respond to whether the adult can see the item. Explicit studies of item visibility in typically developing (TD) children have shown that they are able to respond accurately from around 2 years old (Moll and Tomasello, 2004, 2006). Hobson (1984) examined VPT in adolescents with

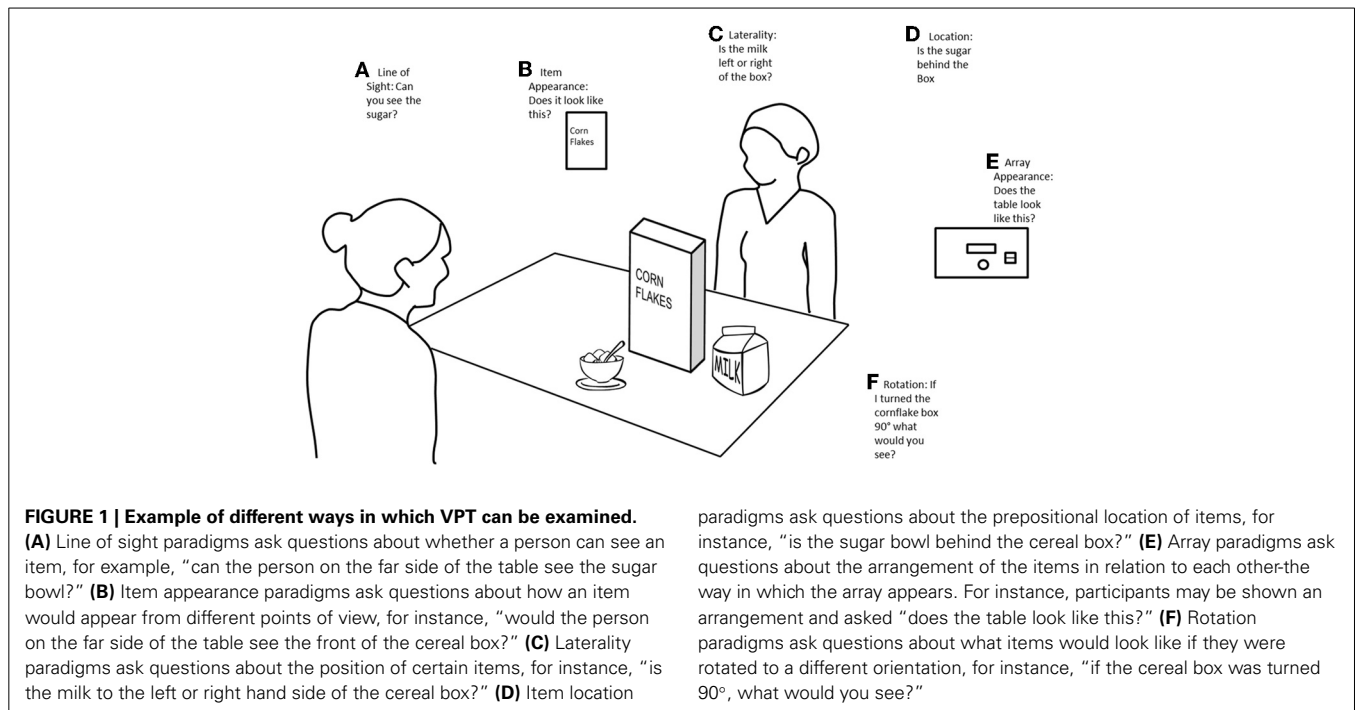


FIGURE 1 | Example of different ways in which VPT can be examined.

(A) Line of sight paradigms ask questions about whether a person can see an item, for example, “can the person on the far side of the table see the sugar bowl?” **(B)** Item appearance paradigms ask questions about how an item would appear from different points of view, for instance, “would the person on the far side of the table see the front of the cereal box?” **(C)** Laterality paradigms ask questions about the position of certain items, for instance, “is the milk to the left or right hand side of the cereal box?” **(D)** Item location

paradigms ask questions about the prepositional location of items, for instance, “is the sugar bowl behind the cereal box?” **(E)** Array paradigms ask questions about the arrangement of the items in relation to each other—the way in which the array appears. For instance, participants may be shown an arrangement and asked “does the table look like this?” **(F)** Rotation paradigms ask questions about what items would look like if they were rotated to a different orientation, for instance, “if the cereal box was turned 90°, what would you see?”

autism and VMA (verbal mental age) matched TD children using a “hide and seek” game paradigm, and found that the ability to perform VPT was intact. Participants were presented with a display which included hiding holes and two figures. The participant had to “hide” their figure from the other, indicating in which hole the figure would need to be placed so that they would not be seen. The participants with autism performed similarly to the ability matched TD children. These results have since been replicated using a similar hiding paradigm (Reed and Peterson, 1990; Tan and Harris, 1991; Reed, 2002). The findings from these studies suggest that children with ASD are able to understand the concept of “hiding” and what other people can see.

VPT has also been examined using line of sight paradigms. Leslie and Frith (1988) used a line of sight paradigm to investigate VPT in children with autism. Participants were presented with a scene in which a doll sat on one side of a cardboard screen and a counter was placed on the same side as the doll, or the opposite side. The child had to respond to whether the doll could see the counter. All of the autistic children were able to complete the task, suggesting that they had a basic understanding of what the doll could and could not see.

Baron-Cohen (1989) used a line of sight paradigm to examine VPT in children with autism and a group of TD children. Children were presented with a task in which an experimenter would orient their gaze or body toward one of six items surrounding the child and the child would have to identify which item the experimenter was looking to. The results showed that 92.5% of the children with ASD passed the task compared to 94.4% of TD children, suggesting VPT to be intact in the ASD group. Baron-Cohen’s study has been replicated since, though findings have not been quite as clear. Leekam et al. (1997) compared a group of ASD children to a group of VMA matched typical children on

Baron-Cohen’s perspective taking task. Though results showed no significant difference between the groups, there was a ceiling effect in the TD group (100%) whereas the ASD group scored on average much lower (66.6%). They also found that VMA was a significant predictor of performance, with those of lower VMA showing more difficulty with the task.

Warreyn et al. (2005) also conducted a replication of Baron-Cohen (1989) and found that young children with autism performed worse on the VPT task compared to age matched TD children. Similarly to Leekam et al. (1997), they found VMA to be a significant predictor of VPT ability. The authors suggested that VPT may develop later in children with autism and that they may be delayed compared to TD children.

All of the studies presented above (Hobson, 1984; Leslie and Frith, 1988; Baron-Cohen, 1989; Reed and Peterson, 1990; Tan and Harris, 1991; Leekam et al., 1997; Reed, 2002; Warreyn et al., 2005) can be classified as Level 1 VPT tasks on the basis that they examine line of sight.

VPT has also been examined using questions about item appearance. Mizuno et al. (2011) used a paradigm similar to that of Masangkay et al. (1974), in which adults with autism were shown a picture card with two sides. Participants were asked to identify which side they would see or another person would see in two different VPT conditions. In the first condition participants were asked a “what” question (“*what can I see?*” or “*what can Sarah see?*”) vs. “*What can you see?*”). In the second condition they were asked a “who” question (i.e., “*who will see the carrot?*”). Results showed that participants with autism were slower in the “what” condition than in the “who” condition. The authors argued that this was a result of difficulty switching between personal pronouns (“*what can you see?*” requires the participant to make the link between “you” being themselves), which people

with autism often find difficult (Lee et al., 1994). As the study uses a classic VPT1 paradigm, it seems most appropriate to label this a VPT1 task.

Hobson (1984) compared children with autism to a group of younger, VMA matched typical children. To examine VPT, Hobson used an object appearance task in which children had to identify the viewpoint of a third person (a doll). Typical and ASD children were presented with a cube which had a different color on each vertical face. The child was given a chance to familiarize themselves with the cube. Once familiarized the experimenter would place a doll (Fred) at one side of the cube and ask “*Fred sits here, which colour can he see?*” or “*place Fred so he can see the ____.*” The child was then given a second doll (Mary) and asked to “*put Mary so that Mary sees the same as Fred sees.*” Results showed that there was no significant effect of group, with the ASD children performing similarly to the typical children. Hobson did find a significant effect of verbal ability in the ASD group, with higher functioning ASD children performing better. This is consistent with the findings from Warreyn et al. (2005) and (Leekam et al., 1997), and suggests that verbal ability may be an important predictor of VPT. It is also worth noting that neither group performed at ceiling level in Hobson’s task meaning any group differences should be clear. As the task could be completed using a VPT1 strategy in which participants use line of sight to respond rather than performing a first person transformation it seems appropriate to define this as a level one VPT task.

Reed and Peterson (1990) also examined VPT in children with autism alongside ToM using an item appearance paradigm. Thirteen ASD children and 13 VMA matched TD children were tested on their ability to rotate a familiar item (a toy) so that the experimenter could see a distinct feature (i.e., “*turn it so that I can see the nose*”). Four different toys were presented and children had to score 100% across all four trials to pass. In contrast the cognitive perspective taking task required the children to perform the Sally-Anne ToM task (Baron-Cohen et al., 1985). The authors found that the children with autism performed similarly to the typical children in the VPT task, but worse in the cognitive perspective taking task. The authors concluded that it could not be the social aspect of ToM that participants with autism had difficulty with, as the VPT task was also social and that poor ToM may be a result of impaired abstract thinking. These findings suggested that VPT and mentalizing are dissociable abilities, with VPT tapping into a different process than ToM. However, the authors found a ceiling effect amongst both the typical and autistic children in the VPT task. This makes it possible that group differences may have been masked due to the task being particularly easy for both groups of participants. This task was classified as a VPT2 task by the authors on the basis that it meets criteria for two people viewing an object from different vantage points (Flavell et al., 1981). However, participants could also use a basic line of sight (VPT1) strategy (turning the item until the feature (i.e., nose) was in the line of sight of the viewer) to respond. The distinction between level one and two VPT are blurred in this task, and it may be more appropriate to label this a VPT1 task.

Tan and Harris (1991) examined VPT in children with autism using an item location task. Twenty children with autism and 20 VMA matched TD children were tested on their ability to

identify the view one of two soft dolls would have of a third object (i.e., *which object would John say was “in front?”*). The authors also measured the children’s ToM using a desire understanding task, presenting the children with scenarios in which someone was offered food that they did or did not like. Children had to respond to whether the person would be happy or unhappy with the offer. There was no significant effect of group on either task, with the autistic children performing similarly to the typical children on both VPT and desire understanding. As with Reed and Peterson’s task, Tan and Harris also found a ceiling effect across both groups of participants which may have masked any group differences. The authors concluded that a global social deficit in autism is unlikely, and that impairment may be related to process and task specific delays. As this task measures how two people seeing a given object may view it differently due to a change in orientation or location (i.e., for Mary, the pencil is in front of the block, whereas for John the pencil is behind the block) it can be considered a VPT2 task.

Yirmiya et al. (1994) examined VPT in children with ASD using an object rotation paradigm in which children were presented with familiar item (toys) on a rotating table. The task required both object rotation and item appearance (“how would this look to me”). ASD children were compared to age and IQ matched TD children on their ability to turn a turntable containing 3 or 10 items so that it matched the point of view of the experimenter. Children were instructed to “turn it around so that you will see it from where you are in the same way that I see it from where I am” or “turn it around until you see it in the exact same way that I see it now from where I am standing.” They found that children with ASD showed a higher number of errors than the typical children. Errors were further categorized into two different types: incorrect (in which the answer was simply wrong) or egocentric (in which the child displayed the turntable with their own point of view). Children with autism were found to display more incorrect errors in the 10 item trials, and more egocentric errors in the 3 item trials. This suggests that the 10 item trials were more reliant on memory, as if both trial types were equated for difficulty you would expect to see similar types of errors across both. This task demands the calculation of two different viewpoints and is clearly a VPT2 task, but as the authors note it has heavy memory demands which may limit performance.

Hamilton et al. (2009) used a related paradigm to examine VPT, mental rotation and ToM ability in a group of ASD children compared to verbal ability matched TD children. Two further groups of TD children were also included in the study, a typical mid-age range group and a typical older group. For the VPT task children were presented with the toy on the turntable and asked to identify their own point of view on the answer sheet. The toy was then covered and a doll placed at another spot on the table. The child was asked to identify the view of the toy the doll would have when the pot was lifted. For the mental rotation task children were shown a toy on a turntable and asked to identify which picture on their answer sheet matched their view. The toy was then covered and rotated and the child asked to identify which view they would see when the pot was lifted. ToM was assessed using a battery of different ToM tasks, including diverse desires and the Sally-Anne task (Baron-Cohen et al., 1985). Results showed that

the children with ASD were significantly worse on the VPT trials compared to the typical children, but performed better on the mental rotation task. It was also found that VPT was significantly predicted by ToM score, suggesting mentalizing is important for perspective taking. The authors suggested that VPT relies on the same cognitive systems as ToM. This is the only study reviewed which includes both a social and non-social spatial task, as well as a measure of ToM. The task attempts to integrate different task demands (viewer and item rotation, item appearance questions) making it possible to start pinpointing specific difficulties with VPT. The use of a control spatial (non-social) task also allows the authors to make clear conclusions about which aspects of VPT that people with autism find difficult (social as opposed to the spatial). We suggest that as the task explicitly requires participants to say what one object would look like from two different points of view, with no line of sight information available (the target was covered with a pot), that this be classified as a VPT2 task.

Dawson and Fernald (1987) also examined VPT in children with autism using an object rotation paradigm in which children had to orient an item a certain way for the experimenter to see it. No control group was included in the study. Participants were presented with cards, blocks and various picture and asked to orient it “so the experimenter could see the face/tail etc . . .” None of the children scored at ceiling level on the task, and performance correlated with social skills, but without a control group it is hard to interpret this data.

David and colleagues examined VPT and ToM in high functioning adults with Asperger syndrome compared to age and IQ matched TD adults. Participants completed two tasks, one examined VPT and the other examined ToM. In the ToM task participants were presented with a virtual image of a person with one item either side of them. The person could be displaying one of three possible body, face and hand postures (positive, neutral, or negative) toward one of the objects. An example of a positive hand gesture would be pointing, whereas negative would be holding the hand out with the palm facing forwards (similar to a “stop” signal). The participant’s task was to identify which object the other person desired (mentalizing for other) or which they would desire themselves (mentalizing for self). In the VPT task the participant was presented with the same image of the person with two objects, one of which was elevated. The participant had to identify which object was elevated from their own point of view, or from that of the other person using a laterality judgment (i.e., the item on *my left* is higher). Measures of speed and accuracy were taken from each participant. In the ToM task results showed that the ASD participants were significantly slower and less accurate at identifying the correct answer when mentalizing for other. They were also trending toward slower mentalizing for self (as accuracy on this task was subjective accuracy could not be measured). There were no differences found between groups for speed or accuracy in the VPT task, for self or other. The authors acknowledged that the VPT task may have been too easy compared to the mentalizing task which may explain differences across tasks. One limitation is that this task does not require participants to take the visual perspective of the other, but only to judge what is on the left or right. Spatial-transformation tasks (Parsons, 1987; Zacks et al., 1999) requires participants to make laterality judgments

about an item in relation to another person, but it is not clear if these are the same as VPT tasks. Further research is needed into these paradigms in order to assess where they fall in relation to perspective taking.

Similarly, Zwickel et al. (2011) examined VPT and ToM in adults with autism and age and IQ matched TD adults using a laterality judgment paradigm. In the VPT task participants viewed videos of animated triangles (Castelli et al., 2002), and during the videos a dot appeared to the left or right of the triangle. Participants were asked simply “*was the dot on your left or right?*” On incongruent trials a dot on the participant’s left fell on the right of the triangle (or vice versa), while on congruent trials a dot on the participant’s left was also on the left of the triangle (or both on the right). Critically, this congruency only arises if the triangle is perceived as an animate active creature. Both typical and autistic participants showed a congruency effect in this task, demonstrating that they could spontaneously consider the left/right orientation of an animated shape. However, the autistic participants were less good at judging the mental states of the triangles in the same animations. This is consistent with the findings of David et al. (2010). Similarly, it is not clear if this task truly demands calculation of the *visual perspective* of another agent rather than just their orientation. More research is needed to explore the use of visuo-spatial perspective taking paradigms in autism.

EVALUATING VPT IN AUTISM

We have reviewed 13 studies of VPT in autism, and suggest that 7 of these assessed VPT1, 3 assessed VPT2 and 3 were unclear or assessed laterality (see **Table 1**). Of the 7 studies examining VPT1, 5 report no differences between typical and autistic participants while the other 2 find that participants with autism perform worse than typical participants. Of the 3 studies examining VPT2, 2 report group differences and the third does not.

There are several interesting issues arising from this review which can guide future research. One important problem is that the boundary between VPT levels one and two is not always clear. A task might be intended to assess VPT1, but participants might choose to use a VPT2 strategy. Or if a study designed to measure VPT2 could also be completed using line of sight, it is possible that people with autism could pass based on this information. This is particularly the case in studies which name the item which can be seen from a particular location [e.g., *place Fred so he can see the red side*, (Hobson, 1984)]. Here the child need only consider Fred’s line of sight to the red part of the cube, but some children might prefer to consider the relationship of the whole cube to the rest of the scene including the child’s own viewpoint. Thus, this task could be solved by a VPT1 or VPT2 strategy. To minimize this issue, we suggest that line of sight tasks seem to be the clearest way to assess VPT 1 (Leslie and Frith, 1988; Baron-Cohen, 1989; Leekam et al., 1997; Warreyn et al., 2005), whilst item appearance tasks appear to be the best way to assess VPT2 (Hamilton et al., 2009) see **Figures 1A,B**.

A related issue is the use of different strategies by different participants. However well an experimenter designs a task, it is always possible that participants could solve the puzzle in a different way. For example, many VPT tasks could potentially be solved

Table 1 | Summary of studies included in this review.

Study	Participants with ASD				Typical participants			Experimental tasks		
	Number	Mean age	Mental age		Number	Mean age	Mental age	IQ measure	Task	Level Result
Yirmiya et al. (1994)	18	12.5	12.3		14	12.0	12.5	WISC	Turn turntables to match my point of view	2 Sig effect of group (TD > AS)
Hamilton et al. (2009)	23	8.0	4.4		23	4.2	4.8	BPVS	Which bear will Susan see?	2 Sig effect of group (TD > AS) on VPT, AS better at MR
Reed and Peterson (1990)	13	12.0	7.1		13	7.1	–	WISC/WAIS	Turn the turn table so I see the	1 No sig difference between groups
Hobson (1984)	12	13.8			27	5.7	–	WISC	What can "... see?	1 No sig difference between groups
Tan and Harris (1991)	20	12.8	7.7		20	6.1	6.5	PPVT	Which item is in front of...	2 No sig difference between groups
David et al. (2010)	19	36.4	>30		15	31.2	>30	WAIS	Which object is elevated from your/their perspective?	lat No sig difference between groups
Zwicker et al. (2011)	19	37.0	>30		18	39.0	>30	WAIS	Which side of the triangle is the dot?	lat Sig diff on WRONG responses only
Warreyn et al. (2005)	20		4.3		20	5.6	4.7	RDLS	What am I looking at?	1 Sig effect of group (TD > AS)
Reed (2002)	25	12.0	8.2		25	6.8	7.11	PPVT	Can you hide Susie where no-one else can see	1 Sig effect of group (TD > AS)
Leekam et al. (1997)	12	11.8	5.4		12	5.8	–	TROG	Which toy am I looking at	1 No sig effect of group, but trending toward TD > AS
Baron-Cohen (1989)	20	11.1	5.5		27	4.5	–	BPVS	Which toy am I looking at	1 No sig difference between groups
Leslie and Frith (1988)	18	13.1	7.2		12	8.8	6.9	BPVS	Can the doll see the counter	1 No sig difference between groups
Dawson and Fernald (1987)	16	11.1	5.8		–	–	–	WISC and others	Turn it so I can see the ...	mix VPT sig related to social skills

with a purely spatial mental rotation (Zacks and Tversky, 2005). This approach is less efficient, but it is possible that different groups of participants prefer to use different strategies. One way to approach this issue is to consider the use of appropriate control tasks to assess other cognitive skills such as children's memory abilities (especially for complex displays), their language skills (for complex questions) and their abilities to perform spatial transformations. The comparison of an experimental task and a closely matched control task in the method of fine cuts (Frith and Happe, 1994) would allow for close examination of the cognitive components which distinguish the different levels of perspective taking. For example, Surtees et al. (2013) suggests that VPT2 requires an embodied spatial transformation while VPT1 does not. If this is the case, then VPT2 abilities should correlate with performance on other tasks requiring embodiment, but not to mental rotation tasks that do not involve bodies. If different groups of participants use different strategies to perform VPT tasks, this might also emerge in the relationship between their VPT skill and other cognitive skills.

Furthermore, this raises another important question concerning how the social and spatial elements of VPT2 fit together: Does intact VPT require spatial *and* social information, or could it be done using just one of these? If VPT2 can be completed using social *or* spatial information it makes sense that it can be unimpaired even in the face of significant ToM deficits, as participants' could rely on the use of spatial information to complete a task. Langdon and Coltheart (2001) suggested that tasks using questions about item location (i.e., Tan and Harris (1991)) were particularly open to completion via spatial cues making it possible for those with social difficulty to perform. However, if VPT2 requires the integration of both spatial and social information to be effective, then even good spatial ability would not completely compensate for poor social processing. Again, determining the strategies and cognitive mechanisms that different participants use to perform VPT tasks is critical here.

Another issue concerns the participant populations tested. The majority of studies presented in this review were conducted on children, and several on groups of children with impaired cognitive functioning. It is difficult to collect reaction time data from children, meaning that more subtle differences in VPT ability related to an inability to integrate social and spatial information may be missed. The two studies conducted with adults (David et al., 2010; Zwickel et al., 2011) did not find group differences but did not use typical VPT tasks. It is possible that [as found in ToM research (Ozonoff et al., 1991)], high functioning adults may be able to pass VPT. Whether this is due to a better understanding of the questions asked, or the development of an alternative strategy for completion of the tasks is unclear. Both of these suggestions warrant further research and careful consideration of the paradigms used to examine VPT.

There are also issues in the lack of consistency in matching groups. Though some of the studies have used rigorous matching techniques (Yirmiya et al., 1994; Hamilton et al., 2009; David et al., 2010), others took no measure of cognitive ability in their typical participants. Both Reed and Peterson (1990) and Hobson (1984) argue for evidence of unimpaired VPT2 performance in autism. However, they both compared groups of older ASD

children to younger typical children. This suggests that at the very least the participants with autism may be displaying a delay in the development of VPT (similar performance to younger children as opposed to an age matched group) and that it may be inappropriate to label their performance as unimpaired. By comparing ASD participants to both age and ability matched control participants, it becomes possible to make stronger claims as to whether performance on a task is normal, impaired or simply delayed. These findings present a strong case for using carefully chosen control groups in studies looking for evidence of impairment in a population such as autism.

FUTURE DIRECTIONS

Understanding the relationship between VPT and ToM is important. Both of these require the consideration that the other person has a different representation of the world to oneself, either a different visual representation or a different belief. Early studies suggested that VPT is intact in autism while ToM is impaired. This motivated the idea that it is easy to distinguish visual representations of self and other because VPT allows concrete feedback by physically moving to a different location (Leslie (1987)). In contrast, ToM requires more abstract representations which people with ASD find difficult. More recent data suggest that VPT2 and ToM are linked in typical children (Hamilton et al., 2009), in those with specific language impairment (Farrant et al., 2006) and in the brain (Aichhorn et al., 2006). This implies that VPT2 and ToM may share similar underlying cognitive mechanisms. Certainly, many false belief tasks rely on the ability to distinguish what people have seen (Sally did not see Anne move the marble) which draws upon VPT. VPT has been found to activate the temporo-parietal junction, an area commonly found to be activated by ToM tasks (Aichhorn et al., 2006). It has been suggested that ToM may be driven by different mechanisms or strategies in people with autism compared to TD people (Tan and Harris, 1991). We believe it may also be worth considering that this could also be the case for VPT in people with autism. If both are being driven by different mechanisms it may explain why some studies have shown VPT to be unimpaired alongside impaired ToM (Tan and Harris, 1991) and vice versa (Hamilton et al., 2009). Further studies of the relationship between ToM and VPT would be very useful, as would studies examining the cognitive mechanisms which underlie each.

It is also worth considering how researchers can tease apart the specific contributions of social and spatial mechanism in VPT. Several VPT tasks have been used successfully in TD individuals which have allowed researchers to emphasize the spatial or social aspects. The use of these paradigms may provide us with useful information about perspective taking in autism. As described earlier, Samson et al. (2005) investigated the social components of VPT in an implicit perspective taking task. Results showed that participants could not ignore the perspective of an avatar, and made slower responses when the avatar could not see something which the participant could see. Another VPT task with strong social demands is Keysar et al. (2003) director task. In this task the participants stand behind a shelf holding several items while another person stands in front (the director) and gives instructions of which items to choose. Not all items are visible to the

director and so the participant must be able to take the directors perspective into account to avoid choosing items that they cannot see. The authors found that participants were not able to inhibit their own perspective when choosing items and often made incorrect responses. This task has been argued to have a strong ToM component as it relies on the ability to represent someone else's false belief (the director believes the "big jar" is the one they can see, but there is a bigger jar on view to the participant). Both of these tasks would provide interesting ways of measuring the social components of VPT.

Kessler and Thomson (2009) developed a task in which they were able to examine the underlying spatial components present in perspective taking (termed spatial perspective taking, or SPT). Participants were presented with images of a human avatar seated at a table with an item to either side of them (a flower and a gun). The position of the avatar at the table was rotated to be more or less congruent with the position of the participant (providing changes in the angular disparity between the avatar and viewer). Participants had to make laterality judgments in regards to the placements of the items from the avatars viewpoint. The authors found that the larger the angular disparity between then avatar and the viewer, the longer participants took to respond. This demonstrated the underlying spatial transformation that the participant completed in order to put themselves in the place of the avatar, highlighting the importance of spatial mechanisms in perspective taking. These findings show that in order to take a first person perspective, an embodied transformation (where the viewer transforms their body to match that of the avatar or target viewpoint) is often necessary. Mazarella et al. (2013) built upon these findings, investigating the neural underpinnings of spatial perspective taking with a similar experiment which examined SPT under fMRI. In this task participants were scanned whilst making egocentric (what item is on your left) vs. altercentric (which item is on their left) judgments about the placement of items on a table [using the same paradigm as Kessler and Thomson (2009)]. This was designed to tease apart the differences in transforming the self to a different position vs. transforming the self into someone else's position. The authors found that though both types of transformation show similar behavioral data patterns, there was a neural distinction between the areas engaged during egocentric and altercentric perspective taking. This suggests that multiple strategies may be used for putting the self in a different place. These tasks both provide clear mechanisms for teasing out the spatial components of VPT, as well as interesting avenues to explore in people with autism.

Recently, researchers have also begun to examine the link between autistic traits in TD individuals and how this affects perspective taking. As VPT is a sociocognitive ability which impacts on social interaction, it stands to reason that those with poorer social skills might also show poorer perspective taking ability. Three studies have examined this question. First, Kessler and Wang (2012) examined participants using the same task from Kessler and Thomson (2009). A measure of autistic traits in these participants was taken using the Autism Quotient (AQ, Baron-Cohen et al., 2001). The authors found that participants who scored higher on the AQ (Baron-Cohen et al., 2001) showed

more difficulty with performing egocentric transformations in a VPT2 task than low AQ scorers. However these high AQ participants also showed quicker response times. The results further suggest that using an embodied perspective slowed participant's responses at higher angular disparities as it took longer for the participant to transform their body to match that of the target. Participants with poorer embodiment skills did not slow their response times, most likely due to the use of a non-embodied transformation strategy. Using a very similar task, Brunye et al. (2012) also found that high AQ scorers had difficulty with using an egocentric reference frame in VPT2, though in this study these participants showed slower response times. This suggested that these high AQ participants were attempting to use an embodied strategy, but found it more difficult. Finally, Shelton et al. (2012) also found a link between spatial skills and autistic traits. In their study participants were presented with a three mountains (Piaget and Inhelder, 1956) like scene, in which an array of three buildings were visible to participants. A doll was placed facing the array and participants had to respond to which point of view the doll would see. Their study showed that participants with high AQ scores were less accurate than those with low AQ scores.

Together, all these suggest that autistic traits (as well as autism itself) can influence a participant's ability to perform VPT2. These studies are important for two reasons. Firstly they demonstrate that autistic traits (not just a diagnosis of autism) impact on the ability to take another perspective. Secondly, they add weight to the argument that those who find it difficult to complete VPT2 using an embodied perspective may develop an alternative strategy. The findings from these studies provide a strong motivation for considering the types of participant samples used in VPT2 research and measuring traits which could affect performance alongside carefully designed paradigms and tasks.

CONCLUSIONS

From the evidence presented in this review, the majority of studies suggest that whilst VPT1 may be intact in people with autism, VPT2 is impaired. We suggest that this is a result of the cognitive mechanisms involved in the different levels of VPT, with VPT2 drawing on embodied spatial transformations and triadic representations (Surtees et al., 2013) more than VPT1. Future studies should carefully consider the cognitive differences between VPT1 and VPT2. Furthermore, there is evidence to suggest that the ability to perform egocentric transformations (a process which can be seen as the first step in completing VPT2 (Yu and Zacks, 2010) could be impaired in autism, but may also be affected in people with high levels of autistic traits (Brunye et al., 2012; Kessler and Wang, 2012; Shelton et al., 2012). It is clear that more research is needed into the processes related to VPT2 in autism in order to clarify these suggestions. There is a strong case to be made for more inclusion of measures of general spatial ability in studies on VPT and the use of a "fine cuts" technique when designing studies. This will allow researchers to tease apart impairments in the spatial demands of a task vs. the social. The recommendations set out in this review provide a strong motivation for investigating VPT in autism and shed light on why findings so far are inconsistent.

REFERENCES

- Aichhorn, M., Perner, J., Kronbichler, M., Staffen, W., and Ladurner, G. (2006). Do visual perspective tasks need theory of mind. *Neuroimage* 30, 1059–1068. doi: 10.1016/j.neuroimage.2005.10.026
- Baron-Cohen, S. (1989). Perceptual role taking and protodeclarative pointing in autism. *Br. J. Dev. Psychol.* 7, 113–127. doi: 10.1111/j.2044-835X.1989.tb00793.x
- Baron-Cohen, S. (1995). *Mindblindness: an Essay on Autism and Theory of Mind*. Cambridge, MA: MIT Press/Bradford Books.
- Baron-Cohen, S., Jolliffe, T., Mortimore, C., and Robertson, M. (1997). Another advanced test of theory of mind: evidence from very high functioning adults with autism or asperger syndrome. *J. Child Psychol. Psychiatry* 38, 813–822. doi: 10.1111/j.1469-7610.1997.tb01599.x
- Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985). Does the autistic child have a “theory of mind.” *Cognition* 21, 37–46. doi: 10.1016/0010-0277(85)90022-8
- Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., and Clubley, E. (2001). The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J. Autism Dev. Disord.* 31, 5–17. doi: 10.1023/A:1005653411471
- Brunye, T. T., Ditman, T., Giles, G. E., Mahoney, C. R., Kessler, K., and Taylor, H. A. (2012). Gender and autistic personality traits predict perspective-taking ability in typical adults. *Pers. Individ. Differ.* 52, 84–88. doi: 10.1016/j.paid.2011.09.004
- Castelli, F., Frith, C., Happe, F., and Frith, U. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain* 125, 1839–1849. doi: 10.1093/brain/awf189
- David, N., Aumann, C., Bewernick, B., Santos, N., Lehnhardt, F.-G., and Vogeley, K. (2010). Investigation of mentalizing and visuospatial perspective taking for self and other in asperger syndrome. *J. Autism Dev. Disord.* 40, 290–299. doi: 10.1007/s10803-009-0867-4
- Dawson, G., and Fernald, M. (1987). Perspective-taking ability and its relationship to the social-behavior of autistic-children. *J. Autism Dev. Disord.* 17, 487–498. doi: 10.1007/BF01486695
- Eigsti, I. M. (2013). A review of embodiment in autism spectrum disorders. *Front. Psychol.* 4:224. doi: 10.3389/fpsyg.2013.00224
- Farrar, B. M., Fletcher, J., and Maybery, M. T. (2006). Specific language impairment, theory of mind, and visual perspective taking: evidence for simulation theory and the developmental role of language. *Child Dev.* 77, 1842–1853. doi: 10.1111/j.1467-8624.2006.00977.x
- Flavell, J. H. (1977). The development of knowledge about visual perception. *Nebr. Symp. Motiv.* 25, 43–76.
- Flavell, J. H., Everett, B. A., Croft, K., and Flavell, E. R. (1981). Young children’s knowledge about visual perception: further evidence for the Level 1–Level 2 distinction. *Dev. Psychol.* 17, 99–103. doi: 10.1037/0012-1649.17.1.99
- Frith, C. D., and Frith, U. (2007). Social cognition in humans. *Curr. Biol.* 17, R724–R732. doi: 10.1016/j.cub.2007.05.068
- Frith, U. (2001). Mind blindness and the brain in autism. *Neuron* 32, 969–979. doi: 10.1016/S0896-6273(01)00552-9
- Frith, U. (2012). Why we need cognitive explanations of autism. *Q. J. Exp. Psychol. (Hove)* 65, 2073–2092. doi: 10.1080/17470218.2012.697178
- Frith, U., and Happe, F. (1994). Autism – beyond theory of mind. *Cognition* 50, 115–132. doi: 10.1016/0010-0277(94)90024-8
- Gzesh, S. M., and Surber, C. F. (1985). Visual perspective-taking skills in children. *Child Dev.* 56, 1204–1213. doi: 10.2307/1130235
- Hamilton, A. F., Brindley, R., and Frith, U. (2009). Visual perspective taking impairment in children with autistic spectrum disorder. *Cognition* 113, 37–44. doi: 10.1016/j.cognition.2009.07.007
- Happe, F. G. E. (1995). Understanding minds and metaphors – insights from the study of figurative language in autism. *Metaphor Symb. Act.* 10, 275–295. doi: 10.1207/s15327868ms1004_3
- Hobson, R. P. (1984). Early childhood autism and the question of egocentrism. *J. Autism Dev. Disord.* 14, 85–104. doi: 10.1007/BF02408558
- Kessler, K., and Thomson, L. A. (2009). The embodied nature of spatial perspective taking: embodied transformation versus sensorimotor interference. *Cognition* 114, 72–88. doi: 10.1016/j.cognition.2009.08.015
- Kessler, K., and Wang, H. F. (2012). Differently embodied transformations in visuo-spatial perspective taking. *Cogn. Process.* 13, S21–S22.
- Keysar, B., Lin, S., and Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition* 89, 25–41. doi: 10.1016/S0010-0277(03)00064-7
- Langdon, R., and Coltheart, M. (2001). Visual perspective-taking and schizotypy: evidence for a simulation-based account of mentalizing in normal adults. *Cognition* 82, 1–26. doi: 10.1016/S0010-0277(01)00139-1
- Lee, A., Hobson, R. P., and Chiat, S. (1994). I, You, Me, and Autism – an experimental study. *J. Autism Dev. Disord.* 24, 155–176. doi: 10.1007/BF02172094
- Leekam, S., BaronCohen, S., Perrett, D., Milders, M., and Brown, S. (1997). Eye-direction detection: a dissociation between geometric and joint attention skills in autism. *Br. J. Dev. Psychol.* 15, 77–95. doi: 10.1111/j.2044-835X.1997.tb00726.x
- Leslie, A. M. (1987). Pretense and representation – the origins of theory of mind. *Psychol. Rev.* 94, 412–426. doi: 10.1037/0033-295X.94.4.412
- Leslie, A. M., and Frith, U. (1988). Autistic children’s understanding of seeing, knowing and believing. *Br. J. Dev. Psychol.* 6, 315–324. doi: 10.1111/j.2044-835X.1988.tb01104.x
- Masangkay, Z. S., McCluskey, K. A., McIntyre, C. W., Simsknig, J., Vaughn, B. E., and Flavell, J. H. (1974). Early development of inferences about visual percepts of others. *Child Dev.* 45, 357–366. doi: 10.2307/1127956
- Mazzarella, E., Ramsey, R., Conson, M., and Hamilton, A. (2013). Brain systems for visual perspective taking and action perception. *Soc. Neurosci.* 8, 248–267. doi: 10.1080/17470919.2012.761160
- Mizuno, A., Liu, Y. N., Williams, D. L., Keller, T. A., Minshew, N. J., and Just, M. A. (2011). The neural basis of deictic shifting in linguistic perspective-taking in high-functioning autism. *Brain* 134, 2422–2435. doi: 10.1093/brain/awr151
- Moll, H., Carpenter, M., and Tomasello, M. (2007). Fourteen-month-olds know what others experience only in joint engagement. *Dev. Sci.* 10, 826–835. doi: 10.1111/j.1467-7687.2007.00615.x
- Moll, H., and Tomasello, M. (2004). 12- and 18-month-old infants follow gaze to spaces behind barriers. *Dev. Sci.* 7, F1–F9. doi: 10.1111/j.1467-7687.2004.00315.x
- Moll, H., and Tomasello, M. (2006). Level 1 perspective-taking at 24 months of age. *Br. J. Dev. Psychol.* 24, 603–613. doi: 10.1348/026151005X55370
- Ozonoff, S., Pennington, B. F., and Rogers, S. J. (1991). Executive function deficits in high-functioning autistic individuals: relationship to theory of mind. *J. Child Psychol. Psychiatry* 32, 1081–1105. doi: 10.1111/j.1469-7610.1991.tb00351.x
- Parsons, L. M. (1987). Imagined spatial transformation of one’s body. *J. Exp. Psychol. Gen.* 116, 172–191. doi: 10.1037/0096-3445.116.2.172
- Piaget, J., and Inhelder, B. R. (1956). *The Child’s Conception of Space*. London: Routledge and K. Paul.
- Reed, T. (2002). Visual perspective taking as a measure of working memory in participants with autism. *J. Dev. Phys. Disabil.* 14, 63–76. doi: 10.1023/A:1013515829985
- Reed, T., and Peterson, C. (1990). A comparative study of autistic subjects’ performance at two levels of visual and cognitive perspective taking. *J. Autism Dev. Disord.* 20, 555–567. doi: 10.1007/BF02216060
- Samson, D., Apperly, I. A., Kathirgamanathan, U., and Humphreys, G. W. (2005). Seeing it my way: a case of a selective deficit in inhibiting self-perspective. *Brain* 128, 1102–1111. doi: 10.1093/brain/awh464
- Senju, A. (2012). Spontaneous theory of mind and its absence in autism spectrum disorders. *Neuroscientist* 18, 108–113. doi: 10.1177/1073858410397208
- Senju, A., Southgate, V., White, S., and Frith, U. (2009). Mindblind eyes: an absence of spontaneous theory of mind in asperger syndrome. *Science* 325, 883–885. doi: 10.1126/science.1176170
- Shelton, A. L., Clements-Stephens, A. M., Lam, W. Y., Pak, D. M., and Murray, A. J. (2012). Should social savvy equal good spatial skills. The interaction of social skills with spatial perspective taking. *J. Exp. Psychol. Gen.* 141, 199–205. doi: 10.1037/a0024617
- Southgate, V., Senju, A., and Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychol. Sci.* 18, 587–592. doi: 10.1111/j.1467-9280.2007.01944.x
- Surtees, A. D. R., Butterfill, S. A., and Apperly, I. A. (2012). Direct and indirect measures of Level-2 perspective-taking in children and adults. *Br. J. Dev. Psychol.* 30, 75–86. doi: 10.1111/j.2044-835X.2011.02063.x
- Surtees, A., Apperly, I., and Samson, D. (2013). Similarities and

- differences in visual and spatial perspective taking processes. *Cognition* 129, 426–438. doi: 10.1016/j.cognition.2013.06.008
- Tan, J., and Harris, P. L. (1991). Autistic children understand seeing and wanting. *Dev. Psychopathol.* 3, 163–174. doi: 10.1017/S0954579400000055
- Warreyn, P., Roeyers, H., Oelbandt, T., and De Groote, I. (2005). What are you looking at. Joint attention and visual perspective taking in young children with autism spectrum disorder. *J. Dev. Phys. Disabil.* 17, 55–73. doi: 10.1007/s10882-005-2201-1
- Yirmiya, N., Sigman, M., and Zacks, D. (1994). Perceptual perspective-taking and seriation abilities in high-functioning children with autism. *Dev. Psychopathol.* 6, 263–272. doi: 10.1017/S0954579400004570
- Yu, A., and Zacks, J. (2010). The role of animacy in spatial transformations. *Mem. Cogn.* 38, 982–993. doi: 10.3758/MC.38.7.982
- Zacks, J., Rypma, B., Gabrieli, J. D., Tversky, B., and Glover, G. H. (1999). Imagined transformations of bodies: an fMRI investigation. *Neuropsychologia* 37, 1029–1040. doi: 10.1016/S0028-3932(99)00012-3
- Zacks, J. M., and Tversky, B. (2005). Multiple systems for spatial imagery: transformations of objects and bodies. *Spat. Cogn. Comput. Interdiscip. J.* 5, 271–306. doi: 10.1207/s15427633scc0504_1
- Zacks, J. M., Vettel, J. M., and Michelon, P. (2003). Imagined viewer and object rotations dissociated with event-related fMRI. *J. Cogn. Neurosci.* 15, 1002–1018. doi: 10.1162/089892903770007399
- Zwicker, J., White, S. J., Coniston, D., Senju, A., and Frith, U. (2011). Exploring the building blocks of social cognition: spontaneous agency perception and visual perspective taking in autism. *Soc. Cogn. Affect. Neurosci.* 6, 564–571. doi: 10.1093/scan/nsq088
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Received: 28 June 2013; accepted: 18 September 2013; published online: 08 October 2013.
- Citation:** Pearson A, Ropar D and Hamilton AF (2013) A review of visual perspective taking in autism spectrum disorder. *Front. Hum. Neurosci.* 7:652. doi: 10.3389/fnhum.2013.00652
- This article was submitted to the journal *Frontiers in Human Neuroscience*. Copyright © 2013 Pearson, Ropar and Hamilton. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The primacy of social over visual perspective-taking

Henrike Moll* and Derya Kadipasaoglu

Department of Psychology, University of Southern California, Los Angeles, CA, USA

Edited by:

Klaus Kessler, University of Glasgow, UK

Reviewed by:

Ian Apperly, University of Birmingham, UK

Mark Gardner, University of Westminster, UK

***Correspondence:**

Henrike Moll, Department of Psychology, University of Southern California, SGM 501, 3620 South McClintock Ave., Los Angeles, CA 90089, USA
e-mail: hmoll@usc.edu

In this article, we argue for the developmental primacy of social over visual perspective-taking. In our terminology, social perspective-taking involves some understanding of another person's preferences, goals, intentions etc. which can be discerned from temporally extended interactions, including dialog. As is evidenced by their successful performance on various reference disambiguation tasks, infants in their second year of life first begin to develop such skills. They can, for example, determine which of two or more objects another is referring to based on previously expressed preferences or the distinct quality with which these objects were jointly explored. The pattern of findings from developmental research further indicates that this ability emerges sooner than analogous forms of visual perspective-taking. Our explanatory account of this developmental sequence highlights the primary importance of joint attention and the formation of common ground with others. Before children can develop an awareness of what exactly is seen or how an object appears from a particular viewpoint, they must learn to share attention and build common "experiential" ground. Learning about others' as well as one's own "snapshot" perspectives in a literal, i.e., optical sense of the term, is a secondary step that affords an abstraction from all (prior) pragmatic involvement with objects.

Keywords: referential communication, perspective-taking, joint attention, theory of mind (ToM), intersubjectivity

Visual perspective-taking tasks typically entail another agent who embodies the spatial coordinates that the participant has to consider. They are thus at least minimally social in the sense that someone else is co-present and available for social interaction (see Schütz, 1932). In line with this, children with autism, whose difficulties are known to be first and foremost social in nature, struggle to determine how others see things from their viewpoint (Reed, 2002; Hamilton et al., 2009; Yu et al., 2011; but see Hobson, 1984). At the same time, however, a "cold-cognitive" assessment or computation of how objects relate to one another in space is arguably less of a social affair than understanding another's affective, conceptual, or epistemic attitude toward a situation (see Fishbein et al., 1972).

In this article, we adopt the opposition of visuo-spatial and social perspective-taking employed by the editors. We will argue from a developmental approach that social perspective-taking is primary and precedes visual perspective-taking in human ontogeny. Our claim is that children first learn to take perspectives in situations that are not defined by differences in how self and other perceive objects visually but by differences in their experiential backgrounds, i.e., in what they did, witnessed, or heard. It might seem more complex to keep track of another's prior encounters and engagement with things than to compute his instantaneous visuo-spatial relation to an object in the room. Yet, it will become clear that infants readily note and update "experiential records" (Perner and Roessler, 2012, p. 522). Registering and remembering what others did, witnessed, or mentioned is less of a task demand for them than a helpful cue to others' goals and intentions. Per definition, no such cues from prior encounters are

available in visual perspective-taking tasks that revolve entirely around momentary visuo-spatial relations.

First, we will review referential ambiguity tasks that are typically solved in the second year of life. It will become obvious that infants readily rely on others' previous expressions of attitude, their prior attentional engagements with objects, and previous discourse to solve the reference problem. These manifold abilities of infants to establish reference against the background of prior interactions are subsumed under "social perspective-taking."

An overview of studies on visual perspective-taking will show that this ability has its onset noticeably later. Again, it is generally taken for granted that perceptual perspective-taking precedes and serves as a foundation for the "deeper" forms of social perspective-taking (Kessler and Thomson, 2010). The same is suggested by accounts of mutual knowledge according to which physical co-presence is the easiest and least error-prone way to arrive at mutual knowledge (Clark and Marshall, 1981; Schiffer, 1972). These assumptions are seriously called into question by the empirical fact that visual perspective-taking does not precede but follows social perspective-taking ontogenetically.

An excursion into the early development of graphic skills lends further support to the idea that knowledge of visual perspectives is a relatively late cognitive achievement that is derivative of social perspective-taking. We will conclude with a programmatic attempt to explain this developmental sequence with the social and cooperative nature that sets humans apart from other animals.

THE ROLE OF THE EXPERIENTIAL BACKGROUND

PRIOR AFFECTIVE EXPRESSIONS

Affective displays are key indicators as to how people will behave toward objects. In a seminal study, 14- and 18-month-old infants were presented with two food items: crackers and broccoli (Repacholi and Gopnik, 1997). The infants opted for the crackers, whereas an adult displayed the opposite preference. When the adult, without looking at either dish, later requested food from the infants, the younger ones gave her what they themselves liked (crackers), whereas the older ones selected the broccoli.

As was made clear by Perner et al. (2005), understanding perspectives in *sensu strictu*, as evidenced by an explicit acknowledgment of different takes on the self-same thing (“I cannot stand broccoli, but she likes it!”) is not necessary for this test. The infants just had to realize that the other and broccoli “go together,” and so an understanding of objective “person-object couplings” suffices to pass this test. Nonetheless, the older infants were able to learn about the other’s taste preference from her prior expressions. A study by Egyed et al. (2007) confirmed that, in the absence of ostensive cues (which gear infants toward more object-centered interpretations such as “Broccoli is good”; see Gergely et al., 2007), 14-month-olds track specific persons’ affective displays toward objects and expect them to behave in accordance with them later.

“Emotional eavesdropping” (Repacholi and Meltzoff, 2007) provides further support that infants act differently vis-à-vis others depending on their previously expressed affective attitudes. When 18-month-olds witness an adult reprimand another for performing a novel action, the infants later imitated the act less when the adult was present as opposed to when he was absent. Independently of their own desire or interest to perform an act, infants thus alter their behavior as a function of others’ attitudes toward objects and actions.

PRIOR ENGAGEMENT

Infants use various other cues to disambiguate reference. A powerful one is the other’s familiarity with or ignorance of objects and their locations (see O’Neill, 1996, for an influential study with 2- and 2.5-year-olds). In their modification of a word learning study (Akhtar et al., 1996), Tomasello and Haberl (2003) found that 1-year-olds knew which of three objects an adult requested from them based on her prior engagement with the objects. When the adult excitedly asked infants for a toy, 12-month-olds chose the one that was new for the adult because she failed to see it in the past. Even though the infants themselves were equally familiar with all toys, their responses showed that they knew what the other had and had not witnessed a few moments prior.

MacPherson and Moore (2010) directly contrasted what an adult knew with what the infant herself was familiar with. In their study, two objects were mutually familiar for adult and infant, a third object was new for the infant and a fourth was new for the adult, but “old” for the infant. When the adult later excitedly requested a toy, 13-month-olds egocentrically chose what was new for themselves, while 19-month-olds selected the toy that was new for the adult and not for them.

Many studies have not just confirmed that infants readily track others’ experiential backgrounds but have also yielded insights

into the scope and limits of this skill. Joint attention has been shown to play an important role in interactive test situations. Having observed as mere onlookers how an agent engages with objects was insufficient for 14-month-olds to identify what the agent requested from them based on his knowledge vs. ignorance of the different objects. When the infant and the agent jointly engaged around the objects, infants successfully determined which things the other did and did not know (Moll and Tomasello, 2007; Moll et al., 2007, 2008). In contrast to mere onlooking, joint attention makes the co-attenders’ familiarity with the object “mutually transparent” (Eilan, 2005)—it leaves no room for doubt that the object has been registered. Furthermore, it seems that unless the questioner clearly conveys that her excitement is elicited by something that is new for her individually, infants have a general bias to point out what is mutually familiar and unifies self and other in prior bouts of shared experiences (Saylor and Ganea, 2007; Liebal et al., 2009, 2011).

In a study that aimed to test false belief understanding, infants saw an adult as striving for different goals depending on what he witnessed earlier (Buttelmann et al., 2009). After the adult had placed an object in a box, he either attentively watched it being moved to a different container (true belief) or failed to witness the transfer (false belief). When the adult later approached the box from which the object had been removed, 18-month-olds helped him to get the box open (“He must want something else from this box!”) in the true belief, but retrieved the object from its new location in the false belief condition. Again, this demonstrates that infants take what others have witnessed into account when acting and responding toward them. They revert to the background constituted by past experiences and use it to inform them about an agent’s desires, goals, and intentions.

Looking-time studies on false belief understanding further support this idea (e.g., Onishi and Baillargeon, 2005; Surian et al., 2007; Kovács et al., 2010). Even in their first year of life, infants look longer when they see an agent acting in a way that disaccords with his prior perceptual experiences (he acts as if he knows something he did not observe) than when they see him behave in ways that are consistent with what he observed. Whether belief understanding can be captured with this method remains the subject of an ongoing debate (Perner and Ruffman, 2005; Low and Perner, 2012), but what this research unequivocally demonstrates is that infants at a very young age are aware of what others have and have not registered perceptually. The findings also relativize the importance of joint attention suggested by interactive studies, because infants in looking-time tests usually do not jointly attend with the other and, in some cases, have not even reached the age at which they are able to do so. Joint attention might thus only play a critical role when infants have to directly respond to the agent in a communicative or cooperative act, which might require a more explicit understanding of knowledge and ignorance.

PRIOR DISCOURSE

To not talk past, but speak with each other, interlocutors must know what they can and cannot presuppose as mutually given. Part of what defines the mutually given is the shared prior

discourse—what Clark and Marshall (1981) refer to as “linguistic co-presence” in their model of mutual knowledge formation. Anecdotal evidence of “egocentric speech” (Piaget, 1929, 1955) alongside experimental data questioned children’s skills in communicative perspective-taking. Young children tend to use pronouns (e.g., O’Neill and Holmes, 2002) and definite articles (Maratsos, 1976; Power and dal Martello, 1986) without having provided the antecedent. Their descriptions are often not specific enough to allow the listener to discern reference—even after requests for clarification were made, thus challenging effortless communication (see Glucksberg and Krauss, 1967; Deutsch and Pechmann, 1982; Sonnenschein and Whitehurst, 1984). Generally, young children have a tendency to underestimate the informativeness that is needed to communicate effectively (Olson and Torrance, 1987).

At the same time, evidence accumulates that even infants adjust their (speech) behavior according to what has been shared linguistically. For example, 2-year-olds use more informative naming constructions when a referent is new than when it is given, in the sense that it was part of previous discourse. Matthews et al. (2006) had 2-year-olds watch a video of a character performing an action (e.g., a clown jumping) together with an assistant. The assistant mentioned the character to the child. Another adult, who had either participated in the discourse or not then asked children to narrate what happened. In their replies, the children referred to the character more often with a pronoun (instead of a full noun) when their interlocutor had participated in the prior discourse than when he was not part of this discourse (see Nayer and Graham, 2006, for similar results with 3-year-olds). In Clark and Marshall’s (1981) terms, the children tailored their references to the linguistic copresence they shared with their particular interlocutor.

On the side of comprehension, even 1-year-olds are sensitive to what is and is not linguistically co-present. Ganea and Saylor (2007) found that 15- and 18-month-olds rely on a person’s prior verbal reference to an absent object to determine what the same person is speaking of a few moments later. After an adult made clear that she was searching for a particular object (e.g., a puppy), she exclaimed that she knew where “it” was and led the infant to a cabinet. Two objects—the target (puppy) and a distractor—were revealed, and the adult ambiguously asked “Can you get it for me?” Infants at both ages selected the target object, thus showing that they located the referent in the adult’s prior speech. Echoing the findings on prior attentional engagement (e.g., Moll et al., 2008; Liebal et al., 2009), the infants also knew with which particular person they shared the linguistic background: When a different adult than the one who had searched articulated the request, the infants grasped objects randomly.

A further indication that infants keep track of and update records of linguistic co-presence is their appropriate use of elliptical constructions in discourse. In a study by Salomo et al. (2010), 2-year-olds were asked, “What’s the agent doing now?” after watching and hearing verbal descriptions of videos showing either the same action performed on different patients (e.g., a frog feeding a duck vs. a ladybug) or different actions performed on the same patient (e.g., a frog feeding vs. washing a duck). In their answers, the children omitted reference to the patient when

it remained the same and was thus given in the prior discourse. When the patient changed and was thus new, the same children made reference to it with a lexical noun. The children thus knew when null-references were and were not warranted given the discourse background. Additional evidence that 2-year-olds know which information is obligatory vs. optional in speech stems from observations of children who acquire “null-argument” languages; i.e., languages that allow the omission of subjects and objects given the appropriate discourse context (Serratrice, 2005).

Taken together, these findings clearly demonstrate that infants produce and understand gestures and speech acts against the background of their prior interactions with other persons (see Wittgenstein, 2001; Tomasello, 2008). What infuses the gestures and speech acts with meaning is the intersubjectively shared background of prior experiences. Through joint attention, infants construct a common ground (Clark, 1996) with specific other persons, and they discriminate between the dyad-specific common grounds, keeping track of what they have and have not shared with whom. In their attempts to secure reference, they naturally revert to these backgrounds, which becomes particularly obvious under conditions of potential ambiguity.

VISUAL PERSPECTIVE-TAKING: NO HELP FROM THE BACKGROUND

None of the above is available in visual perspective-taking. All that is relevant here are instantaneous viewing angles and momentary spatial relations. The experiential background offers no help to solve referential ambiguity in these tests. In fact, a prerequisite that has to be met to guarantee the validity of these tests is that the candidate objects be “experientially neutral,” i.e., that target and distractor cannot be distinguished by any distinct roles they played in prior interactions. The correct response has to depend entirely on the objects’ visibility (level 1) or mode of presentation (level 2 perspective-taking, see e.g., Flavell, 1992, for the distinction of level 1 vs. level 2) from a particular viewpoint. We will limit our analysis to level 1 visual perspective-taking, i.e., the ability to determine what another can and cannot see. This level of perspective-taking emerges a couple of years prior to level 2, and is structurally similar to the tasks above, which dealt with children’s understanding of what others desired, witnessed, or spoke about. Level 2 is a more effortful, qualitatively distinct (Kessler and Rutherford, 2010), phylogenetically recent (human-specific) skill, that requires an explicit understanding of perspectival differences and, in the absence of autism (see Hamilton et al., 2009), emerges between 4 and 5 years.

Children first exhibit an understanding of what others can and cannot see at around 2 years of age and older. For example, when 24-month-olds witness an adult searching for something, they preferably hand her an object that is blocked from the adult’s view instead of a mutually visible one (Moll and Tomasello, 2006). In a similar task by Nummoo and Bloom (2008), 31-month-olds also mostly selected an object that was hidden from an adult’s view when he pretended to be searching for something. (One should note that there was a confound with gaze direction in this study: The adult looked straight at the visible distractor object when asking “where” the referent was, allowing children to act on a simple

heuristic that people do not search for things at which they are currently looking.)

In one of several tasks administered by Masangkay et al. (1974), children between 2 and 3 correctly judged that an adult sitting across from them could not see an apple depicted on the front of a card held between them. Hughes and Donaldson (1979) found that 3-year-olds knew where to place a doll in a house so that none of several policemen at various positions could see her. In another study, 2.5-, but not 2-year-olds, granted an adult visual access to an object he desired to see by either revealing the object from behind an occluder or moving away the occluder (Lempers et al., 1977). In an “analogy task” developed by Yaniv and Shatz (1990) 3.5-year-olds were able to place a duck so that a doll perceiver saw the same part or side (e.g., its back) of the duck as another doll that looked at an identical duck.

In sum, we find that level 1 perspective-taking as demonstrated by tests using interactive methods emerges between the second and third birthday. This ability comprises percept production (enabling another to see something), percept diagnosis (judging what another sees), and percept deprivation (hiding objects from another). In Clark and Marshall’s (1981) terms, it is now that children have come to understand when mutual knowledge is and is not supported by immediate physical co-presence.

However, children at this age are far from being proficient at visual perspective-taking. On the contrary, striking limitations have been identified. Under the age of 3, children are unable to hide an object from an adult by placing a barrier between her and the object (Flavell et al., 1978; McGuigan and Doherty, 2002). Two-year-olds also struggle to select appropriate referring expressions depending on what their interlocutor can see. While the children in Matthews et al.’s (2006) above-mentioned study successfully tailored their expressions to the prior discourse, they did not adjust their speech accordingly when the adult’s visual access to the video was manipulated. More concretely, they did not produce more full nouns (instead of the less informative pronouns) when the adult failed to see the video compared to when he saw it. In Yaniv and Shatz’s (1990) study, 3-year-olds preferably positioned the duck facing the doll, even when asked to place it so that the doll would see its back. They thus exerted a bias to generate the canonical or good—in this case the frontal—view of the object, irrespective of the instruction (see Light and Nix, 1983; more on a similar phenomenon in children’s drawings below).

A DEVELOPMENTAL LAG

Taken together, these studies point at a developmental lag between social and visual perspective-taking. Infants rely on prior joint perceptual experiences including previously shared discourse at least 1 year before they take into account others’ visuo-spatial relations to the things around them when they discern or establish reference. This is a significant *décalage* given the young age of these children. Contra Clark and Marshall (1981) and contra intuition, immediate physical co-presence does not necessarily facilitate the delineation of common ground. While it is true that physical co-presence often rightly signals that a given object figures in the common ground, the same co-presence can hamper children’s ability to identify what is mutually given from what they have privileged access to as individuals. It can trick them into

falsely assuming that an object they see is perceptually available to the other as well. The strong priority that is ascribed to an *ad-hoc* formation of mutual knowledge based on immediate or potential physical co-presence is thus called into question by this developmental sequence.

That the lag is real and robust becomes particularly obvious in studies in which an understanding of knowledge and ignorance is directly contrasted with visual perspective-taking. Moll et al. (2010) compared 2-year-olds’ ability to detect an adult’s ignorance due to absence vs. impeded vision. When the adult disengaged entirely from her interaction with the child by leaving after having shared two toys with her, the children later knew that the adult was unfamiliar with a third object that they were presented with. But when the adult remained co-present with her visual access to the third object blocked by a barrier as the child explored it, the children later acted as if the adult was familiar with this object. They failed to recognize the barrier’s effect.

A very similar pattern emerged in Nurmsoo and Bloom’s (2008) study. In their second experiment, 31-month-olds had no problem identifying what an adult was looking for when she had hidden one object but was absent when the other was hidden—thus making her ignorant of the second object’s location. By contrast, children this age found it relatively difficult to determine what the adult searched for when he had seen neither placement but was spatially positioned so that he could not see one of the objects (Experiment 1). Similarly, and as mentioned above, while 2-year-olds in Matthews et al.’s (2006) study readily switched to more informative references when an object was not shared in prior discourse with an adult, they failed to adjust the informativeness of their speech accordingly when the referent was blocked from the adult’s sight. Taken as a whole, these studies clearly show that young children can draw the knowledge-ignorance distinction before they solve otherwise identical tests that tap visual perspective-taking.

The same gap has been identified with looking-time measures as well. Again, when this method is applied, infants as young as 7 months show a sensitivity to the manipulation of perceptually induced beliefs (Kovács et al., 2010). They look longer when an agent behaves in a way that is inconsistent with what she witnessed earlier than when her behavior matches her prior observations (someone looks for something where she last saw it). In contrast, the youngest age for which level 1 visual perspective-taking has been documented with the looking-time technique is 13–16 months (Luo and Baillargeon, 2007; Sodian et al., 2007; Luo and Beck, 2010). For example, when 13-month-olds repeatedly see an adult reaching for one of two toys, they form an expectation that he will keep doing so—as evidenced by longer looks when he suddenly reaches for the previously ignored toy. But they only form this expectation when the agent is able to see the alternative object, and thus disprefers it. No extended looks were shown when the non-chosen toy was blocked from the agent’s view, and so simply unseen.

Further indication that 13-month-olds have rudimentary skills in visual perspective-taking stems from a study that is purported to test false belief comprehension (Surian et al., 2007). In this looking-time experiment, a caterpillar’s knowledge of his preferred object’s location was manipulated by the presence/absence

of a barrier impeding the caterpillar's, but not the child's, vision of the object. The authors do not interpret their results in terms of visual perspective-taking. But partly because looking-time measures involve no "task" (the child is not asked or prompted to respond to anything in particular), it remains open which aspect infants mainly reacted to or found harder to process: realizing the barrier's defeating effect on the agent's vision, or keeping track of what he did and did not witness.

In either case, the same developmental lag that is found with interactive response methods becomes manifest when looking times are applied, albeit at a younger age—reflecting the reduced task affordances of this method. The fact that the developmental order pervades different research methods shows its robustness. But it has to be emphasized that the lag is limited to level 1 visual perspective-taking and its corresponding counterparts in social perspective-taking. A more synchronous pattern is found at level 2, which affords an explicit knowledge of the possibility of alternative, and potentially false, views. This knowledge, which spans across visual and social perspectives alike, is formed between 4 and 5 years—supporting the idea of a common cognitive thread that runs through various perspective problems (see Perner et al., 2003; Moll and Meltzoff, 2012). The gap that calls for an explanation thus only exists in the early beginnings of perspective-taking, before a more abstract and uniform understanding of perspectives develops in late preschool.

The pressing question then is how the counterintuitive sequence of visual perspective-taking preceding social perspective-taking observed in the early years can be explained. We will make a first explanatory attempt by addressing the more specific question of why visual perspective-taking might pose a particular challenge (see Moll and Meltzoff, 2012).

SHARED PERCEPTUAL SPACES

We argue that young children have a proclivity to treat social interaction as a sufficient condition for shared perceptual availability: "When you and I are co-present and engaged, you should be able to perceive what I perceive." An impression of a shared perceptual space is induced, and only later overcome once children learn more about and attend more to the specific defeating conditions of perception, such as a blocked line of sight.

Support for the idea that co-presence and social engagement create an illusion of shared perception comes from experimental data with both children and adults. Glucksberg and Krauss (1967) report that preschoolers produced iconic gestures and used demonstratives ("It goes like this!") to describe objects to their conversational partner who sat across from an occluder. The work of Keysar and colleagues shows that even adults have a prepotent tendency to assume that others around them share their perceptual access to objects, even when this is not true (e.g., Keysar et al., 2003; Epley et al., 2004; Keysar, 2007). Consistent with what we know about children, adults are biased to over- rather than underestimate what the other sees or knows (Keysar and Henly, 2002; see also Bernstein et al., 2007). Interestingly, the thicker or richer the common ground shared by two people, the more likely they are to overrate the success of their communicative attempts (Wu and Keysar, 2007). The more that is shared, the less prepared one is to identify when something is not

shared. A vast overlap in what is perceptually accessible weakens the alertness to check if a particular object is mutually given or not. In support of this, it was found that people communicate less informatively to a concrete other person who "co-inhabits" their perceptual space than to a merely imagined interlocutor (Schober, 1993). This is much in line with our developmental finding that corporeal co-presence, and thus a high overlap in what can potentially be turned into an object of shared attention, hampers young children's ability to detect others' ignorance (Moll et al., 2010).

This overestimation effect also helps to explain young children's notoriously poor perspective-taking skills when speaking on the phone. It has long been known that children use manual gestures, demonstratives, and non-specific references during phone conversations (Bordeaux and Willbrand, 1987; Warren and Tate, 1992)—indicating that they are unaware of the fact that they and the things around them cannot be seen. In our interpretation, the shared discourse elicits the false impression of a generally shared perceptual space that spans across different sense modalities, including vision. That is, verbally established co-presence leads to the illusory impression of shared visual perception.

This idea, however, is called into question by experiments suggesting that others' viewpoints make their way into our considerations effortlessly and automatically (Qureshi et al., 2010). When asked how many items they see in a visual array, adults and school-age children are slower and less accurate in their judgments if their visual input mismatches that of another agent who is part of the scene they watch (Samson et al., 2010; Surtees and Apperly, 2012; see also Surtees et al., 2012). Two things can be said to reconcile these findings with our overestimation thesis. Firstly, it is conceivable that once level 1 visual perspective-taking has been practiced for years, it becomes "second nature" or automated. Secondly, the participants' situation differs drastically between the studies. In those studies supporting the overestimation thesis, the child interacts with the other directly, which might let the perspectival differences between them dissolve "in the heat of the moment." In the tasks suggesting automatic perspective-taking, participants have a contemplative, theoretical distance to the other, who figures in the array like an object. This theoretical distance could highlight the other's position in relation to the remaining items in the scene. The two sets of findings thus do not necessarily contradict each other.

A GLANCE AT EARLY PICTURE-MAKING

It was speculated that before children's perception is corrupted by language and thought, they ought to see the world with innocent, i.e., objective eyes (see Matisse, 1953) and even master perfectly the art of drawing in linear perspective (Bühler, 1930; Sully, 1895). But of course, by the time children have the motor skills and motivation to depict objects and events, they have long been language- and concept-using beings who have passed any hypothetical phase of innocent vision (see Costall, 1997, 2001).

When children begin to draw figuratively, they do not faithfully translate three-dimensional objects onto two-dimensional picture planes. They show no intention to depict things exactly the way they appear to them from one fixed point of observation. Drawing does not serve the goal of imitating visual experiences. As a famous dictum says, children "draw what they know, not

what they see” (but see Arnheim, 1974, p. 164, for rightly criticizing the false opposition of seeing and knowing that is employed here)—exhibiting a style dubbed “intellectual realism” (Luquet, 1927). They include aspects and elements in their pictures that cannot be seen from their present perspective and may not be visible from any particular, single viewpoint. They create a “good” or ideal view of objects by depicting features they consider relevant or important and omitting what is irrelevant. The goal is not to produce a correct perspectival reconstruction but to show objects in their typical form and thus to capture their constitutive or essential features. For example, a cup will be depicted in canonical fashion with a handle on its side (ideal for grasping, see Cox, 1991). Likewise, humans are shown in their canonical frontal view with a face including two eyes (ideal for social interaction), whereas trunk, nose and other parts might be left out (Cox, 1997).

Also, young children mostly produce images spontaneously from memory and imagination (Golomb, 2004). When presented with a model to guide their drawing activity, they rarely look up to see what the object exactly looks like. The model serves as a source of inspiration—it provides a theme or motif and is relevant insofar it exemplifies a generic object (Luquet, 1927), but it is not adhered to as an original that ought to be replicated. Again, what this indicates is that children do not intend but fail to draw from a fixed perspective.

In his essay “Perspective as symbolic form,” Panofsky (1927) pointed out that a faithful reconstruction of what is seen from a particular viewpoint affords a severe abstraction from the content of experience. In his own words, it is a modern technique that rests on a motivation to strip away the experiential or “given” space and substitute it with a systematic, purely visual space. An individualistic and somewhat arbitrary factor thereby gets introduced, because one commits to showing the scene from a single, static point of observation. Any ordering according to what is regarded important or relevant has to make way for a strictly geometric ordering.

With this held in mind, it becomes much less puzzling why an awareness of visual perspectives emerges rather late—not just in history, but in ontogeny as well (see Gablik, 1977, for parallels between the history and genetic development of visual art). Though children at age 5 and older can be induced to draw what they see, it is not before 7 or 8 years that they spontaneously create view-dependent images (Davis, 1983; Cox, 1991). Even at this age, their advances are such that they acknowledge partial occlusion and draw only what is visible (e.g., the correct number of faces of a cube), but they still do not depict the visible parts precisely in the way they appear (e.g., with lines converging in a vanishing point; Bremner and Batten, 1991; Cox, 1991).

What is of primary importance to children is to share the world of those around them. Precisely how this shared world presents itself from one specific vantage point is secondary and does not become thematic in the very early stages. First and foremost, drawing serves to “make sense of the world” (Arnheim, 1969, p. 257)—and this is true ontogenetically as well. Young children draw to narrate events and give “shape and order” (Cox, 2005) to their experiences. We want to go further and argue that picture-making primarily serves to make sense of the social world, as one of the first and most frequent motifs is the human figure

(Maitland, 1895; Lark-Horovitz et al., 1939; see Cox, 1993, for an overview). But not just the themes or motifs are social; so is the process of drawing. It is an activity that is typically shown in the presence of another to whom the child narrates as she draws, and for whom she might create the picture as a gift. The graphic product in itself can hardly be interpreted without the accompanying speech in which children reconstruct their experiences and reveal what they intend to draw (Cox, 2005).

In either case, we find that the relatively late onset of taking others’ visual perspectives is paralleled by a late emergence of the use of perspective in drawings. Young children’s pictures document their inattention to specific visual perspectives. Just like there was no motivation to graphically capture objects from specific, transient viewpoints in the early history of visual art (Panofsky, 1927), so do children show no interest in representing things precisely the way they happen to see them. They ignore the contingent ways in which things appear momentarily for the sake of capturing what belongs to an object more generally. This also becomes manifest in perceptual self-reports. When preschoolers are asked to indicate how they perceive a visual array by choosing from among a set of different pictures, they often judge incorrectly and select a picture showing the ideal rather than their own view (Liben and Belknap, 1981; Light and Nix, 1983). The upshot is that children’s drawings are one of several pieces of converging evidence that young children pay little attention to differences in visual perspective. Others are their faulty perceptual reports, their behavior during phone conversations, and, as we have seen, profound struggles with visual perspective-taking—neither of which require graphic skills.

CONCLUDING REMARKS

Humans are extraordinarily *relational* and interdependent beings (MacMurray, 1961). They are adapted to rely on and cooperate with others in a way that is unparalleled in the animal kingdom (Gintis et al., 2003). Especially in the early beginnings, a human individual is entirely dependent on others’ care, attention, and sharing of knowledge (Csibra and Gergely, 2011). What is crucial at this early stage is that the child comes to share the world of those around her. She accomplishes this by jointly attending to things with others. It is in these bouts of joint attention that the child learns about objects: their *gestalts*, functions, and labels etc. Importantly, these are perspective-invariant properties. The focus lies on the object and its qualities, not on the different perspectives from which each co-attender perceives it (Campbell, 2012; Moll and Meltzoff, 2012; Seemann, 2012). Only once it can be taken for granted that we attend to the same thing, is there room, in a second step, to “objectify” the different viewpoints from which each of us perceives the object. As Campbell (2012, p. 428) puts it, “The point is that a grasp of the different perspectives from which a thing may be experienced should not be allowed to take on a life of its own; this grasp of the different perspectives from which a thing may be experienced is always grounded in a prior knowledge of which thing is in question.”

But this merely seems to explain why joint attention precedes knowledge of perspectives, not why children engage in social perspective-taking before visual perspective-taking. However, we think that these two things are related. In joint attention, one’s

knowledge of the object becomes mutually transparent and so does the expression of one's attitude toward it. While the focus of joint attention is the object itself, it simultaneously informs us of the other's knowledge of it as well as her take on it. Joint attention thus directly supports the forms of social perspective-taking discussed in this article, which are critical for cooperative communication and other forms of collaborative activities.

The visuo-spatial positions of the co-attenders, in contrast, remain entirely in the background. Firstly, the viewing angles involved usually bear no significance with regard to the object, its qualities, or the other's attitude toward it. Secondly, given the dynamic character of joint attention, these perspectives rarely remain constant but tend to fluctuate over the course of exploration, as the object gets manipulated and/or the spatial positions changed. Joint attention thus directly paves the way to early forms of social, but not visual perspective-taking.

The picture looks very different for non-human primates that possess simple forms of visual perspective-taking. Chimpanzees have been shown to preferably approach food that is blocked from a dominant individual's sight (Hare et al., 2000), and to seek out locations and motion paths that hide their bodies from competitors (Whiten and Byrne, 1988; Hare et al., 2006). These behaviors are advantageous in potentially antagonistic and risky encounters with conspecifics or predators (Hare and Tomasello, 2004). They are evolutionarily adaptive for animals that are yoked much tighter into the here and now than humans and do not engage in shared intentionality and cooperation. We think that joint attention and cooperation bridge spatial distances between self and other and thus privilege social over visual perspective-taking. The competitive and individualistic mode of operating found in non-human primates, in contrast, makes an awareness of the visibility of resources and one's own body to others critical for survival.

Generally, visuo-spatial perspective-taking is seen as the most basic and embodied form of perspective-taking, that is expected to subserve and function as a model for more mental or higher-cognitive forms, such as imagining how others feel or think about

a certain situation (as is also suggested by spatial metaphors such as "putting oneself in another's position/shoes," see Kessler and Thomson, 2010). The genetic primacy of social over visual perspective-taking that we argued and provided empirical support for is at odds with this idea of visual perspective-taking as the cradle for other kinds of perspective-taking. Being aware of and responsive to others' literal viewpoints can certainly be key in social interaction. To communicate effectively we often have to adjust our speech and non-verbal behavior according to what the other sees or how he sees things—e.g., when we direct him to an object outside of his visual field or ask him to move his left shoulder that is on our right as we stand facing him. But getting a grip on others' visual perspectives takes time ontogenetically, and is not the first skill of its type to emerge.

We tried to show in this article that before children come to know what is seen from which particular viewpoint, they not only bridge perspectival differences in acts of joint attention and deictic reference by the age of 9–12 months, allowing them to create a common ground of shared experience with others. They also readily track and update what others have witnessed, done, and said. This knowledge is foundational for effective communication and other forms of cooperation, as it constitutes the background against which gestures and speech acts are understood and produced. We cited empirical evidence that children develop an awareness of visual perspectives somewhat later. This is not only suggested by the relatively late onset of visual perspective-taking, but is also reflected in children's aperspectival drawings and false perceptual judgments. In our attempt to explain the counterintuitive sequence from social to visual perspective-taking we highlighted the primary importance of forming experiential backgrounds with others for the sake of communication and cooperation. If the developmental trajectory that we traced is informative with regard to the relation between visual and social perspective-taking in cognitively mature human beings remains an open question.

REFERENCES

- Akhtar, N., Carpenter, M., and Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child Dev.* 67, 635–645. doi: 10.2307/1131837
- Arnheim, R. (1969). *Visual Thinking*. Berkeley: University of California Press.
- Arnheim, R. (1974). *Art and Visual Perception: a Psychology of the Creative Eye*. Berkeley, CA: University of California Press.
- Bernstein, D. M., Atance, C., Meltzoff, A., and Loftus, G. R. (2007). Hindsight bias and developing theories of mind. *Child Dev.* 78, 1374–1394. doi: 10.1111/j.1467-8624.2007.01071.x
- Bordeaux, M. A., and Willbrand, M. L. (1987). Pragmatic development in children's telephone discourse. *Discourse Process.* 10, 253–266. doi: 10.1080/01638538709544675
- Bremner, J. G., and Batten, A. (1991). Sensitivity to viewpoint in children's drawings of objects and relations between objects. *J. Exp. Child Psychol.* 52, 375–394. doi: 10.1016/0022-0965(91)90070-9
- Bühler, K. (1930). *The Mental Development of the Child*. London: Kegan Paul.
- Buttelmann, D., Carpenter, M., and Tomasello, M. (2009). Eighteen-month-olds show false belief understanding in an active helping paradigm. *Cognition* 112, 337–342. doi: 10.1016/j.cognition.2009.05.006
- Campbell, J. (2012). "An object-dependent perspective on joint attention," in *Joint Attention: New Developments in Psychology, Philosophy of Mind, and Social Neuroscience*, ed A. Seeman (Cambridge, MA: MIT Press), 415–430.
- Clark, H. H. (1996). *Using Language*. Cambridge, MA: Cambridge University Press. doi: 10.1017/CBO9780511620539
- Clark, H. H., and Marshall, C. R. (1981). "Definite reference and mutual knowledge," in *Elements of Discourse Understanding*, eds A. Joshi, B. Webber and I. Sag (Cambridge: Cambridge University Press), 10–63.
- Costall, A. (1997). Innocence and corruption: conflicting images of child art. *Hum. Dev.* 40, 133–144. doi: 10.1159/000278716
- Costall, A. (2001). *Georges-Henri Luquet, Children's Drawings*. London: Free Associations Press.
- Cox, M. V. (1991). *The Child's Point of View*. New York, NY: Guilford Press.
- Cox, M. V. (1993). *Children's Drawings of the Human Figure*. Hove: Erlbaum.
- Cox, M. V. (1997). *Drawings of People by the Under-5s*. New York, NY: Psychology Press.
- Cox, S. (2005). Intention and meaning in young children's drawing. *J. Art Des. Educ.* 24, 115–125. doi: 10.1111/j.1476-8070.2005.00432.x
- Csibra, G., and Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philos. Trans. R. Soc. B* 366, 1149–1157. doi: 10.1098/rstb.2010.0319
- Davis, A. M. (1983). Contextual sensitivity in young children's drawings. *J. Exp. Child Psychol.* 35, 478–486. doi: 10.1016/0022-0965(83)90022-X
- Deutsch, W., and Pechmann, T. (1982). Social interaction and the development of definite descriptions.

- Cognition* 11, 159–184. doi: 10.1016/0010-0277(82)90024-5
- Egyed, K., Király, I., and Gergely, G. (2007). “Understanding object-referential attitude expressions in 18-month-olds: the interpretation switching function of ostensive-communicative cues,” in *Poster Session Presented at the Biennial Meeting of the Society for Research in Child Development* (Boston).
- Eilan, N. (2005). “Joint attention, communication, and mind,” in *Joint Attention: Communication and Other Minds: Issues in Philosophy and Psychology*, eds C. Hoerl, T. McCormack, and J. Roessler (Oxford: Oxford University Press), 1–33. doi: 10.1093/acprof:oso/9780199245635.001.0001
- Epley, N., Morewedge, C. K., and Keysar, B. (2004). Perspective taking in children and adults: equivalent egocentrism but differential correction. *J. Exp. Soc. Psychol.* 40, 760–768. doi: 10.1016/j.jesp.2004.02.002
- Fishbein, H. D., Lewis, S., and Keiffer, K. (1972). Children’s understanding of spatial relations. *Dev. Psychol.* 7, 21–33. doi: 10.1037/h0032858
- Flavell, J. H. (1992). “Perspectives on perspective-taking,” in *Piaget’s Theory: Prospects and Possibilities*, eds H. Beilin and P. Pufall (Hillsdale, NJ: Erlbaum), 107–139.
- Flavell, J. H., Shipstead, S. G., and Croft, K. (1978). Young children’s knowledge about visual perception: hiding objects from others. *Child Dev.* 49, 1208–1211. doi: 10.2307/1128761
- Gablik, S. (1977). *Progress in Art*. New York, NY: Rizzoli.
- Ganea, P. A., and Saylor, M. M. (2007). Infants’ use of shared linguistic information to clarify ambiguous requests. *Child Dev.* 78, 493–502. doi: 10.1111/j.1467-8624.2007.01011.x
- Gergely, G., Egyed, K., and Király, I. (2007). On pedagogy. *Dev. Sci.* 10, 139–146. doi: 10.1111/j.1467-7687.2007.00576.x
- Gintis, H., Bowles, S., Boyd, R., and Fehr, E. (2003). Explaining altruistic behavior in humans. *Evol. Hum. Behav.* 24, 153–172. doi: 10.1016/S1090-5138(02)00157-5
- Glucksberg, S., and Krauss, R. M. (1967). What do people say after they have learned how to talk? Studies of the development of referential communication. *Merrill Palmer Q. Behav. Dev.* 13, 309–316.
- Golomb, C. (2004). *The Child’s Creation of a Pictorial World*. New York, NY: Psychology Press.
- Hamilton, A., Brindley, R., and Frith, U. (2009). Visual perspective taking impairment in children with autistic spectrum disorder. *Cognition* 113, 37–44. doi: 10.1016/j.cognition.2009.07.007
- Hare, B., and Tomasello, M. (2004). Chimpanzees are more skillful in competitive than in cooperative cognitive tasks. *Anim. Behav.* 68, 571–581. doi: 10.1016/j.anbehav.2003.11.011
- Hare, B., Call, J., Agnetta, B., and Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Anim. Behav.* 59, 771–785. doi: 10.1006/anbe.1999.1377
- Hare, B., Call, J., and Tomasello, M. (2006). Chimpanzees deceive a human by hiding. *Cognition* 101, 495–514. doi: 10.1016/j.cognition.2005.01.011
- Hobson, R. P. (1984). Early childhood autism and the question of egocentrism. *J. Autism Dev. Disord.* 14, 85–104. doi: 10.1007/BF02408558
- Hughes, M., and Donaldson, M. (1979). The use of hiding games for studying the coordination of viewpoints. *Educ. Rev.* 31, 133–140. doi: 10.1080/0013191790310207
- Kessler, K., and Rutherford, H. (2010). The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions. *Front. Psychol.* 1:213. doi: 10.3389/fpsy.2010.00213
- Kessler, K., and Thomson, L. A. (2010). The embodied nature of spatial perspective taking: embodied transformation versus sensorimotor interference. *Cognition* 114, 72–88. doi: 10.1016/j.cognition.2009.08.015
- Keysar, B. (2007). Communication and miscommunication: the role of egocentric processes. *Intercultural Pragmatics* 4, 71–84. doi: 10.1515/IP.2007.004
- Keysar, B., and Henly, A. S. (2002). Speakers’ overestimation of their effectiveness. *Psychol. Sci.* 13, 207–212. doi: 10.1111/1467-9280.00439
- Keysar, B., Lin, S., and Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition* 89, 29–41. doi: 10.1016/S0010-0277(03)00064-7
- Kovács, A. M., Téglás, E., and Endress, A. D. (2010). The social sense: susceptibility to others’ beliefs in human infants and adults. *Science* 330, 1830–1834. doi: 10.1126/science.1190792
- Lark-Horovitz, B., Barnhart, E. N., and Sills, E. M. (1939). *Graphic Work-Sample Diagnosis: an Analytical Method of Estimating Children’s Drawing Ability*. Cleveland, OH: Cleveland Museum of Art.
- Lempers, J. D., Flavell, E. R., and Flavell, J. H. (1977). The development in very young children of tacit knowledge concerning visual perception. *Genet. Psychol. Monogr.* 93, 3–53.
- Liben, L. S., and Belknap, B. (1981). Intellectual realism: implications for investigations of perspective-taking in young children. *Child Dev.* 52, 921–924. doi: 10.2307/1129095
- Liebal, K., Behne, T., Carpenter, M., and Tomasello, M. (2009). Infants use shared experience to interpret pointing gestures. *Dev. Sci.* 12, 264–271. doi: 10.1111/j.1467-7687.2008.00758.x
- Liebal, K., Carpenter, M., and Tomasello, M. (2011). Young children’s understanding of markedness in nonverbal communication. *J. Child Lang.* 38, 888–903. doi: 10.1017/S0305000910000383
- Light, P., and Nix, C. (1983). Own view versus good view in a perspective-taking task. *Child Dev.* 54, 480–483. doi: 10.2307/1129709
- Low, J., and Perner, J. (2012). Editorial—implicit and explicit theory of mind: state of the art. *Br. J. Dev. Psychol.* 30, 1–13. doi: 10.1111/j.2044-835X.2011.02074.x
- Luo, Y., and Baillargeon, R. (2007). Do 12.5-month-old infants consider what objects others can see when interpreting their actions? *Cognition* 105, 489–512. doi: 10.1016/j.cognition.2006.10.007
- Luo, Y., and Beck, W. (2010). Do you see what I see? Infants’ reasoning about others’ incomplete perceptions. *Dev. Sci.* 13, 134–142. doi: 10.1111/j.1467-7687.2009.00863.x
- Luquet, G. H. (1927). *Children’s Drawings (‘Le Dessin Enfantin’)*. A. Costall, Transl. London: Free Association Books.
- MacMurray, J. (1961). *Persons in Relation*. London: Faber and Faber.
- MacPherson, A. C., and Moore, C. (2010). Understanding interest in the second year of life. *Infancy* 15, 324–335. doi: 10.1111/j.1532-7078.2009.00011.x
- Maitland, L. M. (1895). What children draw to please themselves. *Inland Educ.* 1, 77–81.
- Maratsos, M. P. (1976). *The Use of Definite and Indefinite Reference in Young Children: an Experimental Study of Semantic Acquisition*. Cambridge, UK: Cambridge University Press.
- Masangkay, Z. S., McCluskey, K. A., McIntyre, C. W., Sims-Knight, J., Vaughn, B. E., and Flavell, J. H. (1974). The early development of inferences about the visual precepts of others. *Child Dev.* 45, 357–366. doi: 10.2307/1127956
- Matisse, H. (1953). Looking at life with the eyes of a child. *Art Educ.* 6, 5–6. doi: 10.2307/3183775
- Matthews, D., Lieven, E., Theakston, A., and Tomasello, M. (2006). The effect of perceptual availability and prior discourse on young children’s use of referring expressions. *Appl. Psycholinguist.* 27, 403–422. doi: 10.1017/S0142716406060334
- McGuigan, N., and Doherty, M. J. (2002). The relation between hiding skill and judgment of eye direction in preschool children. *Dev. Psychol.* 38, 418–427. doi: 10.1037/0012-1649.38.3.418
- Moll, H., Carpenter, M., and Tomasello, M. (2007). Fourteen-month-old infants know what others experience only in joint engagement with them. *Dev. Sci.* 10, 826–835. doi: 10.1111/j.1467-7687.2007.00615.x
- Moll, H., Carpenter, M., and Tomasello, T. (2010). Social engagement leads 2-year-olds to overestimate others. *Infancy* 16, 248–265. doi: 10.1111/j.1532-7078.2010.00044.x
- Moll, H., and Meltzoff, A. N. (2012). “Perspective-taking and its foundation in joint attention,” in *Joint Attention: New Developments in Psychology, Philosophy of Mind, and Social Neuroscience*, ed A. Seeman (Cambridge, MA: MIT Press), 393–414.
- Moll, H., and Tomasello, M. (2006). Level 1 perspective-taking at 24 months of age. *Br. J. Dev. Psychol.* 24, 603–613. doi: 10.1348/026151005X55370
- Moll, H., and Tomasello, M. (2007). How 14- and 18-month-olds know what others have experienced. *Dev. Psychol.* 43, 309–317. doi: 10.1037/0012-1649.43.2.309
- Moll, H., Richter, N., Carpenter, M., and Tomasello, M. (2008). Fourteen-month-olds know what ‘we’ have shared in a special way. *Infancy* 13, 90–101. doi: 10.1080/15250000701779402
- Nayer, S. L., and Graham, S. A. (2006). Children’s communicative strategies in novel and familiar word situations. *First Lang.* 26, 403–420. doi: 10.1177/0142723706064834
- Nurmsoo, E., and Bloom, P. (2008). Preschoolers’ perspective taking in word learning: do they blindly follow eye gaze? *Psychol. Sci.* 19, 211–215. doi: 10.1111/j.1467-9280.2008.02069.x
- Olson, D. R., and Torrance, N. G. (1987). Language, literacy, and mental states. *Discourse Process.* 10, 157–168. doi: 10.1080/01638538709544667

- O'Neill, D. K., and Holmes, A. (2002). Young preschoolers' ability to reference story characters: the contribution of gesture and character speech. *First Lang.* 22, 73–103.
- O'Neill, D. K. (1996). Two-year-old children's sensitivity to a parent's knowledge state when making requests. *Child Dev.* 67, 659–677. doi: 10.2307/1131839
- Onishi, K. H., and Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science* 308, 255–258. doi: 10.1126/science.1107621
- Panofsky, E. (1927). *Perspective as Symbolic Form*. Cambridge, MA: MIT Press.
- Perner, J., Brandl, J. L., and Garnham, A. (2003). What is a perspective problem? Developmental issues in belief ascription and dual identity. *Facta Philos.* 5, 355–378.
- Perner, J., and Roessler, J. (2012). From infants' to children's appreciation of belief. *Trends Cogn. Sci.* 16, 519–525. doi: 10.1016/j.tics.2012.08.004
- Perner, J., and Ruffman, T. (2005). Infants' insight into the mind: how deep? *Science* 308, 214–216. doi: 10.1126/science.1111656
- Perner, J., Zauner, P., and Sprung, M. (2005). "What does 'that' have to do with points of view? Conflicting desires and 'want' in German," in *Why Language Matters for Theory of Mind*, eds J. Astington and J. Baird (Oxford: Oxford University Press), 220–244.
- Piaget, J. (1929). *The Child's Conception of the World*. London: Routledge and Kegan Paul.
- Piaget, J. (1955). *The Child's Construction of Reality*. London: Routledge and Kegan Paul.
- Power, R., and dal Martello, M. F. (1986). The use of the definite and indefinite articles by Italian preschool children. *J. Child Lang.* 13, 145–154. doi: 10.1017/S0305000900000350
- Qureshi, A. W., Apperly, I. A., and Samson, D. (2010). Executive function is necessary for perspective selection, not level-1 visual perspective selection: evidence from a dual-task study of adults. *Cognition* 117, 230–236. doi: 10.1016/j.cognition.2010.08.003
- Reed, T. (2002). Visual perspective taking as a measure of working memory in participants with autism. *J. Dev. Phys. Disabil.* 14, 63–76. doi: 10.1023/A:1013515829985
- Repacholi, B. M., and Gopnik, A. (1997). Early reasoning about desires: evidence from 14- and 18-month-olds. *Dev. Psychol.* 33, 12–21. doi: 10.1037/0012-1649.33.1.12
- Repacholi, B. M., and Meltzoff, A. N. (2007). Emotional eavesdropping: infants selectively respond to indirect emotional signals. *Child Dev.* 78, 503–521. doi: 10.1111/j.1467-8624.2007.01012.x
- Salomo, D., Lieven, E., and Tomasello, M. (2010). Young children's sensitivity to new and given information when answering predicate-focus questions. *Appl. Psycholinguistics* 31, 101–115. doi: 10.1017/S014271640999018X
- Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., and Bodley Scott, S. E. (2010). Seeing it their way: what other people see is calculated by low-level and early acting processes. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 1255–1266.
- Saylor, M. M., and Ganea, P. A. (2007). Infants interpret ambiguous requests for absent objects. *Dev. Psychol.* 43, 696–704. doi: 10.1037/0012-1649.43.3.696
- Schiffer, S. (1972). *Meaning*. Oxford: Oxford University Press.
- Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition* 47, 1–24. doi: 10.1016/0010-0277(93)90060-9
- Schütz, A. (1932). *Der Sinnhafte Aufbau der Sozialen Welt: Eine Einleitung in die Verstehende Soziologie*. Vienna: Springer. doi: 10.1007/978-3-7091-3108-4
- Seemann, A. (2012). "Joint attention: towards a relational account," in *Joint Attention: New Developments in Psychology, Philosophy of Mind, and Social Neuroscience*, ed A. Seeman (Cambridge, MA: MIT Press), 183–202.
- Serratrice, L. (2005). The role of discourse pragmatics in the acquisition of subjects in Italian. *Appl. Psycholinguist.* 26, 437–462. doi: 10.1017/S0142716405050241
- Sodian, B., Thoermer, C., and Metz, U. (2007). Now I see it but you don't: 14-month-olds can represent another person's visual perspective. *Dev. Sci.* 10, 199–204. doi: 10.1111/j.1467-7687.2007.00580.x
- Sonnenschein, S., and Whitehurst, G. J. (1984). Developing referential communication: a hierarchy of skills. *Child Dev.* 55, 1936–1945. doi: 10.2307/1129940
- Sully, J. (1895). *Studies of Childhood*. New York, NY: D. Appleton and Co. doi: 10.1037/11376-000
- Surian, L., Caldi, S., and Sperber, D. (2007). Attribution of beliefs to 13-month-old infants. *Psychol. Sci.* 18, 580–586. doi: 10.1111/j.1467-9280.2007.01943.x
- Surtees, A. D., and Apperly, I. A. (2012). Egocentrism and automatic perspective taking in children and adults. *Child Dev.* 83, 452–460. doi: 10.1111/j.1467-8624.2011.01730.x
- Surtees, A. D., Noordzij, M. L., and Apperly, I. A. (2012). Sometimes losing your self in space: children's and adults' spontaneous use of multiple spatial reference frames. *Dev. Psychol.* 48, 185–191. doi: 10.1037/a0025863
- Tomasello, M. (2008). *Origins of Human Communication*. Cambridge, MA: MIT Press.
- Tomasello, M., and Haberl, K. (2003). Understanding attention: 12- and 18-month-olds know what's new for other persons. *Dev. Psychol.* 39, 906–912. doi: 10.1037/0012-1649.39.5.906
- Warren, A., and Tate, C. (1992). "Egocentrism in children's telephone conversations," in *Private Speech: From Social Interaction to Self-Regulation*, eds R. Diaz and L. Berk (Hillsdale, NJ: Lawrence Erlbaum), 245–264.
- Whiten, A., and Byrne, R. W. (1988). Tactical deception in primates. *Behav. Brain Sci.* 11, 233–244. doi: 10.1017/S0140525X00049682
- Wittgenstein, L. (2001). *Philosophical Investigations*. Oxford: Wiley-Blackwell.
- Wu, S., and Keysar, B. (2007). The effect of communication overlap on communication effectiveness. *Cogn. Sci.* 31, 169–181. doi: 10.1080/03640210709336989
- Yaniv, I., and Shatz, M. (1990). Heuristics of reasoning and analogy in children's visual perspective taking. *Child Dev.* 61, 1491–1501. doi: 10.2307/1130758
- Yu, Y., Su, Y., and Chan, R. (2011). "The relationship between visual perspective taking and imitation impairments in children with autism," in *A Comprehensive Book on Autism Spectrum Disorders*, ed M. Mohammadi (Rijeka: InTech), 369–384.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 03 June 2013; accepted: 22 August 2013; published online: 10 September 2013.

Citation: Moll H and Kadipasaoglu D (2013) The primacy of social over visual perspective-taking. *Front. Hum. Neurosci.* 7:558. doi: 10.3389/fnhum.2013.00558

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2013 Moll and Kadipasaoglu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Old cortex, new contexts: re-purposing spatial perception for social cognition

Carolyn Parkinson* and Thalia Wheatley

Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH, USA

Edited by:

Antonia Hamilton, University of Nottingham, UK

Reviewed by:

Ivan Toni, Radboud University, Netherlands

Anna M. Borghi, Institute of Cognitive Sciences and Technologies, University of Bologna, Rome, Italy

*Correspondence:

Carolyn Parkinson, Department of Psychological and Brain Sciences, Dartmouth College, 6207 Moore Hall, Hanover, NH 03755, USA
e-mail: carolyn.parkinson@gmail.com

Much of everyday mental life involves information that we cannot currently perceive directly, from contemplating the strengths of friendships to reasoning about the contents of other minds. Despite their primacy to everyday human functioning, and in particular, to human sociality, the mechanisms that support abstract thought are poorly understood. An explanatory framework that has gained traction recently in cognitive neuroscience is exaptation, or the re-purposing of evolutionarily old circuitry to carry out new functions. We argue for the utility of applying this concept to social cognition. Convergent behavioral and neuroscientific evidence suggests that humans co-opt mechanisms originally devoted to spatial perception for more abstract domains of cognition (e.g., temporal reasoning). Preliminary evidence suggests that some aspects of social cognition also involve the exaptation of substrates originally evolved for processing physical space. We discuss the potential for future work to test more directly if cortical substrates for spatial processing were exapted for social cognition, and in so doing, to improve our understanding of how humans evolved mechanisms for navigating an exceptionally complex social world.

Keywords: exaptation, neural reuse, social neuroscience, metaphor, spatial cognition, perspective taking, social distance, posterior parietal cortex

EXAPTATION AND HUMAN COGNITION

Our thoughts often include information outside of the current sensory environment, from imagined futures to the contents of other minds. However, the mechanisms supporting abstract cognition remain poorly understood. An explanatory framework that has gained traction recently (Gallese and Lakoff, 2005; Dehaene and Cohen, 2007; Anderson, 2010) involves exaptation: co-opting existing morphological features for novel functions (Gould and Vrba, 1982). New cognitive capacities may have emerged over the course of evolution when brain regions originally devoted to specific functions were repurposed and recombined in novel ways to process additional kinds of information (Anderson, 2010). Analogous cortical recycling processes may occur during development whereby cultural inventions co-opt circuitry evolved for older aspects of cognition (Dehaene and Cohen, 2007). Exaptation and cortical recycling provide plausible neural bases for proposals that representational resources originally devoted to space were co-opted to process more abstract information (e.g., Boroditsky, 2011).

If our ability to reason about abstract concepts resulted from evolutionary “tinkering” (Jacob, 1977) with neural mechanisms originally developed for operating on physical space, then demanding that these mechanisms handle conflicting inputs pertaining to their new and old functions simultaneously should create response conflict. Further, evidence from clinical and neuroimaging studies should suggest shared substrates for these functions. Both kinds of evidence are accumulating with respect to several domains of abstract cognition, most widely in studies relating temporal and numerical processing to spatial cognition (Hubbard et al., 2005; Bonato et al., 2012). Here, we highlight the potential for our understanding of the mechanisms underlying

abstract social cognition to benefit from a similar approach, and review evidence that these mechanisms may be best understood in terms of the kinds of computations (e.g., distance judgments, perspective taking), rather than the domains of knowledge, that they involve.

SOCIALITY AND HUMAN BRAIN EVOLUTION

Humans come into the world seemingly hardwired to detect and connect with other minds (Wheatley et al., 2012), and maintaining this predisposition is closely tied to healthy development (Pavlova, 2012). Effectively perceiving and interpreting social cues is particularly crucial for humans, who must navigate an exceptionally flexible system of relationships with conspecifics (Fiske, 1991). The ability to meet the intensive computational demands of humans’ complex social environment (e.g., forging alliances, sharing intentions, tactical deception; Harcourt, 1988, 1989; Tomasello et al., 2005) is thought to have been a driving force for cortical expansion during evolution (Dunbar, 1998). Consistent with this hypothesis, feats of social cognition presumed to be uniquely human, such as sharing intentions (Tomasello et al., 2005) and representing others’ beliefs (representational theory of mind, RTOM; Call and Tomasello, 2008) involve cortical areas that underwent the most evolutionary expansion (e.g., lateral posterior parietal cortex, PPC; Van Essen et al., 2001; particularly the temporoparietal junction, TPJ; Saxe, 2006; Redcay et al., 2010). These aspects of social cognition (e.g., RTOM) tend to involve information that cannot be perceived directly (e.g., false beliefs), and are functionally (Gobbini et al., 2007) and structurally (Parkinson and Wheatley, 2012) dissociable from older social processes (e.g., motor resonance), suggesting they either involve entirely new structures or structures previously devoted to non-social functions. Gould and Vrba

(1982) suggested that exaptation of complex traits would likely be followed by secondary adaptations to further support new functions. Consistent with PPC circuitry evolved for dealing with space having been exapted, then expanded, to support abstract social cognition, the PPC has an evolutionarily old role in spatial perception (it encodes space in our distant relatives, e.g., rats; Nitz, 2006), processes both social and spatial information in humans and other primates (Yamazaki et al., 2009), and has expanded (Van Essen et al., 2001) and formed new connections (Mantini et al., 2013) in humans as it came to support evermore abstract aspects of social cognition. Recent computational modeling experiments support the notion that human brain expansion was driven by the cognitive demands of human sociality (Dávid-Barrett and Dunbar, 2013). Importantly, large brain size has a great metabolic cost; the human brain accounts for 2% of body mass but requires 20% of the energy that we consume (Clark and Sokoloff, 1999). In order to outweigh the considerable metabolic cost of the larger brain that they require, the cognitive mechanisms supporting human sociality must have conferred substantial adaptive benefits.

However, compared to other domains of abstract cognition (e.g., mathematics; Hubbard et al., 2005), little is known about how social forms of abstract cognition (e.g., representing beliefs or one's place in a social network) relate to evolutionarily older aspects of cognition. This may be due to several factors. First, compared to cognitive neuroscience, social cognitive neuroscience is a young field (Ochsner, 2007); many aspects of social cognition have simply been studied less extensively than other aspects of cognition. Second, early social cognitive neuroscience research often assumed a modular view of the brain (Bergeron, 2007), and involved searching for encapsulated brain areas devoted to processing particular contents (Kihlstrom, 2010). If one understands an aspect of cognition to be supported by a domain-specific module, attempting to relate that aspect of cognition to other mental phenomena may not be considered a particularly worthwhile endeavor. More recently, brain areas (e.g., TPJ; fusiform face area) previously implicated in various facets of social information processing (e.g., RTOM; face perception) have been found to perform similar operations (e.g., reorienting attention, Mitchell, 2008; visual object encoding, Hanson et al., 2004) on diverse contents. Consistent with the suggestion that social cognition and physical perception involve common computations (Zaki, 2013), the functional significance of brain areas involved in social cognition may often be best characterized in terms of the operations they perform across multiple domains of information.

LINGUISTIC MAPPINGS BETWEEN ABSTRACT COGNITION AND SPATIAL PERCEPTION

One window into the cognitive operations supporting abstract thought is the language we use to describe them (Lakoff and Johnson, 1980). The spatialization of form hypothesis (Lakoff, 1987) specifically highlights the widespread use of spatial words (e.g., “outside,” “far”) to describe conceptual relations, suggesting that spatial schemata structure mental representations. Abstract relations may be represented in terms of space because unlike spatial relationships, they must be imagined rather than observed (Evans, 2006). We can observe two people sitting close together, gaze direction, or moving a vehicle forward, but can only imagine

the closeness of a friendship, a belief, or moving a meeting forward (Casasanto et al., 2010). In this view, phrases like “close friendship” or “far from the truth” are not mere figures of speech, but rather, figures of thought that reveal the structure of mental representations (Lakoff, 1986). The extent to which representational overlap between space and abstract domains results from exaptation during evolution, metaphoric structuring acquired during development, or some combination of these processes, remains an open question. With respect to social processing, the recruitment of brain areas involved in reorienting visual attention (TPJ) while congenitally blind individuals perform RTOM tasks (Bedny et al., 2009) suggests that functional overlap between social and visuospatial processes may be an innately predisposed result of evolutionary exaptation that is now reflected in linguistic metaphors for mentalizing (e.g., “Try to see things from my point of view”).

The domain of abstract cognition that has been studied most extensively in terms its relation to space is time. Cross-linguistic studies indicate that people around the world use spatial language to describe time (Boroditsky, 2011); the intuition to represent time analogously to space may be evolutionarily predisposed. Do all languages employ spatial language to describe social relationships (e.g., “close friend”) and RTOM? Are mappings consistent across languages? Some cross-linguistic variability exists in spatiotemporal metaphors, but certain mappings (future = forward) are nearly ubiquitous, likely due to shared aspects of human physiology and experience. Similarly, some English spatial metaphors for social relationships (familiarity = closeness) may stem from the tendency to give personal space to others based on the “closeness” of relationships (Hayduk, 1983). To our knowledge, metaphoric mappings between spatial and social relationships or between visuospatial and social perspective taking have not been subjected to exhaustive cross-linguistic analysis. Thus, whether or not humans around the world use space to structure mental representations of the magnitude and traversal of social distances remains an open question.

BEHAVIORAL EVIDENCE FOR MAPPINGS BETWEEN ABSTRACT COGNITION AND SPATIAL PERCEPTION

Behavioral mappings between space and abstract cognition have been most extensively studied with respect to time and number. Number and space are associated implicitly; according to the spatial numerical association of response codes (SNARC) effect, people are faster to respond regarding small numbers on the left side of space, and large numbers on the right side of space, even for tasks unrelated to magnitude (Dehaene et al., 1993). Similar associations have been documented between number and elevation (Pecher and Boot, 2011; Lugli et al., 2013). Representational overlap between space and number appears to comprise a universal human intuition (Dehaene et al., 2008), and can be documented outside of the laboratory. When thinking about numbers, more than 10% of individuals report automatically accessing mental “number forms” consisting of spatial layouts (Seron et al., 1992). It has even been suggested that on the scale of motoric action, time, space, and quantity are processed by an analog magnitude system (Walsh, 2003), which was co-opted to process discrete number (Bueti and Walsh, 2009).

Stimulus-response compatibility codes also exist for time (Ishihara et al., 2008; Sell and Kaschak, 2011). Additionally, people tend to spontaneously sway forward while imagining the future and backward while imagining the past, suggesting that representations of movement through space are automatically activated during imagined movement through time (Miles et al., 2010). Monkeys (Merritt et al., 2010) and infants without exposure to relevant linguistic or sensorimotor mappings (Srinivasan and Carey, 2010) exhibit representational overlap between spatial extent and temporal duration (but not all magnitudes), suggesting that spatiotemporal mappings originate from common processing mechanisms, independently of sensorimotor grounding or linguistic correspondences.

Social and spatial information are also behaviorally associated. Visual perspective taking and mentalizing abilities are positively correlated (Flavell et al., 1986; Hamilton et al., 2009). People readily convert judgments of social compatibility into physical distances (Yamakawa et al., 2009). Words characterizing close social distances (e.g., “us,” “friend”) are associated with close locations, and words characterizing remote social distances (e.g., “them,” “enemy”) are associated with far spatial locations (Bar-Anan et al., 2007). Additionally, consistent with the suggestion that out-group members are construed as being physically distant from oneself (except following threat, Xiao and Van Bavel, 2012), Jones et al. (1981) found that out-group members are rated as more homogeneous (i.e., having a narrower range of personal characteristics) than in-group members. Similarly, powerful individuals, who see themselves as exceptionally distinctive, construe others as exceptionally distant and homogeneous (Fiske, 1993; Lee and Tiedens, 2001). It may be parsimonious to represent social and spatial distances analogously: Construal level theory of psychological distance (Liberman and Trope, 2008) posits that spatial, temporal and social egocentric distance share a common psychological meaning – distance from the self in the here and now.

Although extant research highlights a possible relationship between mental representations of social and spatial information, more research is needed to explore this possibility, and address several remaining questions, such as: is there a hierarchy of egocentric psychological distance domains, in which some are more primary than others? Do we spontaneously access representations of moving through space when traversing “social” distances or perspective taking, like during mental time travel? Are spatial representations activated explicitly when thinking about social relationships in everyday life, as they are for many individuals when thinking about numbers? Exploring questions like these will lead to an improved understanding of the mechanisms involved in abstract social cognition.

NEUROSCIENTIFIC EVIDENCE FOR MAPPINGS BETWEEN ABSTRACT COGNITION AND SPATIAL PERCEPTION

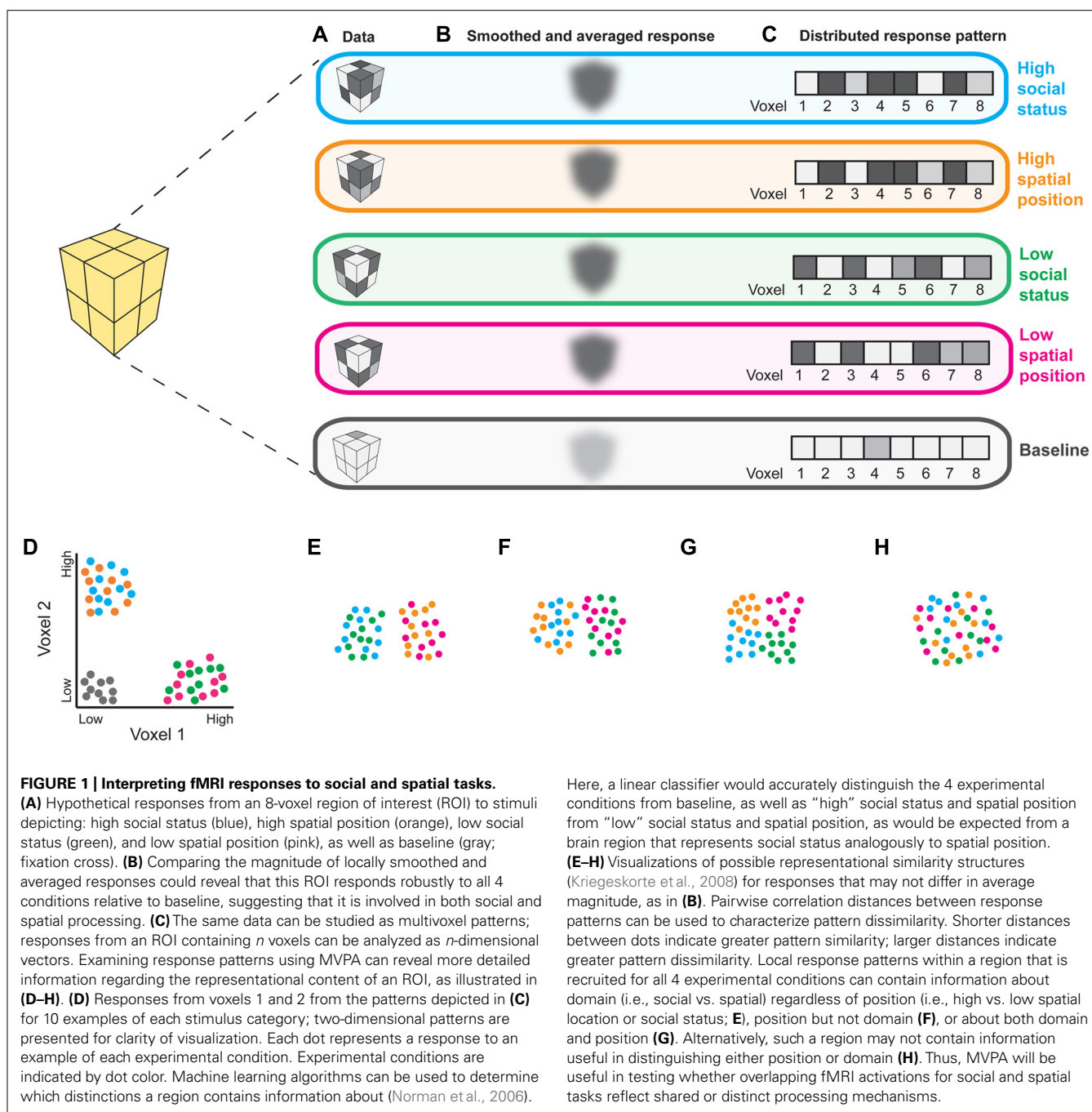
If spatial processing were repurposed for abstract cognition, one would expect overlapping neural substrates. Past research suggests that PPC systems for sensorimotor control and cognition largely overlap (Creem-Regehr, 2009). As the PPC expanded in size over the course of human evolution (Van Essen et al., 2001), it appears to have expanded in function as well, leading to suggestions that mechanisms originally devoted to representing

peripersonal space were repurposed to perform analogous operations on new contents. According to this theory, mechanisms previously dedicated to representing spatial information about the current sensory environment were first co-opted to represent simulations of peripersonal space in the past and future to support episodic memory and prospection, and later, to represent information in increasingly abstract frames of reference (Yamazaki et al., 2009). A growing body of neuroimaging and neuropsychological evidence suggests that representations of spatial and abstract information, including aspects of social cognition, are associated in the PPC.

Functional magnetic resonance imaging (fMRI) studies in humans implicate the PPC in representing perceptual, temporal, social and conceptual frames of reference (Yamazaki et al., 2009). Importantly, most of these results are based on overlapping activations from univariate contrasts, which could reflect shared neural codes or nearby but distinct codes for different kinds of information (Peelen and Downing, 2007). Multivariate pattern analysis (MVPA), which compares distributed patterns of activity between experimental conditions, rather than regionally smoothed and averaged responses, may better characterize brain regions’ representational contents (**Figure 1**). The few studies that have used MVPA to compare spatial and abstract cognition in the PPC support the suggestion that representations of spatial information “scaffold” those of more abstract information. A pattern classifier trained only to distinguish PPC responses to leftward vs. rightward saccades can distinguish mental addition from subtraction (Knops et al., 2009). Additionally, position and valence words can be decoded by a classifier trained only on patterns of PPC activity corresponding to visual elevation (Quadflieg et al., 2011). Because MVPA can reveal information about underlying cognitive structures (**Figure 1**), this approach will be valuable in elucidating whether use of spatial language in describing abstract social concepts reflects true representational similarities or linguistic bottlenecks that push people to use metaphors in the absence of adequate domain-specific terminology (such bottlenecks have been demonstrated in olfaction; Yeshurun and Sobel, 2010).

Further research is needed to characterize the relationship between the PPC’s involvement in social and spatial cognition. For instance, the TPJ is recruited both when subjects reason about others’ false beliefs and positions in space (Abraham et al., 2008), suggesting that this region may perform similar computations on visuospatial and social contents. MVPA could be used to more directly test this possibility. Similarly, judgments about hierarchy and social distance recruit areas of the PPC involved in self-referential physical distance processing (Chiao et al., 2009; Yamakawa et al., 2009). Does this brain region represent “high” social status and “close” social distances analogously to how it represents “high” spatial location and “close” spatial distances? Again, characterizing the representational structure of the PPC with MVPA could elucidate this question (**Figure 1**).

Neuropsychological data also suggests a close relationship between representations of spatial and abstract information in the PPC. Patients with left hemineglect following right PPC damage often also neglect the “left” side of the mental number line (Zorzi et al., 2002), whereas PPC lesion patients without neglect show no numerical deficits (Vuilleumier et al., 2004). Remarkably,



normal numerical processing is restored in neglect patients following interventions utilizing adaptation to leftward-shifting prism glasses that restore visual attention to the previously neglected side of space (Rossetti et al., 2004). Patients with hemispatial neglect exhibit analogous distortions of temporal processing, systematically overestimating temporal durations (Basso et al., 1996; Calabria et al., 2011). Spatiotemporal mappings appear to be supported by the PPC in healthy individuals, as they are diminished following transcranial magnetic stimulation to this region (Oliveri et al., 2009). Neuropsychological studies relating spatial and abstract cognition have focused primarily on non-social domains

of abstract cognition (e.g., time, number) and space. However, Samson et al. (2004) reported impaired mentalizing in patients with focal lesions to the inferior PPC. To our knowledge, no studies have tested if PPC damage is associated with abnormal representations of one's social network.

One limitation of neuroscientific evidence relating space and other domains of cognition is that data are available only from individuals in industrialized societies, and many of the corresponding behavioral phenomena are malleable to cultural learning (Dehaene et al., 1993). Although the tendency to map various domains of knowledge onto spatial representations appears to

comprise a universal intuition (Dehaene et al., 2008; Parkinson et al., 2012), the nature of these mappings is often subject to cultural variation (Hung et al., 2008; Boroditsky and Gaby, 2010). Even two weeks of tool use engenders white and gray matter changes in the macaque PPC (Hubbard et al., 2005; Iriki, 2005). Lifelong immersion in cultures emphasizing metaphors and analogical reasoning no doubt impacts neural representations. Although the work summarized here is drawn from studies conducted in several countries, more cross-cultural work, especially that involving direct cross-cultural comparisons, is required to better understand how representational overlap between spatial and social cognition arises in the brain.

COMPARING SPATIAL REPRESENTATIONS BETWEEN DOMAINS OF KNOWLEDGE

Importantly, although multiple domains of abstract cognition appear to co-opt mechanisms for spatial processing, different exaptations could have arisen separately, and may operate differently. There is a paucity of research investigating how different domains of knowledge that use space as a “reference domain” relate to one another. Different processes may have independently come to co-opt circuitry originally for spatial computations because such an arrangement was efficient and likely given pre-existing anatomical and functional constraints (Cantlon et al., 2009). Consistent with this suggestion,

a recent study comparing spatial representations of number and pitch within individuals suggests that spatial representations are idiosyncratic to specific domains of knowledge (Beecham et al., 2009). Thus, although past work relating spatial cognition to non-social aspects of abstract cognition will be informative for future studies aimed at characterizing the relationship between spatial perception and social cognition, this will not be a trivial endeavor.

CONCLUSION

Convergent evidence from behavior, neuropsychology, and neuroimaging suggest that humans use knowledge about space to scaffold mental representations of abstract information. Whereas most investigations have focused on non-social domains of abstract cognition, less work has explored the relationship between abstract aspects of social cognition (e.g., social distance evaluation, mentalizing) and spatial perception. Given the substantial progress that has stemmed from using this approach to characterize the mechanisms that support non-social domains of abstract cognition, we predict that relating abstract social cognition to spatial perception will be similarly fruitful. Further, given the centrality of sociality to human health and brain evolution (Dunbar, 1998), better understanding the mechanisms involved in social cognition is essential to understanding the human brain more generally.

REFERENCES

- Abraham, A., Werning, M., Rakoczy, H., von Cramon, D. Y., and Schubotz, R. I. (2008). Minds, persons, and space: an fMRI investigation into the relational complexity of higher-order intentionality. *Conscious. Cogn.* 17, 438–450. doi: 10.1016/j.concog.2008.03.011
- Anderson, M. L. (2010). Neural reuse: a fundamental organizational principle of the brain. *Behav. Brain Sci.* 33, 245–266. doi: 10.1017/S0140525X10000853
- Bar-Anan, Y., Liberman, N., Trope, Y., and Algom, D. (2007). Automatic processing of psychological distance: evidence from a Stroop task. *J. Exp. Psychol. Gen.* 136, 610–622. doi: 10.1037/0096-3445.136.4.610
- Basso, G., Nichelli, P., Frassinetti, F., and di Pellegrino, G. (1996). Time perception in a neglected space. *Neuroreport* 7, 2111–2114. doi: 10.1097/00001756-199609020-00009
- Bedny, M., Pascual-Leone, A., and Saxe, R. R. (2009). Growing up blind does not change the neural bases of theory of mind. *Proc. Natl. Acad. Sci. U.S.A.* 106, 11312–11317. doi: 10.1073/pnas.0900010106
- Beecham, R., Reeve, R. A., and Wilson, S. J. (2009). Spatial representations are specific to different domains of knowledge. *PLoS ONE* 4:e5543. doi: 10.1371/journal.pone.0005543
- Bergeron, V. (2007). Anatomical and functional modularity in cognitive science: shifting the focus. *Philos. Psychol.* 20, 175–195. doi: 10.1080/09515080701197155
- Bonato, M., Zorzi, M., and Umiltà, C. (2012). When time is space: evidence for a mental time line. *Neurosci. Biobehav. Rev.* 36, 2257–2273. doi: 10.1016/j.neubiorev.2012.08.007
- Boroditsky, L. (2011). “How languages construct time,” in *Space, Time and Number in the Brain*, eds S. Dehaene and E. Brannon (London: Elsevier), 333–341.
- Boroditsky, L., and Gaby, A. (2010). Remembrances of times East: absolute spatial representations of time in an Australian Aboriginal community. *Psychol. Sci.* 21, 1635–1639. doi: 10.1177/0956797610386621
- Bueti, D., and Walsh, V. (2009). The parietal cortex and the representation of time, space, number, and other magnitudes. *Philos. Trans. R. Soc. B Biol. Sci.* 364, 1831–1840. doi: 10.1098/rstb.2009.0028
- Calabria, M., Jacquin-Courtois, S., Miozzo, A., Rossetti, Y., Padovani, A., Cotelli, M., et al. (2011). Time perception in spatial neglect: a distorted representation. *Neuropsychology* 25, 193–200. doi: 10.1037/a0021304
- Call, J., and Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends Cogn. Sci.* 12, 187–192. doi: 10.1016/j.tics.2008.02.010
- Cantlon, J. F., Platt, M. L., and Brannon, E. M. (2009). Beyond the number domain. *Trends Cogn. Sci.* 13, 83–91. doi: 10.1016/j.tics.2008.11.007
- Casasanto, D., Fotakopoulou, O., and Boroditsky, L. (2010). Space and time in the child’s mind: evidence for a cross-dimensional asymmetry. *Cogn. Sci.* 34, 387–405. doi: 10.1111/j.1551-6709.2010.01094.x
- Chiao, J., Harada, T., Oby, E., Li, Z., and Parrish, T. (2009). Neural representations of social status hierarchy in human inferior parietal cortex. *Neuropsychologia* 47, 354–363. doi: 10.1016/j.neuropsychologia.2008.09.023
- Clark, D. D., and Sokoloff, L. (1999). “Circulation and energy metabolism of the brain,” in *Basic Neurochemistry: Molecular, Cellular and Medical Aspects*, eds G. J. Siegel, B. W. Agranoff, R. W. Albers, S. K. Fisher, and M. D. Uhler (Philadelphia, PA: Lippincott), 637–670.
- Creem-Regehr, S. H. (2009). Sensory-motor and cognitive functions of the human posterior parietal cortex involved in manual actions. *Neurobiol. Learn. Mem.* 91, 166–171. doi: 10.1016/j.nlm.2008.10.004
- Dávid-Barrett, T., and Dunbar, R. I. M. (2013). Processing power limits social group size: computational evidence for the cognitive costs of sociality. *Proc. Biol. Sci.* 280, 20131151. doi: 10.1098/rspb.2013.1151
- Dehaene, S., Bossini, S., and Giraux, P. (1993). The mental representation of parity and number magnitude. *J. Exp. Psychol. Gen.* 122, 371–396. doi: 10.1037/0096-3445.122.3.371
- Dehaene, S., and Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron* 56, 384–398. doi: 10.1016/j.neuron.2007.10.004
- Dehaene, S., Izard, V., Spelke, E., and Pica, P. (2008). Log or linear? Distinct intuitions of the number scale in Western and Amazonian Indigenous cultures. *Science* 320, 1217–1220. doi: 10.1126/science.1156540
- Dunbar, R. I. M. (1998). The social brain hypothesis. *Evolut. Anthropol.* 6, 178–190. doi: 10.1002/(SICI)1520-6505(1998)6:5<178::AID-EVAN5>3.0.CO;2-8
- Evans, V. (2006). *The Structure of Time: Language, Meaning and Temporal Cognition*. Amsterdam: Brill.
- Fiske, A. P. (1991). *Structures of Social Life: The Four Elementary Forms of Human Relations*. New York: Free Press.
- Fiske, S. T. (1993). Controlling other people: the impact of power on stereotyping. *Am. Psychol.* 48, 621–628. doi: 10.1037/0003-066X.48.6.621

- Flavell, J. H., Green, F. L., and Flavell, E. R. (1986). Development of knowledge about the appearance-reality distinction. *Monogr. Soc. Res. Child Dev.* 51, 1–87. doi: 10.2307/1165866
- Gallese, V., and Lakoff, G. (2005). The brain's concepts: the role of the sensory-motor system in conceptual knowledge. *Cogn. Neuropsychol.* 22, 455–479. doi: 10.1080/02643290442000310
- Gobbini, M. I., Koralek, A. C., Bryan, R. E., Montgomery, K. J., and Haxby, J. V. (2007). Two takes on the social brain: a comparison of theory of mind tasks. *J. Cogn. Neurosci.* 19, 1803–1814. doi: 10.1162/jocn.2007.19.11.1803
- Gould, S., and Vrba, E. (1982). Exaptation – a missing term in the science of form. *Paleobiology* 8, 4–15.
- Hamilton, A. F., Brindley, R., and Frith, U. (2009). Visual perspective taking impairment in children with autistic spectrum disorder. *Cognition* 113, 37–44. doi: 10.1016/j.cognition.2009.07.007
- Hanson, S. J., Matsuka, T., and Haxby, J. V. (2004). Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: is there a “face” area? *Neuroimage* 23, 156–166. doi: 10.1016/j.neuroimage.2004.05.020
- Harcourt, A. H. (1988). “Alliances in contests and social intelligence,” in *Machiavellian Intelligence*, eds R. Byrne and A. Whiten (Oxford: Oxford University Press), 142–152.
- Harcourt, A. H. (1989). “Sociality and competition in primates and non-primates,” in *Comparative Socioecology*, eds V. Standen and R. Foley (Oxford: Blackwell Scientific), 223–242.
- Hayduk, L. A. (1983). Personal space: where we now stand. *Psychol. Bull.* 94, 293–335. doi: 10.1037/0033-2909.94.2.293
- Hubbard, E. M., Piazza, M., Pinel, P., and Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nat. Rev. Neurosci.* 6, 435–448. doi: 10.1038/nrn1684
- Hung, Y.-H., Hung, D. L., Tzeng, O. J. L., and Wu, D. H. (2008). Flexible spatial mapping of different notations of numbers in Chinese readers. *Cognition* 106, 1441–1450. doi: 10.1016/j.cognition.2007.04.017
- Irki, A. (2005). “A prototype of homo faber: a silent precursor of human intelligence in the tool-using monkey brain,” in *From Monkey Brain to Human Brain: A Fyssen Foundation Symposium*, eds S. Dehaene, J. R. Duhamel, M. D. Hauser, and G. Rizzolatti (Cambridge: MIT Press), 253–272.
- Ishihara, M., Keller, P. E., Rossetti, Y., and Prinz, W. (2008). Horizontal spatial representations of time: evidence for the STEARC effect. *Cortex* 44, 454–461. doi: 10.1016/j.cortex.2007.08.010
- Jacob, F. (1977). Evolution and tinkering. *Science* 196, 1161–1166. doi: 10.1126/science.860134
- Jones, E. E., Wood, G. C., and Quattrone, G. A. (1981). Perceived variability of personal characteristics in in-groups and out-groups: the role of knowledge and evaluation. *Pers. Soc. Psychol. Bull.* 7, 523–528. doi: 10.1177/014616728173024
- Kihlstrom, J. F. (2010). Social neuroscience: the footprints of Phineas Gage. *Soc. Cogn.* 28, 757–783. doi: 10.1521/soco.2010.28.6.757
- Knops, A., Thirion, B., Hubbard, E. M., Michel, V., and Dehaene, S. (2009). Recruitment of an area involved in eye movements during mental arithmetic. *Science* 324, 1583–1585. doi: 10.1126/science.1171599
- Kriegeskorte, N., Mur, M., and Bandettini, P. (2008). Representational similarity analysis: connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* 2:4. doi: 10.3389/neuro.06.004.2008
- Lakoff, G. (1986). A figure of thought. *Metaphor Symb.* 1, 215–225. doi: 10.1207/s15327868ms0103_4
- Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: Chicago Press.
- Lakoff, G., and Johnson, M. (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.
- Lee, E., and Tiedens, L. Z. (2001). Is it lonely at the top? The interdependence and interdependence of power holders. *Res. Org. Behav.* 23, 43–91. doi: 10.1016/S0191-3085(01)23003-2
- Liberman, N., and Trope, Y. (2008). The psychology of transcending the here and now. *Science* 322, 1201–1205. doi: 10.1126/science.1161958
- Lugli, L., Baroni, G., Anelli, F., Borghi, A. M., and Nicoletti, R. (2013). Counting is easier while experiencing a congruent motion. *PLoS ONE* 8:e64500. doi: 10.1371/journal.pone.0064500
- Mantini, D., Corbetta, M., Romani, G. L., Orban, G. A., and Vanduffel, W. (2013). Evolutionarily novel functional networks in the human brain? *J. Neurosci.* 33, 3259–3275. doi: 10.1523/JNEUROSCI.4392-12.2013
- Merritt, D. J., Casasanto, D., and Brannon, E. M. (2010). Do monkeys think in metaphors? Representations of space and time in monkeys and humans. *Cognition* 117, 191–202. doi: 10.1016/j.cognition.2010.08.011
- Miles, L. K., Nind, L. K., and Macrae, C. N. (2010). Moving through time. *Psychol. Sci.* 21, 222–223. doi: 10.1177/0956797609359333
- Mitchell, J. P. (2008). Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cereb. Cortex* 18, 262–271. doi: 10.1093/cercor/bhm051
- Nitz, D. A. (2006). Tracking route progression in the posterior parietal cortex. *Neuron* 49, 747–756. doi: 10.1016/j.neuron.2006.01.037
- Norman, K. A., Polyn, S. M., Detre, G. J., and Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn. Neurosci.* 10, 424–430. doi: 10.1016/j.tics.2006.07.005
- Ochsner, K. N. (2007). “Social cognitive neuroscience: historical development, core principles, and future promise,” in *Social Psychology: A Handbook of Basic Principles*, eds A. Kruglanski and E. T. Higgins (New York: Guilford Press), 38–66.
- Oliveri, M., Koch, G., Salerno, S., Torriero, S., Gerfo Lo, E., and Caltagirone, C. (2009). Representation of time intervals in the right posterior parietal cortex: implications for a mental time line. *NeuroImage* 46, 1173–1179. doi: 10.1016/j.neuroimage.2009.03.042
- Parkinson, C., Kohler, P., Sievers, B., and Wheatley, T. (2012). Associations between auditory pitch and visual elevation do not depend on language: evidence from a remote population. *Perception* 41, 854–861. doi: 10.1068/p7225
- Parkinson, C., and Wheatley, T. (2012). Relating anatomical and social connectivity: white matter microstructure predicts emotional empathy. *Cereb. Cortex* doi: 10.1093/cercor/bhs347 [Epub ahead of print].
- Pavlova, M. A. (2012). Biological motion processing as a hallmark of social cognition. *Cereb. Cortex* 22, 981–995. doi: 10.1093/cercor/bhr156
- Pecher, D., and Boot, I. (2011). Numbers in space: differences between concrete and abstract situations. *Front. Psychol.* 2:121. doi: 10.3389/fpsyg.2011.00121
- Peelen, M., and Downing, P. (2007). Using multi-voxel pattern analysis of fMRI data to interpret overlapping functional activations. *Trends Cogn. Sci.* 11, 4–5. doi: 10.1016/j.tics.2006.10.009
- Quadflieg, S., Etzel, J. A., Gazzola, V., Keysers, C., Schubert, T. W., Waiter, G. D., et al. (2011). Puddles, parties, and professors: linking word categorization to neural patterns of visuospatial coding. *J. Cogn. Neurosci.* 23, 2636–2649. doi: 10.1162/jocn.2011.21628
- Redcay, E., Dodel-Feder, D., Pearson, M. J., Mavros, P. L., Kleiner, M., Gabrieli, J. D., et al. (2010). Live face-to-face interaction during fMRI: a new tool for social cognitive neuroscience. *Neuroimage* 50, 1639–1647. doi: 10.1016/j.neuroimage.2010.01.052
- Rossetti, Y., Jacquin-Courtois, S., Rode, G., Ota, H., Michel, C., and Boisson, D. (2004). Does action make the link between number and space representation? Visuo-manual adaptation improves number bisection in unilateral neglect. *Psychol. Sci.* 15, 426–430. doi: 10.1111/j.0956-7976.2004.00696
- Samson, D., Apperly, I. A., Chiavarino, C., and Humphreys, G. W. (2004). Left temporoparietal junction is necessary for representing someone else's belief. *Nat. Neurosci.* 7, 499–500. doi: 10.1038/nn1223
- Saxe, R. (2006). Uniquely human social cognition. *Curr. Opin. Neurobiol.* 16, 235–239. doi: 10.1016/j.conb.2006.03.001
- Sell, A. J., and Kaschak, M. P. (2011). Processing time shifts affects the execution of motor responses. *Brain Lang.* 117, 39–44. doi: 10.1016/j.bandl.2010.07.003
- Seron, X., Pesenti, M., Noël, M. P., Deloche, G., and Cornet, J. A. (1992). Images of numbers, or “When 98 is upper left and 6 sky blue”. *Cognition* 44, 159–196. doi: 10.1016/0010-0277(92)90053-K
- Srinivasan, M., and Carey, S. (2010). The long and the short of it: on the nature and origin of functional overlap between representations of space and time. *Cognition* 116, 217–241. doi: 10.1016/j.cognition.2010.05.005
- Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* 28, 675–735. doi: 10.1017/S0140525X05000129
- Van Essen, D. C., Lewis, J. W., Drury, H. A., Hadjikhani, N., Tootell, R. B., Bakircioglu, M., et al. (2001). Mapping visual cortex in

- monkeys and humans using surface-based atlases. *Vis. Res.* 41, 1359–1378. doi: 10.1016/S0042-6989(01)00045-1
- Vuilleumier, P., Ortigue, S., and Brugger, P. (2004). The number space and neglect. *Cortex* 40, 399–410. doi: 10.1016/S0010-9452(08)70134-5
- Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends Cogn. Sci.* 7, 483–488. doi: 10.1016/j.tics.2003.09.002
- Wheatley, T., Kang, O., Parkinson, C., and Looser, C. (2012). From mind perception to mental connection: synchrony as a mechanism for social understanding. *Soc. Pers. Psychol. Compass* 6, 589–606. doi: 10.1111/j.1751-9004.2012.00450.x
- Xiao, Y. J., and Van Bavel, J. J. (2012). See your friends close and your enemies closer: social identity and identity threat shape the representation of physical distance. *Pers. Soc. Psychol. Bull.* 38, 959–972. doi: 10.1177/0146167212442228
- Yamakawa, Y., Kanai, R., Matsumura, M., and Naito, E. (2009). Social distance evaluation in human parietal cortex. *PLoS ONE* 4:e4360. doi: 10.1371/journal.pone.0004360
- Yamazaki, Y., Hashimoto, T., and Iriki, A. (2009). The posterior parietal cortex and non-spatial cognition. *F1000 Biol. Rep.* 1, 74. doi: 10.3410/B1-74
- Yeshurun, Y., and Sobel, N. (2010). An odor is not worth a thousand words: from multidimensional odors to unidimensional odor objects. *Annu. Rev. Psychol.* 61, 219–241. doi: 10.1146/annurev.psych.60.110707.163639
- Zaki, J. (2013). Cue integration: a common framework for social cognition and physical perception. *Perspect. Psychol. Sci.* 8, 296–312. doi: 10.1177/1745691613475454
- Zorzi, M., Priftis, K., and Umiltà, C. (2002). Brain damage: neglect disrupts the mental number line. *Nature* 417, 138–139. doi: 10.1038/417138a

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 July 2013; accepted: 17 September 2013; published online: 08 October 2013.

Citation: Parkinson C and Wheatley T (2013) Old cortex, new contexts: re-purposing spatial perception for social cognition. *Front. Hum. Neurosci.* 7:645. doi: 10.3389/fnhum.2013.00645

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2013 Parkinson and Wheatley. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Common brain areas engaged in false belief reasoning and visual perspective taking: a meta-analysis of functional brain imaging studies

Matthias Schurz^{1,2*}, Markus Aichhorn^{1,2}, Anna Martin^{1,2} and Josef Perner^{1,2}

¹ Center for Neurocognitive Research, University of Salzburg, Salzburg, Austria

² Department of Psychology, University of Salzburg, Salzburg, Austria

Edited by:

Sarah H. Creem-Regehr, University of Utah, USA

Reviewed by:

Christian Sorg, Klinikum rechts der Isar Technische Universität München, Germany
Elliot Berkman, University of Oregon, USA

*Correspondence:

Matthias Schurz, Center for Neurocognitive Research and Department of Psychology, University of Salzburg, Hellbrunnerstr. 34, 5020 Salzburg, Austria
e-mail: matthias.schurz@sbg.ac.at

We performed a quantitative meta-analysis of functional neuroimaging studies to identify brain areas which are commonly engaged in social and visuo-spatial perspective taking. Specifically, we compared brain activation for visual-perspective taking to activation for false belief reasoning, which requires awareness of perspective to understand someone's mistaken belief about the world which contrasts with reality. In support of a previous account by Perner and Leekam (2008), our meta-analytic conjunction analysis found common activation for false belief reasoning and visual perspective taking in the left but not the right dorsal temporo-parietal junction (TPJ). This fits with the idea that the left dorsal TPJ is responsible for representing different perspectives in a domain-general fashion. Moreover, our conjunction analysis found activation in the precuneus and the left middle occipital gyrus close to the putative Extrastriate Body Area (EBA). The precuneus is linked to mental-imagery which may aid in the construction of a different perspective. The EBA may be engaged due to imagined body-transformations when another's viewpoint is adopted.

Keywords: neuroimaging meta-analysis, theory of mind, false belief, visual perspective taking, temporo-parietal junction

INTRODUCTION

Being able to adopt another person's perspective is an important feature of human social cognition. In the last decade and a half, functional neuroimaging studies have sought to identify the neural mechanisms underlying this ability. Two lines of research have emerged. One group of studies has looked at perspective relevant processes in the context of visuo-spatial cognition, typically by asking about the visual experience arising from a different point of view (visual perspective taking). Studies in this field can be divided into level 1 and 2 visual perspective taking (Masangkay et al., 1974; Flavell et al., 1981). Level 1 perspective taking refers to the ability to distinguish what people can and cannot see, e.g., that two persons looking at different sides of a piece of paper see different things. Level 2 perspective taking refers to the ability to understand that when two persons look at an object from different viewpoints or angles, they arrive at different and maybe contradictory descriptions. Besides research on visual perspective taking, another group of studies has looked at perspective relevant processing in social contexts. The terms "mentalizing", "mind reading" or "theory of mind" refer to our ability to think about the mental states—such as thoughts and beliefs—of ourselves and others (Premack and Woodruff, 1978). A way to test children's ability to attribute mental states to others is the false belief task. Children are told a story in which a character, Mistaken Max, fails to witness how his chocolate is unexpectedly transferred from one location to another. Therefore, he believes that the chocolate is still in its original location. Children have to predict whether he

will look for the chocolate in its original or in its new location. To arrive at the correct answer—its original location—children have to take into account that Mistaken Max holds a false belief about the location of the chocolate, which contrasts with their own knowledge about its real location.

Developmental research showed that the ability to make correct level 2 visual perspective judgments emerges about 2 years later than the ability to master these judgments at level 1 (Masangkay et al., 1974). At the same time when children start to master level 2 judgments—at around 4 to 5 years of age—they also start to pass the false belief test (Wimmer and Perner, 1983). Hamilton et al. (2009) found that theory of mind performance (assessed with a set of tasks including the false belief task) significantly predicted performance on a level 2 visual perspective task in a sample of 4–8 year old children. In contrast, neither performance on a mental rotation task nor children's verbal mental age showed a relation to level 2 visual perspective taking. One explanation (e.g., Perner et al., 2003; Perner and Rössler, 2012) for this link between level 2 visual perspective and false belief understanding is that both tasks require an understanding of perspective, i.e., that different persons can have different views on/or beliefs about one and the same state of affairs. In addition, for both tasks children must be able to intentionally switch to another perspective.

Brain activation for false belief reasoning was mostly studied by presenting short stories to adult participants, and results show a consistent network of brain areas activated (see e.g., Saxe and Kanwisher, 2003; Perner et al., 2006; Saxe and Powell,

2006), including the medial prefrontal cortex (mPFC), bilateral temporal poles, the precuneus, and bilateral temporo-parietal junction (TPJ) areas. The mPFC was linked to the processing of socially and emotionally relevant information about other people that is contained in the stories, but not specifically linked to the processing of belief (Aichhorn et al., 2006; Saxe, 2006; Saxe and Powell, 2006). For example, an fMRI study found that the mPFC was equally engaged by stories about a person's thoughts and by stories about a person's physical appearance or bodily sensations (Saxe and Powell, 2006). The temporal poles were linked to the retrieval of social semantic knowledge from long-term memory, which takes place because participants read stories about persons in social situations (e.g., Gallagher and Frith, 2003; Ross and Olson, 2010). Based on its engagement in visuo-spatial mental imagery (e.g., Ghaem et al., 1997; Hanakawa et al., 2003), it was assumed that the precuneus subserves mental imagery to represent another person's perspective in theory of mind tasks (Cavanna and Trimble, 2006). Similarly, the TPJ areas were linked to the representation of mental and non-mental perspectives (Perner et al., 2006; Perner and Leekam, 2008). The right TPJ was specifically linked to the representation of beliefs, as it was found to respond more strongly when reading statements about a person's thoughts than when reading statements about physical appearance or bodily sensations (Saxe and Powell, 2006), and also compared to reading statements about a person's emotions and perceptions (Zaitchik et al., 2010). In another study, an interesting observation was made for the left TPJ. Perner et al. (2006) presented a novel condition—false sign stories—in addition to the standard false belief and photo control stories. An example for a false sign story is: “The sign to the monastery points to the path through the woods. While playing, the children make the sign point to the golf course. According to the sign the monastery is now in the direction of the . . . golf course / woods”. False sign stories present a very different problem than false belief stories which require figuring out an internal and unobservable mental state of another person. False sign stories simply require reflecting upon the external and directly observable world—the direction to where a sign is pointing. Nevertheless, Perner et al. (2006) found equally high activation for false sign stories as for false belief stories in the left TPJ (but not in the right TPJ). A significantly lower level of activation was found for photo control stories. These findings suggest that the left TPJ is responsible for an operation that is common to reasoning about false belief and false signs: processing of a perspective difference, regardless of whether it is an unobservable inner state (belief) or a visible state (where the sign points). Both in the case of false belief stories and false sign stories, two contrasting perspectives of one and the same state of affairs are involved: the belief of a person that contrasts with one's own knowledge of reality, or the location to which a sign is pointing that contrasts with one's own knowledge about the real location of a target.

In comparison to the detailed picture that has already emerged for the neural correlates of false belief reasoning, brain imaging evidence on visual perspective taking is relatively scattered and has been discussed less extensively. To our knowledge, no systematic review or meta-analysis of visual perspective studies has been done yet. When contrasting judgments about another person's

perspective with judgments about one's own perspective (level 1 and 2 taken together) studies mainly found activation in three areas: (i) Lateral prefrontal cortices (e.g., Vogeley et al., 2004; Aichhorn et al., 2006; David et al., 2006, 2008; Dumontheil et al., 2010; Mazzarella et al., 2013), (ii) bilateral parietal and temporo-parietal areas (Vogeley et al., 2004; David et al., 2006, 2008; Kaiser et al., 2008; Mazzarella et al., 2013) and (iii) the precuneus (Vogeley et al., 2004; Kaiser et al., 2008; Dumontheil et al., 2010). The lateral prefrontal cortices—in particular the inferior frontal gyri—are engaged by cognitive control in interference tasks, such as the color-word Stroop task or stimulus-response reversal studies (for review see Derrfuss et al., 2005). Likewise, researchers linked these areas to the inhibition of the irrelevant own perspective when making visual perspective judgments (McCleery et al., 2011; Ramsey et al., 2013). Activation in temporo-parietal areas was linked to the representation of perspectives, and in particular to the representation of differences in perspective and ownership of perspective (McCleery et al., 2011). In addition, Ramsey et al. (2013) suggested that superior parietal areas are engaged in perspective selection, i.e., in choosing the relevant over the irrelevant perspective. This was assumed to take place in cooperation with lateral prefrontal areas, forming a functional network that is sometimes referred to as the “fronto-parietal control network” (e.g., Vincent et al., 2008). The precuneus was rarely mentioned when discussing the neurocognitive processes subserving visual perspective taking, although it is implicated in multiple forms of visuo-spatial mental imagery (e.g., Ghaem et al., 1997; Hanakawa et al., 2003).

Our review of the neuroimaging literature on false belief reasoning and visual perspective taking showed that both discussed the left TPJ and the precuneus as candidate areas for representing perspectives and perspective differences. Only little functional imaging research has addressed this connection. To our knowledge, no study has directly compared activation for false belief reasoning to activation for visual perspective taking. Aichhorn et al. (2006) measured brain activation for level 2 visual perspective taking, and asked participants to judge the spatial arrangement of two objects (e.g., “the block is in front of the pole”) from the viewpoint of an avatar. The authors found brain activation for level 2 perspective taking—compared to making the same judgments from one's own viewpoint—in an area of the left TPJ that was also activated in a number of earlier studies on theory of mind (e.g., Gallagher et al., 2000; Ruby and Decety, 2003; Saxe and Kanwisher, 2003). Aichhorn et al. (2006) therefore concluded that the left TPJ is responsible for representing different perspectives and is commonly engaged by tasks which require such processing. However, a more recent study provided evidence against this interpretation. David et al. (2008) asked participants to either make a visual perspective or a mentalizing (preference) judgment with respect to two objects in front of an avatar. The avatar was facing participants. In the level 2 perspective judgment, participants were asked which of the two objects (left or right) was elevated from the avatar's point of view. For example, if the elevated object was on the left from the avatar's point of view, this implied that it was shown on the right side of the image to participants. In the preference judgment, participants were asked to judge which object the avatar would prefer—based

on his gestures (e.g., pointing at one object) and facial expression. During the judgments participants always indicated the object as seen from their own perspective. The comparison of brain activation between the tasks showed two completely distinct networks of brain activation and no overlap in the left TPJ. Therefore, David et al. (2008) concluded that visual perspective taking and mentalizing rely on different cortical mechanisms.

Studies that compared brain activation for visual perspective taking and mentalizing show contradictory results. However, these studies never directly compared activation for false belief reasoning with visual perspective taking. As we have outlined above, both developmental research and neurocognitive theories speak for a functional link between these tasks. The present study evaluates the functional overlap between false belief reasoning and visual perspective taking by means of a quantitative meta-analysis of brain imaging studies. To increase statistical power, we analyze both level 1 and 2 visual perspective taking studies in our meta-analysis. Based on the reviewed literature, we expect to find a functional overlap in the left TPJ and in the precuneus.

METHODS

We performed key-word searches in the databases PubMed, Science Citation Index, and PsycInfo. The first criterion of our search was that studies included one of the key-words “neuroimaging” or “fMRI” or “PET”. For our false belief meta-analysis, the second criterion was that studies further included the key-words “false belief” or “theory of mind”. For our visual perspective taking meta-analysis, the second criterion was that studies included the key-words “perspective taking” or “visual perspective” or “viewer rotation”¹. In a second step, we extended our literature samples by searching the reference lists of recent meta-analyses on theory of mind and social cognition (Mar, 2011; Bzdok et al., 2012; Denny et al., 2012; Murray et al., 2012) as well as the reference lists of most recent publications on visual perspective taking (Lambrey et al., 2012; Mazzarella et al., 2013; Ramsey et al., 2013).

We then applied a number of methodological selection-criteria to the literature identified by our search (see e.g., Radua et al., 2012). Studies were only selected if they had performed a whole brain analysis and reported activation coordinates in standard space (MNI or Talairach). We ensured that the same threshold throughout the whole brain was used within each included study, in order to avoid biases toward liberally thresholded brain regions. This does not mean that different studies should employ the same threshold. We included 25 studies ($N = 419$) in our meta-analysis on false belief and 14 studies ($N = 216$) in our meta-analysis on visual perspective taking. We used Effect-Size Signed Differential Mapping (ES-SDM) software, version 2.31 for meta-analysis (Radua et al., 2010, 2012; <http://www.sdmproject.com>). ES-SDM uses standard effect size and variance-based meta-analytic calculations. Based on the reported t -values and the sample size of a study, ES-SDM creates a map of effect-sizes (Hedge’s g values) and their variances. Variance is estimated from the map

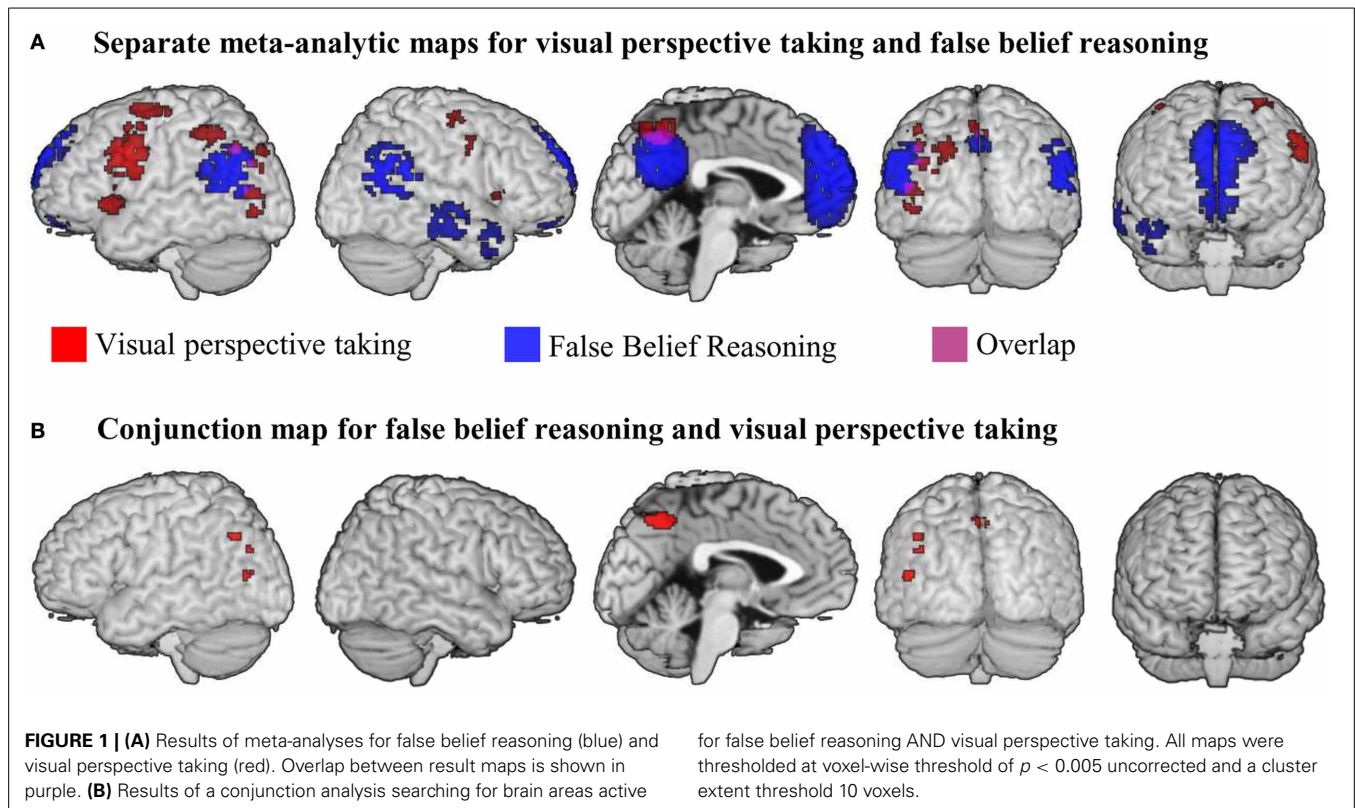
of effect-sizes and the sample size of the study. Effect-sizes are exactly calculated for those voxels containing a peak reported in the results table of an original study. For the rest of the voxels, an effect-size is estimated depending on the distance to close peaks (<20 mm) by means of an unnormalized Gaussian kernel. In the present analysis, we used the recommended Gaussian kernel with a FWHM of 20 mm. A validation study which compared the results of coordinate based ES-SDM meta-analysis to the results of a standard voxel-wise GLM analysis of the same original data (Radua et al., 2012) found that this FWHM provided an optimal balance between sensitivity and specificity. For statistical-analysis, all foci were transformed to Talairach space which is the native space of the software, by using the matrix transformations proposed by Lancaster et al. (2007). We calculated a mean analysis for each task-group. Calculation of the meta-analytic mean map is implemented by a random-effects model in which each study is weighted by the inverse of the sum of its variance plus an estimate of between-study heterogeneity. The latter is obtained by the DerSimonian-Laird method (DerSimonian and Laird, 1986). This approach enables studies with larger sample size or lower variability to contribute more and that effects are assumed to randomly vary between samples. The statistical significance was assessed by a permutation test; 100 random maps were generated with the same number of input foci as included in the to-be-tested map (see Radua et al., 2012). Finally, the meta-analytic maps were thresholded using a voxel-level (height) threshold of $p < 0.005$ (uncorrected) and a cluster-level (extent) threshold of 10 voxels. This uncorrected threshold was found to optimally balance sensitivity and specificity, and to be an approximate equivalent to a corrected threshold of $p < 0.05$ in original neuroimaging studies (Radua et al., 2012). We performed a conjunction analysis (see Figure 1B) with the “image calculator” utility in SPM8 (www.fil.ion.ucl.ac.uk). Conjoint activation is determined by a voxel-wise combination of results by a logical AND function. For convenience, we report all activations in MNI-space.

RESULTS

FALSE BELIEF REASONING

Studies on false-belief reasoning mainly used two types of tasks. One group of studies contrasted stories about false belief with stories about an outdated photograph. We give some examples in Table 1. In total, we found 15 studies (reported in 14 publications) that relied on this type of contrast (Saxe and Kanwisher, 2003; Saxe and Wexler, 2005; Perner et al., 2006; Saxe and Powell, 2006; Saxe et al., 2006; Young et al., 2007, 2010, 2011; Kliemann et al., 2008; Mitchell, 2008; Aichhorn et al., 2009; Young and Saxe, 2009; Dodell-Feder et al., 2011; Lee et al., 2011). In the false belief story a short text passage is presented, which involves a person holding a false belief. A test question asks participants about the belief or its behavioral consequences. In the control task, a short text passage describes a photograph (or a similar physical representation) of the past, together with a note about how things shown on the photograph have changed by now. Participants are asked what is shown on the photo. Another more heterogeneous group of studies presented similar stories about false belief. In this group of studies, however, stories of different length and richness were presented, and different types of control stories were used.

¹Viewer rotation’ refers to an imagined change of one’s own point of view, i.e., imagining oneself rotating around an object in space, arriving at a new viewpoint on it.



We give some examples for false belief studies of our second group in the lower part of **Table 1**. The common element of these studies is that they present a story (sentence or cartoon format) about a person that holds a false belief as activation tasks. Participants are asked a question which relates to the false belief of the person. In the control condition, again a story about a person is presented, but here the person does not hold a false belief. Participants are asked about non-mental state information in the story. In total, we found 10 studies that relied on this type of contrast (Fletcher et al., 1995; Happé et al., 1996; Gallagher et al., 2000; Nieminen-von Wendt et al., 2003; Hynes et al., 2006; Kobayashi et al., 2006, 2007; Gobbi et al., 2007; Abraham et al., 2010; Jimura et al., 2010). We pooled the two groups of tasks that present stories about false belief into one single meta-analysis (total $n = 25$).

We performed a meta-analysis on the reported activation maps for the contrast false belief stories > control stories. Results are shown in blue in **Figure 1A** and are listed in **Table 2**. The largest cluster of meta-analytic convergence was found in the mPFC, including parts of dorsal and ventral mPFC and the anterior cingulate gyrus. Another large cluster of convergence was found in precuneus and posterior cingulate gyrus bilaterally. Further clusters of convergent activation were found in bilateral temporoparietal areas, spanning across parts of middle and superior temporal gyri up to the inferior parietal lobule (up to $z = 42$). Two smaller clusters of convergence were found in anterior parts of the right temporal lobe.

VISUAL PERSPECTIVE TAKING

Compared to the large number of imaging studies on false belief reasoning, relatively few imaging studies on visual perspective

taking exist. We identified three groups of visual perspective tasks in the literature: level 1 visual perspective taking (3 studies), level 2 visual perspective taking (5 studies), and level 2 imagined viewer rotation (6 studies). Due to the small sample-sizes of these task-groups, it was not possible to perform individual meta-analyses. We therefore decided to merge the different visual perspective tasks into a pooled analysis, which gave us a large enough sample for quantitative meta-analytic calculations ($n = 14$). Later on (see section Region of interest based review), we provide a complementary results overview for individual task-types.

Table 3 gives task-descriptions for all visual-perspective taking studies in our meta-analysis. Level 1 visual perspective taking studies typically present a scene with an avatar and a number of objects. Participants are asked how many of these objects the avatar can see (while some of the objects are behind the avatars' back). In the control conditions of level 1 visual perspective taking studies, participants are asked how many objects they can see themselves². Level 2 visual perspective taking tasks also typically present a scene with an avatar and a number of objects. However, here the avatar is able to see all of the objects in the

²A recent level 1 visual perspective taking study (Ramsey et al., 2013) did not look at the contrast other > self, but at interactions between perspective taking (self vs. other) and the consistency of perspectives (i.e., do self and other perspectives differ in the task-response that they require?). Ramsey et al.'s (2013) results show that the consistency between perspectives is an important and previously ignored determinant of brain activation. However, these data go beyond the scope of the present meta-analysis and cannot be synthesized with results from other studies in our meta-analysis. Therefore, we did not include Ramsey et al.'s (2013) study in our analysis.

Table 1 | Examples for false belief reasoning tasks.

Author	Activation task	Control task
FALSE BELIEF VIGNETTES VS. PHOTO CONTROL VIGNETTES (3 EXAMPLES OUT OF 15 STUDIES)		
Aichhorn et al., 2009	Read a short vignette involving a person holding a false belief. Predict the behavior of that person based on her belief. e.g.: <i>'Julia sees the ice cream van go to the lake. She doesn't see that the van turns off to the town hall. Therefore, Julia will look for the ice cream van at the ... ?' (lake or town hall).</i>	Read a short vignette involving a photograph of the past, and a description how things shown on the photo have changed by now. Answer a question about the outdated scene shown on the photo. e.g.: <i>'Julia takes a picture of the ice- van in front of the pond. The ice cream van changes to the market place; the picture gets developed. On the picture, the ice- van is at the ... ?' (pond or market place).</i>
Saxe and Kanwisher, 2003	Read a short vignette involving a person holding a false belief. Answer a question about her belief. e.g.: <i>'John told Emily that he had a Porsche. Actually, his car is a Ford. Emily doesn't know anything about cars so she believed John. When Emily sees John's car, she thinks it is a ... ?' (Porsche or Ford).</i>	Read a false-photograph vignette. Answer a question concerning the outdated content in the photo. e.g.: <i>'A photograph was taken of an apple hanging on a tree branch. The film took half an hour to develop. In the meantime, a strong wind blew the apple to the ground. The developed photograph shows the apple on the ... ?' (tree or ground).</i>
Lee et al., 2011	Read a short vignette involving a person holding a false belief. Answer a question about her belief. e.g.: <i>'David knows that Ethan is very scared of spiders. Ethan, alone in the attic, sees a shadow move and thinks it is a burglar. David hears Ethan cry for help. David assumes that Ethan thinks he has seen ... ?' (a spider or a burglar).</i>	Read a false-photograph vignette. Answer a question concerning the outdated content in the photo. e.g.: <i>'Amy made a drawing of a treehouse three years ago. That was before the storm. We built a new treehouse last summer, but we painted it red instead of blue. The treehouse in Amy's drawing is ... ?' (red or blue).</i>
FALSE BELIEF STORIES VS. CONTROL STORIES (2 EXAMPLES OUT OF 10 STUDIES)		
Abraham et al., 2010	Read a story describing a character's mental state (belief/desire) about an object. Then the object in reality is described. Indicate if the character is surprised/delighted by the reality. e.g.: <i>'Thomas believes that there will be a lot of sugar in the chocolate pudding.–The chocolate pudding does not taste sweet.–Would that surprise Thomas?' (yes or no).</i>	Read a story describing a non-mental attribute of a group of persons. Then a particular member of this group is introduced. Indicate if that member holds the non-mental attribute (syllogistic reasoning).e.g.: <i>'All students at the dance academy own more than three pairs of dance shoes.–Sonja studies dance at the academy.–Does Sonja own more than three pairs of dance shoes?' (yes or no).</i>
Fletcher et al., 1995	Read a story about a character performing an action. The reason for that action must be inferred from his/her false belief (sometimes also ignorance). Explain (silently) why he did that. e.g.: <i>'A burglar who has just robbed a shop is making his getaway. As he is running home, a policeman on his beat sees him drop his glove. He doesn't know the man is a burglar, he just wants to tell him he dropped his glove. But when the policeman shouts out to the burglar, "Hey, you! Stop!," the burglar turns around, sees the policeman and gives himself up. He puts his hands up and admits that he did the break-in at the local shop.'</i>	Read a story about a character performing an action. The reason for that action can be inferred from the information in the story and is not related to false belief. Explain (silently) why he did that. e.g.: <i>'A burglar is about to break into a jewelers' shop. He skillfully picks the lock on the shop door. Carefully he crawls under the electronic detector beam. If he breaks this beam it will set off the alarm. Quietly he opens the door of the store-room and sees the gems glittering. As he reaches out, however, he steps on something soft. He hears a screech and something small and furry runs out past him, toward the shop door. Immediately, the alarm sounds.'</i>

scene, but views them from a different angle. Participants are asked to indicate the relative position of one object from the avatar's viewpoint. In the control condition of level 2 visual perspective tasks, participants are asked about the relative location of one object from their own perspective. The last type of visual-perspective taking tasks in our meta-analysis, level 2 imagined viewer rotation tasks, typically present an array of objects and ask to imagine viewing this array from a different angle. Then, participants are asked to indicate the relative position of one object from the imagined viewpoint. Two types of control tasks are frequently used in studies on imagined viewer rotation. In one type of control task, participants have to indicate the relative position of one

object in the array as seen from their actual viewpoint (similar to the control conditions in level 2 visual perspective taking tasks). In another type of control task, a so-called object rotation task, participants are asked to imagine rotating the array around its vertical axis (e.g., with their right hand), and then indicate the current position of one object from their viewpoint.

We performed a meta-analysis on the reported activations for all three types of visual perspective taking compared to their respective control condition. **Figure 1A** shows clusters of reliable meta-analytic convergence for visual perspective taking in red, and results are listed in **Table 2**. The largest cluster of convergent activation was found in the left lateral prefrontal cortex, with its

Table 2 | Results of meta-analyses for False Belief Reasoning and Visual Perspective Taking.

#	Hem	Label	Cluster center			BA	z-val.	vx	Individual foci			Label
			x	y	z				x	y	z	
FALSE BELIEF REASONING (25 STUDIES)												
1	R	Precuneus	6	−59	35	7	5.00	1282	4	−42	33	R cingulate
									0	−72	38	L cuneus
									0	−59	35	L precuneus
									0	−42	35	L cingulate
2	R	Supramarginal	62	−45	21	40	7.70	886	55	−67	15	R mid. temporal
									47	−61	39	R angular
									56	−56	25	R sup. temporal
									58	−52	42	R inf. parietal
3	R	Sup. temporal	51	−9	−9	22	3.21	326	61	−23	−17	R mid. temporal
									61	−16	−15	R mid. temporal
4	R	Sup. temporal	46	11	−24	38	2.606	207	48	8	−33	R sup. temporal
									40	11	−22	R inf. frontal
5	L	Mid. temporal	−57	−65	27	39	6.58	917	−50	−61	45	L inf. parietal
									−44	−61	40	L angular
									−59	−56	29	L sup. temporal
									−55	−47	28	L supramarginal
6	L	Sup. frontal	−5	60	21	9	6.42	2083	−18	47	29	L sup. frontal
									−5	60	28	L med. frontal
									3	48	−7	R ant. cingulate
									4	50	33	R med. frontal
									6	58	23	R sup. frontal
VISUAL PERSPECTIVE TAKING (14 STUDIES)												
1	R	Precentral	41	−8	54	4	2.69	72	58	1	32	R precentral
2	R	Insula	44	15	2	13	2.53	21				
3	L	Inf. occipital	−48	−80	2	18	2.74	71	−46	−77	11	L mid. occipital
4	L	Mid. temporal	−39	−73	32	39	2.53	101	−31	−81	33	L sup. occipital
									−18	−63	40	L precuneus
5		Precuneus	0	−53	52	7	2.76	187	0	−66	54	precuneus
									4	−54	50	R precuneus
6	L	Cerebellum	−27	−53	−35		2.73	175	−23	−55	−28	L cerebellum
									−24	−59	−38	L cerebellum
7	L	Inf. parietal	−37	−43	47	40	2.95	252	−41	−59	42	L angular
									−37	−60	33	L mid. temporal
8	L	Mid. frontal	−42	15	33	9	3.60	1054	−33	−5	62	L precentral
									−42	5	34	L precentral
									−38	7	13	L insula
									−42	8	27	L inf. frontal
									−44	14	−5	L insula
FALSE BELIEF and VISUAL PERSPECTIVE TAKING												
1	R	Precuneus	4	−54	50	7	2.66	46	6	−56	46	R precuneus
2	L	Mid. temporal	−39	−73	32	39	2.49	4				
3	L	Mid. occipital	−49	−72	13	37	2.32	12				
4	L	Angular	−41	−59	42	39	2.46	19	−41	−63	40	L angular
5		precuneus	0	−53	52	7	2.76	63	0	−66	54	precuneus
									−5	−51	52	L precuneus

peak in the left middle frontal gyrus. The cluster further included parts of the inferior frontal gyrus, the insula and the precentral gyrus. In the right hemisphere, lateral prefrontal activation was substantially smaller compared to the left. Two small clusters of activation were found, located in the right precentral gyrus and right insula. Larger clusters were found in the left inferior parietal

lobule and in the precuneus. The left inferior parietal cluster included parts of the angular gyrus and the posterior middle temporal gyrus. The precuneus cluster spanned both hemispheres. In addition to the left inferior parietal area, two other clusters of convergence were found in left temporo-parietal areas. One was located in the left posterior middle temporal gyrus extending into

Table 3 | ROI-based follow-up review: + signs denote that a study reported activation within 20 mm distance to a peak of our meta-analytic conjunction (20 mm corresponds to the smoothness of meta-analysis).

First author	PREC	ANG	OCC	Persp. task	Control task
LEVEL 1 VISUAL PERSPECTIVE TAKING					
Vogele et al., 2004	+	+		You see a scene including an avatar and a number of objects. Indicate how many objects the avatar can see.	You see a scene including an avatar and a number of objects. Indicate how many objects you see.
Kaiser et al., 2008	+	+		You see a scene including an avatar and a number of objects. Indicate how many objects the avatar can see.	You see a scene including an avatar and a number of objects. Indicate how many objects you see.
Dumontheil et al., 2010*	+		+	You see a scene including an avatar and a number of objects. Follow the avatar's instruction (e.g., 'move the large object up'). This instruction is dependent on the avatar's perspective (e.g., he can't see the largest object in the scene).	You see a scene including an avatar and a number of objects. Follow the avatar's instruction (e.g., 'moves the large object up'). As a rule, if the avatar has a male (female) voice, you can only move certain objects.
LEVEL 2 VISUAL PERSPECTIVE TAKING					
Aichhorn et al., 2006				You see a scene including an avatar and two objects. Indicate their relative spatial arrangement e.g., 'Block is in front of the pole' from the viewpoint of an avatar.	You see a scene including an avatar and two objects. Indicate their relative spatial arrangement e.g., 'Block is in front of the pole' from your own viewpoint.
David et al., 2006		+		You see a scene including two avatars facing you. You and the avatars play a ball-tossing game. Take the perspective of one of them and indicate in which direction (left or right) he must throw the ball to pass it to the other avatar.	You see a scene including two avatars facing you. You and the avatars play a ball-tossing game. Indicate from your own perspective in which direction (left or right) you must throw the ball to pass it to one of the avatars.
David et al., 2008		+		You see a scene including an avatar facing you and two objects, located between you and the avatar. Indicate from his perspective which object (left or right) is elevated.	You see a scene including an avatar facing you and two objects, located between you and the avatar. Indicate from your perspective which object (left or right) is elevated.
Kockler et al., 2010			+	You see a scene including an avatar and one object. Indicate from the avatar's perspective if the object is to his left or right.	You see a scene including an avatar and one object. Indicate from your own perspective if the object is to your left or right.
Mazzarella et al., 2013		+		You see a table with an object on it. An avatar stands next to the table. Indicate from his perspective if the object is to his left or right.	You see a table with an object on it. An avatar stands next to the table. Indicate from your perspective if the object is to your left or right.
LEVEL 2 IMAGINED VIEWER ROTATION					
Creem et al., 2001	+			You see an array of four objects. Imagine being located in the array's center. Then imagine that your body position is rotated to a certain degree. Perform a relative location judgment (which object is on your right?).	You see an array of four objects. Imagine being located in the array's center. Stick to your actual body orientation, do not imagine a rotation. Perform a relative location judgment (which object is on your right?).
Zacks et al., 2003		+	+	You see an array of four objects. Imagine viewing the array from a different angle (i.e., imagine a self-rotation around the array). Indicate if a particular object is now on the left or right side of the array.	You see an array of four objects. Imagine that the array rotates along its vertical axis, while your own position remains the same. Indicate if a particular object is now on the left or right side of the array.
Wraga et al., 2005			+	You see a Shepard-Metzler object. Imagine viewing the object from a different angle (i.e., imagine a self-rotation around the array). Indicate if you now can see a particular side of the object.	You see a Shepard-Metzler object. Imagine rotating the object along. Indicate if you now can see a particular side of the object.

(Continued)

Table 3 | Continued

First author	PREC	ANG	OCC	Persp. task	Control task
Creem-Regehr et al., 2007	+		+	You see an array of 6 objects surrounding a hand. Imagine viewing the hand from the position of one of the objects (i.e., imagine a self-rotation to this position). Indicate if the thumb is now to your left or right.	You see an array of 6 objects surrounding a hand. Indicate from your actual viewpoint if the thumb is to your left or right.
Wraga et al., 2010	+	+	+	You see a Shepard-Metzler object. Imagine viewing the object from a different angle (i.e., imagine a self-rotation around the array). Indicate if you now can see a particular side of the object.	You see a Shepard-Metzler object. Imagine holding the object in the right hand and rotating it in a specific way. Indicate if you now can see a particular side of the object.
Lambrey et al., 2012				You see a table with objects on it. Imagine rotating yourself around it to the position of an avatar / arrow (activation collapsed). Memorize the arrangement of objects from this perspective (tested later).	You see a table with objects on it. Imagine rotating the table until one object is in front of an avatar / arrow (activation collapsed). Memorize the current arrangement of objects from your own perspective (tested later).

PREC ... Precuneus $x = 0$, $y = -53$, $z = 52$; ANG ... Angular Gyrus $x = -41$, $y = -59$, $z = 42$; OCC ... Middle Occipital Gyrus $x = -49$, $y = -72$, $z = 13$;

*Dumontheil et al.'s (2010) study could also be classified as a level 2 perspective task. The picture stimuli used in the task show a level 1 perspective difference. However, the task also presents statements (e.g., "move the large ball up") that have to be interpreted from another person's perspective. A correct interpretation requires understanding that the other person has a different perspective of the entire scene (from his perspective, one particular ball is the largest of all, whereas from one's own point of view, another ball is the largest of all).

the superior occipital gyrus; the other located in the left inferior and middle occipital gyri, near the location of the Extrastriate Body Area (EBA, Downing et al., 2001). Finally, a cluster of convergent activation was found in the left cerebellum (not visible in **Figure 1A** because of its location buried underneath the cerebellar surface).

CONJUNCTION ANALYSIS

Our conjunction analysis determined which brain areas showed convergent activation for both false belief reasoning and visual perspective taking. Results are listed in **Table 2** and illustrated in **Figure 1B**. The largest areas of convergence for both meta-analyses were found in bilateral precuneus, with a slightly larger cluster in the left compared to the right precuneus. Further conjoined clusters of convergence were found in the left TPJ (angular gyrus and the posterior middle temporal gyrus) and the left middle occipital gyrus corresponding to the EBA.

REGION OF INTEREST BASED REVIEW

We followed-up the findings of our meta-analytic conjunction by a region of interest (ROI) based review. This approach does not include a statistical comparison. However, it gives an overview of which visual perspective taking studies contributed to the meta-analytic findings. We selected three peaks from our meta-analytic conjunction as ROIs: precuneus ($x = 0$, $y = -53$, $z = 52$), left dorsal TPJ/angular gyrus ($x = -41$, $y = -59$, $z = 42$) and left middle occipital gyrus ($x = -49$, $y = -72$, $z = 13$). ROIs were created by a 20 mm spherical volume around the peak coordinates. This radius corresponds to the size of the smoothing (FWHM) used by our meta-analysis. The other peaks from our conjunction analysis (left posterior middle temporal gyrus, right precuneus) did not enter our ROI analysis because they were

located at too close distance to the three other ROIs, and were therefore practically not separable from them.

For each study, we checked if any of the reported activation coordinates fell within the 20 mm sphere around the three peak coordinates in the left precuneus, the left angular gyrus, and the left middle occipital gyrus. **Table 3** summarizes the results of this review. It lists each study and indicates with a '+' symbol if a study reported activation within a ROI. Contributions to the meta-analytic peak activation in the left angular gyrus were balanced over the three types of visual perspective taking: level 1 visual perspective (2/3 studies), level 2 visual perspective (3/5 studies), and level 2 viewer rotation (2/6 studies). Contributions to the peak activation in the left middle occipital gyrus were relatively weak for level 1 visual perspective (1/3 studies) and level 2 visual perspective (1/5 studies), but more substantial for level 2 viewer rotation (4/6 studies). Contributions to the meta-analytic peak in the left precuneus were relatively strong for level 1 visual perspective (3/3 studies), moderate for 2 viewer rotation (3/6 studies), and completely absent for level 2 visual perspective (0/5 studies).

DISCUSSION

We meta-analyzed brain activation for false-belief reasoning and visual perspective taking and looked for common brain areas engaged by these tasks with a conjunction analysis. We expected to find common activation in the left TPJ, based on our hypothesis that this area is implicated in processing perspective differences (Perner and Leekam, 2008). Our results confirm this expectation, as we found two clusters in the dorsal left TPJ (angular gyrus and posterior middle temporal gyrus) that were reliably engaged both in false belief and in visual perspective processing. In addition to these clusters, our meta-analysis revealed common areas in the left middle occipital gyrus and in the precuneus for false belief reasoning and visual perspective taking. In the next sections, we

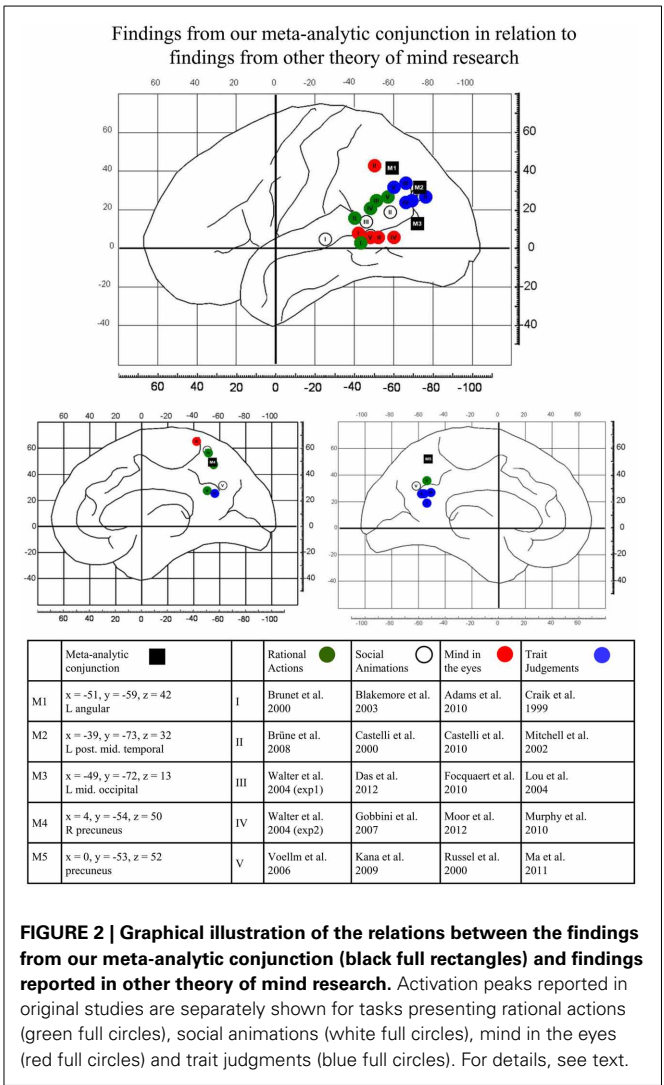
will discuss the potential functional roles of these locations of convergent brain activation.

LEFT TEMPORO-PARIETAL JUNCTION

Our meta-analytic conjunction found two clusters of conjoint activation for visual perspective taking and false belief reasoning in the left dorsal TPJ, one in the angular gyrus at $z = 42$ and another one in the left posterior middle temporal gyrus at $z = 32$. No overlap in activation was found for right TPJ areas. These findings support the functional account of TPJ areas reviewed in our introduction (Perner et al., 2006; Perner and Leekam, 2008). In this view, the right TPJ is mostly responsible for belief-desire reasoning. Accordingly, our meta-analysis found activation only for false belief reasoning here, and no activation for visual perspective taking. The left TPJ, on the other hand, is thought to be involved in processing of alternative perspectives in a domain-general way. In support of this idea, we found an overlap in brain activation between visual perspective taking and false belief reasoning here.

An interesting aspect of the found overlap between visual perspective taking and false belief reasoning relates to its location within the left TPJ. Literature reviews have shown that different theory of mind tasks engage different parts of the left TPJ (Gobbini et al., 2007; Perner and Leekam, 2008; Bahnemann et al., 2010). For example, Perner and Leekam (2008) report that theory of mind tasks which require processing of a perspective difference—as for example the false belief tasks—engage more dorsal parts of the left TPJ, whereas theory of mind tasks that do not require such processing only engage more ventral parts located around the posterior Superior Temporal Sulcus (pSTS). This distinction is also relevant for the interpretation of David et al.’s (2008) study, which failed to find a functional overlap between visual perspective taking and theory of mind. David et al. (2008) used a preference judgment task to test theory of mind. Different from the false belief task, this task does not require processing of a perspective difference. Preferences are specific relations between a person and an object (e.g., “Max does not like apples”, “I do like apples,” but there is no difference in perspective, for Max and I have the same view. We both know that he hates apples and I like apples. Consequently, it becomes clear that—based on the functional distinction between dorsal and ventral pSTS made in literature reviews—one would not expect an overlap with visual perspective taking (in left dorsal TPJ). Whereas visual perspective taking should engage the left dorsal TPJ, the preference decision task should engage other areas more ventrally in the TPJ and in pSTS. Indeed, David et al. (2006) found activation for the preference decision task only in the right pSTS, and activation in more dorsal parietal areas for visual perspective taking. Conversely, the present meta-analysis looked at a theory of mind task that does present a perspective difference (false belief) and did find a functional overlap in the left dorsal TPJ with visual perspective taking.

To check whether the proposed functional distinction between dorsal and ventral TPJ can be linked to our observed activations for visual perspective taking, we performed an informal review which compares activation for different theory of mind studies to activation for visual perspective taking as found in our meta-analysis. In **Figure 2**, we indicate the results of our



conjunction analysis between false belief reasoning and visual perspective taking by black boxes. In addition, we tentatively summarize temporo-parietal findings from popular theory of mind tasks by reviewing the peak-activations found in temporo-parietal areas for 5 studies per task-type. Green circles indicate locations for rational actions (Brunet et al., 2000; Walter et al., 2004; Voellm et al., 2006; Brüne et al., 2008). These tasks typically have a non-verbal format and present a cartoon-story about a person in the activation tasks. Participants are then asked about the goal of the person in the story, i.e., to predict what will happen next. In the control task, questions about non-mental aspects of the stories are asked (e.g., physical causality). White circles in **Figure 2** indicate activation-peaks reported for social animations (Castelli et al., 2000; Blakemore et al., 2003; Gobbini et al., 2007; Kana et al., 2009; Das et al., 2012). These studies typically present video animations of simple geometrical shapes (see Heider and Simmel, 1944). In the activation condition, the animations portray actions which are typical for an intentional or social interaction. In the control condition, the animations show random or purely mechanical movements. For

each movie, participants are asked to explain what is shown. Red circles in **Figure 2** show activations for the so-called “mind in the eyes” tasks (after Baron-Cohen et al., 1999). The reviewed studies (Russell et al., 2000; Adams et al., 2010; Castelli et al., 2010; Focquaert et al., 2010; Moor et al., 2012) typically show in the activation task a photograph of a pair of eyes and ask which of two adjectives (e.g., “concerned” vs. “unconcerned”) best describes the mental state of the person. In the control tasks, again a photo of eyes is shown and participants are asked to indicate the gender of the depicted person.

Rational actions, social animations and mind in the eyes all do not require awareness of perspectives or processing of perspective differences for task performance. Consistent with Perner and Leekam’s (2008) theorizing, **Figure 2** shows that activations of these three task-types are mostly located ventrally and anteriorly to our conjunctions results. However, activation for another type of theory of mind task—judgments about another person’s personality traits—shows some overlap with the dorsal TPJ areas identified by our conjunction analysis. Trait judgment tasks (Craik et al., 1999; Mitchell et al., 2002; Lou et al., 2004; Murphy et al., 2010; Ma et al., 2011) typically present personality trait-adjectives. In the activation task, participants are asked to indicate whether the adjective describes a particular person or not. In the control tasks, participants perform a non-mental state related task on similar trait words (e.g., *is this word written in upper- or lower-case?*).

As false-belief reasoning and visual-perspective taking, trait judgments may also require awareness of perspective, but for different reasons. Traits indicate habitual patterns of behavior, thought, and emotion. They are characteristic for a person when the person’s habits deviate from the norm. For instance, a person is called “anxious” or “nervous” (Mitchell et al., 2002) if she tends to be concerned about situations where one normally has no reason to be anxious, i.e., the person takes a deviant perspective on how dangerous or challenging a situation is. Or a person is “stubborn” (Murphy et al., 2010) if she refuses to change her opinion or position on a subject when objectively (from the judging person’s point of view) it is time to give up. So, many traits result from habitually biased perspectives, and trait judgments are judgments about whether a person habitually takes a different perspective on certain aspects of life.

PRECUNEUS

Although the precuneus is part of the typical set of brain areas active in theory of mind tasks (see e.g., Mar, 2011; Bzdok et al., 2012), relatively little has been said about its functional role in processing mental states of others. Several lines of research show that the area is implicated in mental imagery, i.e., the construction of a visual scene in absence of the appropriate external stimulus (Thomas, 2010). Studies found activation in the precuneus for the imagined execution of movements (e.g., Hanakawa et al., 2003), mental simulation of routes (Ghaem et al., 1997), mental imagery in deductive reasoning (e.g., Knauff et al., 2003) and for processing of intervals between tones in music perception (e.g., Platel et al., 1997). In their review on the precuneus, Cavanna and Trimble (2006) suggested that the main function of the precuneus in theory of mind is mental imagery to represent the perspective

of another person. This function would be compatible with our finding that this area is engaged both in false belief reasoning and in visual perspective taking. Unexpectedly, however, we observed in our follow-up review that the precuneus tended to be engaged only by level 1 perspective and level 2 imagined viewer rotation tasks, but not by level 2 visual perspective taking tasks. Based on the assumption that activation in the precuneus reflects mental imagery to represent another’s perspective, we would clearly expect activation also for level 2 visual perspective tasks. Contrary to that, we did not find such activation in any of the five reviewed level 2 visual perspective taking studies.

LEFT MIDDLE OCCIPITAL GYRUS

Although activation in the lateral occipital cortex can be found for multiple forms of visual object recognition and visuo-spatial processing, we are particularly interested by the fact that our cluster in the left middle occipital gyrus is in good correspondence to the location of the EBA, with an euclidian distance of 5 mm to the coordinates reported in the seminal paper by Downing et al. (2001). The EBA was traditionally considered as a category-selective region for the visual processing of static images of the human body. Saxe et al. (2005) found that while the right EBA shows preferential activity for allocentric views on body-parts (i.e., the typical view we have on others), the left EBA is equally active for egocentric (i.e., the typical view we have on ourselves) and allocentric views on body-parts. Astafiev et al. (2004) found that the EBA is also engaged when participants perform movements (e.g., arm movement) in the absence of visual feedback. The authors interpreted these results as showing that in addition to a visual recognition function, the EBA is also engaged in maintaining our bodily representation by integrating visual, spatial attention, and sensory-motor signals. Recently, it has also been found that the EBA is engaged by imagined body movements. For example, Iseki et al. (2008) found activation in the EBA when participants were asked to imagine walking around in a room while they were actually lying in the fMRI scanner. (Deen and McCarthy, 2010) found activation in the EBA when participants read stories including passages about human movements (for example, ‘... on Christmas morning, Johnny ran down the stairs to the tree ...’) compared to control stories (‘... Susan is sympathetic to children with disabilities ...’).

Altogether, research on the EBA suggests that this area is involved in maintaining a bodily self-representation, and that this process is also engaged when one imagines a body movement of oneself or others. We speculate that activation in the EBA found in our meta-analysis may reflect imagined bodily transformations related to adopting a different visual perspective. We want to note that our follow-up review found that activation in the EBA mainly stemmed from level 2 imagined viewer rotation studies. This kind of task clearly invites imagining a movement of one’s own body.

BRAIN CONNECTIVITY

To give a complementary characterization of our main findings, we take a look at their structural and functional connectivity profiles.

For a characterization of connectivity of the left dorsal TPJ, we refer to the work by Caspers et al. (2011) who present

structural connectivity fingerprints from probabilistic fiber tract analyses for different parts of the left inferior parietal lobe. The activation peak from our conjunction analysis ($x = -42$, $y = -59$, $z = 42$) falls into the left angular gyrus and more precisely, in the cytoarchitectonic area PGa according to the Jülich Histological Atlas (Caspers et al., 2006, 2008) which is accessible with the software fslview (<http://fsl.fmrib.ox.ac.uk/fsl/fslview/>). The connectivity fingerprint for the area PGa is presented in Caspers et al. (2011, p. 371). In the left hemisphere, area PGa shows strong structural connectivity to (i) lateral prefrontal areas, in particular areas of the inferior frontal gyrus (ii) posterior occipito-temporal areas and posterior fusiform areas (iii) areas of the insula and (iv) parts of the superior parietal lobe. In addition, moderate connectivity is found to more anterior parts of the temporal gyrus and posterior cingulate gyrus/ventral precuneus.

For a characterization of precuneus connectivity, we rely on results from a recent resting-state functional connectivity analysis of this area (Zhang and Li, 2012). Results show that more dorsal parts of the precuneus are strongly linked to lateral occipital, superior parietal as well as lateral prefrontal areas in both hemispheres. More ventral parts of the precuneus are strongly linked to bilateral lingual gyri and the calcarine sulcus, bilateral inferior parietal lobuli (in particular the angular gyri) and the ventral mPFC. The activation peak from our conjunction analysis lies on the border between ventral and dorsal precuneus as defined by Zhang and Li (2012).

Taken together, connectivity data show that our three main findings, the left dorsal TPJ (corresponding to the angular gyrus and area PGa), the precuneus and the left middle occipital gyrus (roughly corresponding to the posterior occipito-temporal cortex) are structurally and functionally connected to each other. Via the precuneus, the left TPJ is also connected indirectly to the right TPJ, and from this perspective, it is evident that the left hemispheric network found in our meta-analysis is linked to a right hemispheric homologue network. Of particular interest, the connectivity fingerprint for the left TPJ area found in our conjunction analysis shows that this area is linked both to fronto-parietal

areas (lateral prefrontal cortex, superior parietal lobe) which we only found in our meta-analysis of visual perspective taking, and to anterior temporal areas which we only found in our meta-analysis on false belief reasoning. It is tempting to speculate that this may reflect how a domain general function—processing of a perspective difference—can be applied to different problems (social versus spatial). However, direct evidence from task-based functional connectivity studies is needed to justify such a claim.

CONCLUSION

To identify brain areas which are commonly engaged in social and visuo-spatial perspective taking, we performed a meta-analysis on false belief reasoning and visual perspective taking. False belief is a case of social cognition that requires processing of a perspective difference to understand someone's mistaken belief about the world which contrasts with reality. We found common activation for false belief reasoning and visual perspective taking in the left but not right dorsal TPJ. This fits with the idea that the left dorsal TPJ is responsible for representing different perspectives in a domain-general fashion (e.g., Perner and Leekam, 2008). In addition, we found common activation for false belief reasoning in the precuneus and the left middle occipital gyrus. Common activation in the precuneus can be linked to mental imagery which may support both social and visuo-spatial scene construction, whereas common activation in the left middle occipital gyrus—falling into the EBA—can be linked to imagining a change in one's body position in order to get another's point of view.

AUTHOR CONTRIBUTIONS

Matthias Schurz, Josef Perner and Markus Aichhorn designed and planned this work. Matthias Schurz and Anna Martin implemented the meta-analysis. Matthias Schurz and Josef Perner wrote the manuscript.

ACKNOWLEDGEMENTS

Matthias Schurz would like to thank Fabio Richlan for sharing his methodological expertise in the field of meta-analysis.

REFERENCES

- Abraham, A., Rakoczy, H., Werning, M., von Cramon, D. Y., and Schubotz, R. I. (2010). Matching mind to world and vice versa: functional dissociations between belief and desire mental state processing. *Soc. Neurosci.* 5, 1–18. doi: 10.1080/17470910903166853
- Adams, R. B. Jr., Rule, N. O., Franklin, R. G. Jr., Wang, E., Stevenson, M. T., Yoshikawa, S., et al. (2010). Cross-cultural reading the mind in the eyes: an fMRI investigation. *J. Cogn. Neurosci.* 22, 97–108. doi: 10.1162/jocn.2009.21187
- Aichhorn, M., Perner, J., Kronbichler, M., Staffen, W., and Ladurner, G. (2006). Do visual perspective tasks need theory of mind? *Neuroimage* 30, 1059–1068. doi: 10.1016/j.neuroimage.2005.10.026
- Aichhorn, M., Perner, J., Weiss, B., Kronbichler, M., Staffen, W., and Ladurner, G. (2009). Temporoparietal junction activity in theory-of-mind tasks: falseness, beliefs, or attention. *J. Cogn. Neurosci.* 21, 1179–1192. doi: 10.1162/jocn.2009.21082
- Astafiev, S. V., Stanley, C. M., Shulman, G. L., and Corbetta, M. (2004). Extrastriate body area in human occipital cortex responds to the performance of motor actions. *Nat. Neurosci.* 7, 542–548. doi: 10.1038/nn1241
- Bahnemann, M., Dziobek, I., Prehn, K., Wolf, I., and Heekeren, H. R. (2010). Sociotopy in the temporoparietal cortex: common versus distinct processes. *Soc. Cogn. Affect. Neurosci.* 5, 48–58. doi: 10.1093/scan/nsp045
- Baron-Cohen, S., Ring, H. A., Wheelwright, S., Bullmore, E. T., Brammer, M. J., Simmons, A., et al. (1999). Social intelligence in the normal and autistic brain: an fMRI study. *Eur. J. Neurosci.* 11, 1891–1898. doi: 10.1046/j.1460-9568.1999.00621.x
- Blakemore, S. J., Boyer, P., Pachot-Clouard, M., Meltzoff, A., Segebarth, C., and Decety, J. (2003). The detection of contingency and animacy from simple animations in the human brain. *Cereb. Cortex* 13, 837–844. doi: 10.1093/cercor/13.8.837
- Brüne, M., Lissek, S., Fuchs, N., Witthaus, H., Peters, S., Nicolas, V., et al. (2008). An fMRI study of theory of mind in schizophrenic patients with “passivity” symptoms. *Neuropsychologia* 46, 1992–2001. doi: 10.1016/j.neuropsychologia.2008.01.023
- Brunet, E., Sarfati, Y., Hardy-Baylé, M. C., and Decety, J. (2000). A PET investigation of the attribution of intentions with a nonverbal task. *Neuroimage* 11, 157–166. doi: 10.1006/nimg.1999.0525
- Bzdok, D., Schillbach, L., Vogeley, K., Schneider, K., Laird, A. R., Langner, R., et al. (2012). Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. *Brain Struct. Funct.* 217, 783–796. doi: 10.1007/s00429-012-0380-y
- Caspers, S., Eickhoff, S. B., Geyer, S., Scheperjans, F., Mohlberg, H., Zilles, K., et al. (2008). The human inferior parietal lobule in stereotaxic space. *Brain Struct. Funct.* 212,

- 481–495. doi: 10.1007/s00429-008-0195-z
- Caspers, S., Eickhoff, S. B., Rick, T., von Kapri, A., Kühlen, T., Huang, R., et al. (2011). Probabilistic fibre tract analysis of cytoarchitecturally defined human inferior parietal lobule areas reveals similarities to macaques. *Neuroimage* 58, 362–380. doi: 10.1016/j.neuroimage.2011.06.027
- Caspers, S., Geyer, S., Schleicher, A., Mohlberg, H., Amunts, K., and Zilles, K. (2006). The human inferior parietal cortex: cytoarchitectonic parcellation and interindividual variability. *Neuroimage* 33, 430–448. doi: 10.1016/j.neuroimage.2006.06.054
- Castelli, F., Happé, F., Frith, U., and Frith, C. (2000). Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage* 12, 314–315. doi: 10.1006/nimg.2000.0612
- Castelli, I., Baglio, F., Blasi, V., Alberoni, M., Falini, A., Liverta-Sempio, O., et al. (2010). Effects of aging on mindreading ability through the eyes: an fMRI study. *Neuropsychologia* 48, 2586–2594. doi: 10.1016/j.neuropsychologia.2010.05.005
- Cavanna, A. E., and Trimble, M. R. (2006). The precuneus: a review of its functional anatomy and behavioural correlates. *Brain* 129, 564–583. doi: 10.1093/brain/awl004
- Craik, F. I. M., Moroz, T. M., Moscovitch, M., Stuss, D. T., Winocur, G., Tulving, E., et al. (1999). In search of the self: a positron emission tomography study. *Psychol. Sci.* 10, 26–34. doi: 10.1111/1467-9280.00102
- Creem, S. H., Downs, T. H., Wraga, M., Harrington, G. S., Proffitt, D. R., and Downs, J. H. 3rd. (2001). An fMRI study of imagined self-rotation. *Cogn. Affect. Behav. Neurosci.* 1, 239–249. doi: 10.3758/CABN.1.3.239
- Creem-Regehr, S. H., Neil, J. A., and Yeh, H. J. (2007). Neural correlates of two imagined egocentric transformations. *Neuroimage* 35, 916–927. doi: 10.1016/j.neuroimage.2006.11.057
- Das, P., Lagopoulos, J., Coulston, C. M., Henderson, A. F., and Malhi, G. S. (2012). Mentalizing impairment in schizophrenia: a functional MRI study. *Schizophr. Res.* 134, 158–164. doi: 10.1016/j.schres.2011.08.019
- David, N., Aumann, C., Santos, N. S., Bewernick, B. H., Eickhoff, S. B., Newen, A., et al. (2008). Differential involvement of the posterior temporal cortex in mentalizing but not perspective taking. *Soc. Cogn. Affect. Neurosci.* 3, 279–289. doi: 10.1093/scan/nsn023
- David, N., Bewernick, B. H., Cohen, M. X., Newen, A., Lux, S., Fink, G. R., et al. (2006). Neural representations of self versus other: visual-spatial perspective taking and agency in a virtual ball-tossing game. *J. Cogn. Neurosci.* 18, 898–910. doi: 10.1162/jocn.2006.18.6.898
- Deen, B., and McCarthy, G. (2010). Reading about the actions of others: biological motion imagery and action congruency influence brain activity. *Neuropsychologia* 48, 1607–1615. doi: 10.1016/j.neuropsychologia.2010.01.028
- Denny, B. T., Kober, H., Wager, T. D., and Ochsner, K. N. (2012). A meta-analysis of functional neuroimaging studies of self- and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *J. Cogn. Neurosci.* 24, 1742–1752. doi: 10.1162/jocn_a_00233
- Derrfuss, J., Brass, M., Neumann, J., and von Cramon, D. Y. (2005). Involvement of the inferior frontal junction in cognitive control: meta-analyses of switching and Stroop studies. *Hum. Brain Mapp.* 25, 22–34. doi: 10.1002/hbm.20127
- DerSimonian, R., and Laird, N. (1986). Meta-analysis in clinical trials. *Control. Clin. Trials* 7, 177–188. doi: 10.1016/0197-2456(86)90046-2
- Dodell-Feder, D., Koster-Hale, J., Bedny, M., and Saxe, R. (2011). fMRI item analysis in a theory of mind task. *Neuroimage* 55, 705–712. doi: 10.1016/j.neuroimage.2010.12.040
- Downing, P. E., Jiang, Y., Shuman, M., and Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science* 293, 2470–2473. doi: 10.1126/science.1063414
- Dumontheil, I., Küster, O., Apperly, I. A., and Blakemore, S. J. (2010). Taking perspective into account in a communicative task. *Neuroimage* 52, 1574–1583. doi: 10.1016/j.neuroimage.2010.05.056
- Flavell, J. H., Everett, B. A., Croft, K., and Flavell, E. R. (1981). Young children's knowledge about visual perception: further evidence for the Level 1 – Level 2 distinction. *Dev. Psychol.* 17, 99–103. doi: 10.1037/0012-1649.17.1.99
- Fletcher, P. C., Happé, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S., et al. (1995). Other minds in the brain: a functional imaging study of “theory of mind” in story comprehension. *Cognition* 57, 109–128. doi: 10.1016/0010-0277(95)00692-R
- Focquaert, F., Steven-Wheeler, M. S., Vanneste, S., Doron, K. W., and Platek, S. M. (2010). Mindreading in individuals with an empathizing versus systemizing cognitive style: an fMRI study. *Brain Res. Bull.* 83, 214–222. doi: 10.1016/j.brainresbull.2010.08.008
- Gallagher, H. L., and Frith, C. D. (2003). Functional imaging of ‘theory of mind’. *Trends Cogn. Sci.* 7, 77–83. doi: 10.1016/S1364-6613(02)00025-6
- Gallagher, H. L., Happé, F., Brunswick, N., Fletcher, P. C., Frith, U., and Frith, C. D. (2000). Reading the mind in cartoons and stories: an fMRI study of ‘theory of mind’ in verbal and nonverbal tasks. *Neuropsychologia* 38, 11–21. doi: 10.1016/S0028-3932(99)00053-6
- Ghaem, O., Mellet, E., Crivello, F., Tzourio, N., Mazoyer, B., Berthoz, A., et al. (1997). Mental navigation along memorized routes activates the hippocampus, precuneus, and insula. *Neuroreport* 8, 739–744. doi: 10.1097/00001756-199702100-00032
- Gobbini, M. I., Koralek, A. C., Bryan, R. E., Montgomery, K. J., and Haxby, J. V. (2007). Two takes on the social brain: a comparison of theory of mind tasks. *J. Cogn. Neurosci.* 19, 1803–1814. doi: 10.1162/jocn.2007.19.11.1803
- Hamilton, A. F., Brindley, R., and Frith, U. (2009). Visual perspective taking impairment in children with autistic spectrum disorder. *Cognition* 113, 37–44. doi: 10.1016/j.cognition.2009.07.007
- Hanakawa, T., Immisch, I., Toma, K., Dimyan, M. A., Van Gelderen, P., and Hallett, M. (2003). Functional properties of brain areas associated with motor execution and imagery. *J. Neurophysiol.* 89, 989–1002. doi: 10.1152/jn.00132.2002
- Happé, F., Ehlers, S., Fletcher, P., Frith, U., Johansson, M., Gillberg, C., et al. (1996). ‘Theory of mind’ in the brain. Evidence from a PET scan study of Asperger syndrome. *Neuroreport* 8, 197–201. doi: 10.1097/00001756-199612200-00040
- Heider, F., and Simmel, M. (1944). An experimental study of apparent behavior. *Am. J. Psychol.* 57, 243–259. doi: 10.2307/1416950
- Hynes, C. A., Baird, A. A., and Grafton, S. T. (2006). Differential role of the orbital frontal lobe in emotional versus cognitive perspective-taking. *Neuropsychologia* 44, 374–383. doi: 10.1016/j.neuropsychologia.2005.06.011
- Iseki, K., Hanakawa, T., Shinozaki, J., Nankaku, M., and Fukuyama, H. (2008). Neural mechanisms involved in mental imagery and observation of gait. *Neuroimage* 41, 1021–1031. doi: 10.1016/j.neuroimage.2008.03.010
- Jimura, K., Konishi, S., Asari, T., and Miyashita, Y. (2010). Temporal pole activity during understanding other persons’ mental states correlates with neuroticism trait. *Brain Res.* 1328, 104–112. doi: 10.1016/j.brainres.2010.03.016
- Kaiser, S., Walther, S., Nennig, E., Kronmüller, K., Mundt, C., Weisbrod, M., et al. (2008). Gender-specific strategy use and neural correlates in a spatial perspective taking task. *Neuropsychologia* 46, 2524–2531. doi: 10.1016/j.neuropsychologia.2008.04.013
- Kana, R. K., Keller, T. A., Cherkassky, V. L., Minshew, N. J., and Just, M. A. (2009). Atypical frontal-posterior synchronization of theory of mind regions in autism during mental state attribution. *Soc. Neurosci.* 4, 135–152. doi: 10.1080/17470910802198510
- Kliemann, D., Young, L., Scholz, J., and Saxe, R. (2008). The influence of prior record on moral judgment. *Neuropsychologia* 46, 2949–2957. doi: 10.1016/j.neuropsychologia.2008.06.010
- Knauff, M., Fangmeier, T., Ruff, C. C., and Johnson-Laird, P. N. (2003). Reasoning, models, and images: behavioral measures and cortical activity. *J. Cogn. Neurosci.* 15, 559–573. doi: 10.1162/0899290321662949
- Kockler, H., Scheef, L., Tepest, R., David, N., Bewernick, B. H., Newen, A., et al. (2010). Visuospatial perspective taking in a dynamic environment: perceiving moving objects from a first-person-perspective induces a disposition to act. *Conscious. Cogn.* 19, 690–701. doi: 10.1016/j.concog.2010.03.003
- Kobayashi, C., Glover, G. H., and Temple, E. (2006). Cultural and linguistic influence on neural bases of ‘theory of mind’: an fMRI study with Japanese bilinguals. *Brain Lang.* 98, 210–220. doi: 10.1016/j.bandl.2006.04.013
- Kobayashi, C., Glover, G. H., and Temple, E. (2007). Children’s and adults’ neural bases of verbal and nonverbal ‘theory of mind’. *Neuropsychologia* 45, 1522–1533. doi: 10.1016/j.neuropsychologia.2006.11.017

- Lambrey, S., Doeller, C., Berthoz, A., and Burgess, N. (2012). Imagining being somewhere else: neural basis of changing perspective in space. *Cereb. Cortex* 22, 166–174. doi: 10.1093/cercor/bhr101
- Lancaster, J. L., Tordesillas-Gutierrez, D., Martinez, M., Salinas, F., Evans, A., Zilles, K., et al. (2007). Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. *Hum. Brain Mapp.* 28, 1194–1205. doi: 10.1002/hbm.20345
- Lee, J., Quintana, J., Nori, P., and Green, M. F. (2011). Theory of mind in schizophrenia: exploring neural mechanisms of belief attribution. *Soc. Neurosci.* 6, 569–581. doi: 10.1080/17470919.2011.620774
- Lou, H. C., Luber, B., Crupain, M., Keenan, J. P., Nowak, M., Kjaer, T. W., et al. (2004). Parietal cortex and representation of the mental self. *Proc. Natl. Acad. Sci. U.S.A.* 101, 6827–6832. doi: 10.1073/pnas.0400049101
- Ma, N., Vanderkerckhove, M., van Overwalle, F., Seurinck, R., and Fias, W. (2011). Spontaneous and intentional trait inferences recruit a common mentalizing network to a different degree: spontaneous inferences activate only its core areas. *Soc. Neurosci.* 6, 123–138. doi: 10.1080/17470919.2010.485884
- Mar, R. A. (2011). The neural bases of social cognition and story comprehension. *Annu. Rev. Psychol.* 62, 103–134. doi: 10.1146/annurev-psych-120709-145406
- Mazzarella, E., Ramsey, R., Conson, M., and Hamilton, A. (2013). Brain systems for visual perspective taking and action perception. *Soc. Neurosci.* 8, 248–267. doi: 10.1080/17470919.2012.761160
- Masangkay, Z. S., McCluskey, K. A., McIntyre, C. W., Sims-Knight, J., Vaughn, B. E., and Flavell, J. H. (1974). Early development of inferences about visual percepts of others. *Child Dev.* 45, 357–366. doi: 10.2307/1127956
- McCleery, J. P., Surtees, A. D., Graham, K. A., Richards, J. E., and Apperly, I. A. (2011). The neural and cognitive time course of theory of mind. *J. Neurosci.* 31, 12849–12854. doi: 10.1523/JNEUROSCI.1392-11.2011
- Mitchell, J. P. (2008). Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cereb. Cortex* 18, 262–271. doi: 10.1093/cercor/bhm051
- Mitchell, J. P., Heatherton, T. F., and Macrae, C. N. (2002). Distinct neural systems subserve person and object knowledge. *Proc. Natl. Acad. Sci. U.S.A.* 99, 15238–15243. doi: 10.1073/pnas.232395699
- Moor, B. G., Macks, Z. A., Gueroglu, B., Rombouts, S. A., Molen, M. W., and Crone, E. A. (2012). Neurodevelopmental changes of reading the mind in the eyes. *Soc. Cogn. Affect. Neurosci.* 7, 44–52. doi: 10.1093/scan/nsr020
- Murphy, E. R., Brent, B. K., Benton, M., Pruitt, P., Diwadkar, V., Rajarethinam, R. P., et al. (2010). Differential processing of metacognitive evaluation and the neural circuitry of the self and others in schizophrenia: a pilot study. *Schizophr. Res.* 116, 252–258. doi: 10.1016/j.schres.2009.11.009
- Murray, R. J., Schaer, M., and Debbané, M. (2012). Degrees of separation: a quantitative neuroimaging meta-analysis investigating self-specificity and shared neural activation between self- and other-reflection. *Neurosci. Biobehav. Rev.* 36, 1043–1059. doi: 10.1016/j.neubiorev.2011.12.013
- Nieminen-von Wendt, T., Metsähonkala, L., Kulomäki, T., Aalto, S., Autti, T., Vanhala, R., et al. (2003). Changes in cerebral blood flow in Asperger syndrome during theory of mind tasks presented by the auditory route. *Eur. Child Adolesc. Psychiatry* 12, 178–189. doi: 10.1007/s00787-003-0337-z
- Perner, J., Aichhorn, M., Kronbichler, M., Staffen, W., and Ladurner, G. (2006). Thinking of mental and other representations: the roles of left and right temporo-parietal junction. *Soc. Neurosci.* 1, 245–258. doi: 10.1080/17470910600989896
- Perner, J., Brandl, J. L., and Garnham, A. (2003). What is a perspective problem? Developmental issues in belief ascription and dual identity. *Facta Philosophica* 5, 355–378.
- Perner, J., and Leekam, S. (2008). The curious incident of the photo that was accused of being false: issues of domain specificity in development, autism, and brain imaging. *Q. J. Exp. Psychol.* 61, 76–89. doi: 10.1080/17470210701508756
- Perner, J., and Rössler, J. (2012). From infants' to children's appreciation of belief. *Trends Cogn. Sci.* 16, 519–525. doi: 10.1016/j.tics.2012.08.004
- Platel, H., Price, C., Baron, J. C., Wise, R., Lambert, J., Frackowiak, R. S., et al. (1997). The structural components of music perception. A functional anatomical study. *Brain* 120, 229–243. doi: 10.1093/brain/120.2.229
- Premack, D., and Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behav. Brain Sci.* 1, 515–526. doi: 10.1017/S0140525X00076512
- Radua, J., Mataix-Cols, D., Phillips, M. L., El-Hage, W., Kronhaus, D. M., Cardoner, N., et al. (2012). A new meta-analytic method for neuroimaging studies that combines reported peak coordinates and statistical parametric maps. *Eur. Psychiatry* 27, 605–611. doi: 10.1016/j.eurpsy.2011.04.001
- Radua, J., van den Heuvel, O. A., Surguladze, S., and Mataix-Cols, D. (2010). Meta-analytical comparison of voxel-based morphometry studies in obsessive-compulsive disorder vs. other anxiety disorders. *Arch. Gen. Psychiatry* 67, 701–711. doi: 10.1001/archgenpsychiatry.2010.70
- Ramsey, R., Hansen, P., Apperly, I., Samson, D. (2013). Seeing it my way or your way: frontoparietal brain areas sustain viewpoint-independent perspective selection processes. *J. Cogn. Neurosci.* 25, 670–684. doi: 10.1162/jocn_a_00345
- Ross, L. A., and Olson, I. R. (2010). Social cognition and the anterior temporal lobes. *Neuroimage* 49, 3452–3462. doi: 10.1016/j.neuroimage.2009.11.012
- Ruby, P., and Decety, J. (2003). What you believe versus what you think they believe? A neuroimaging study of conceptual perspective taking. *Eur. J. Neurosci.* 17, 2475–2480. doi: 10.1046/j.1460-9568.2003.02673.x
- Russell, T. A., Rubia, K., Bullmore, E. T., Soni, W., Suckling, J., Brammer, M. J., et al. (2000). Exploring the social brain in schizophrenia: left prefrontal underactivation during mental state attribution. *Am. J. Psychiatry* 157, 2040–2042. doi: 10.1176/appi.ajp.157.12.2040
- Saxe, R. (2006). Why and how to study theory of mind with fMRI. *Brain Res.* 1079, 57–65. doi: 10.1016/j.brainres.2006.01.001
- Saxe, R., Jamal, N., and Powell, L. (2005). My body or yours? The effect of visual perspective on cortical body representations. *Cereb. Cortex* 16, 178–182. doi: 10.1093/cercor/bhi095
- Saxe, R., and Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in “theory of mind”. *Neuroimage* 19, 1835–1842. doi: 10.1016/S1053-8119(03)00230-1
- Saxe, R., and Powell, L. (2006). It's the thought that counts: specific brain regions for one component of theory of mind. *Psychol. Sci.* 17, 692–699. doi: 10.1111/j.1467-9280.2006.01768.x
- Saxe, R., Schulz, L. E., and Jiang, Y. V. (2006). Reading minds versus following rules: dissociating theory of mind and executive control in the brain. *Soc. Neurosci.* 1, 284–298. doi: 10.1080/17470910601000446
- Saxe, R., and Wexler, A. (2005). Making sense of another mind: The role of the right temporo-parietal junction. *Neuropsychologia* 43, 1391–1399. doi: 10.1016/j.neuropsychologia.2005.02.013
- Thomas, N. J. T. (2010). “Mental imagery,” in *Stanford Encyclopedia of Philosophy*, ed Zalta, E. N. (Available online at: <http://plato.stanford.edu/entries/mental-imagery>, assessed 26.06.2013).
- Vincent, J. L., Kahn, I., Snyder, A. Z., Raichle, M. E., and Buckner, R. L. (2008). Evidence for a frontoparietal control system revealed by intrinsic functional connectivity. *J. Neurophysiol.* 100, 3328–3342. doi: 10.1152/jn.90355.2008
- Voellm, B. A., Taylor, A. N., Richardson, P., Corcoran, R., Stirling, J., McKie, S., et al. (2006). Neuronal correlates of theory of mind and empathy: a functional magnetic resonance imaging study in a nonverbal task. *Neuroimage* 29, 90–98. doi: 10.1016/j.neuroimage.2005.07.022
- Vogele, K., May, M., Ritzl, A., Falkai, P., Zilles, K., and Fink, G. R. (2004). Neural correlates of first-person perspective as one constituent of human self-consciousness. *J. Cogn. Neurosci.* 16, 817–827. doi: 10.1162/089892904970799
- Walter, H. W., Adenzato, M., Ciaramidaro, A., Enrici, I., Lorenzo, P., and Bara, B. G. (2004). Understanding intentions in social interaction: the role of the anterior paracingulate cortex. *J. Cogn. Neurosci.* 16, 1854–1863. doi: 10.1162/0898929042947838
- Wimmer, H., and Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13, 103–128. doi: 10.1016/0010-0277(83)90004-5
- Wraga, M., Boyle, H. K., Flynn, C. M. (2010). Role of motor processes in extrinsically encoding mental transformations. *Brain Cogn.* 74, 193–202. doi: 10.1016/j.bandc.2010.07.005
- Wraga, M., Shephard, J. M., Church, J. A., Inati, S., and Kosslyn, S. M. (2005). Imagined rotations of self versus objects: an fMRI study. *Neuropsychologia* 43, 1351–1361. doi: 10.1016/j.neuropsychologia.2004.11.028

- Young, L., Cushman, F., Hauser, M., and Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proc. Natl. Acad. Sci. U.S.A.* 104, 8235–8240. doi: 10.1073/pnas.0701408104
- Young, L., and Saxe, R. (2009). An fMRI investigation of spontaneous mental state inference for moral judgment. *J. Cogn. Neurosci.* 21, 1396–1405. doi: 10.1162/jocn.2009.21137
- Young, L., Dodell-Feder, D., and Saxe, R. (2010). What gets the attention of the temporo-parietal junction? An fMRI investigation of attention and theory of mind. *Neuropsychologia* 48, 2658–2664. doi: 10.1016/j.neuropsychologia.2010.05.012
- Young, L., Scholz, J., and Saxe, R. (2011). Neural evidence for “intuitive prosecution”: the use of mental state information for negative moral verdicts. *Soc. Neurosci.* 6, 302–315. doi: 10.1080/17470919.2010.529712
- Zacks, J. M., Vettel, J. M., and Michelon, P. (2003). Imagined viewer and object rotations dissociated with event-related fMRI. *J. Cogn. Neurosci.* 15, 1002–1018. doi: 10.1162/089892903770007399
- Zaitchik, D., Walker, C., Miller, S., LaViolette, P., Feczko, E., and Dickerson, B. C. (2010). Mental state attribution and the temporo-parietal junction: an fMRI study comparing belief, emotion, and perception. *Neuropsychologia* 48, 2528–2536. doi: 10.1016/j.neuropsychologia.2010.04.031
- Zhang, S., and Li, C. S. (2012). Functional connectivity mapping of the human precuneus by resting state fMRI. *Neuroimage* 59, 3548–3562. doi: 10.1016/j.neuroimage.2011.11.023
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Received: 15 July 2013; accepted: 07 October 2013; published online: 01 November 2013.
- Citation: Schurz M, Aichhorn M, Martin A and Perner J (2013) Common brain areas engaged in false belief reasoning and visual perspective taking: a meta-analysis of functional brain imaging studies. *Front. Hum. Neurosci.* 7:712. doi: 10.3389/fnhum.2013.00712
- This article was submitted to the journal *Frontiers in Human Neuroscience*. Copyright © 2013 Schurz, Aichhorn, Martin and Perner. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The economy of social resources and its influence on spatial perceptions

Elizabeth B. Gross* and Dennis Proffitt

Department of Psychology, University of Virginia, Charlottesville, VA, USA

Edited by:

Sarah H. Creem-Regehr, University of Utah, USA

Reviewed by:

Cordula Vesper, Central European University, Hungary

David A. Dunning, Cornell University, USA

*Correspondence:

Elizabeth B. Gross, Department of Psychology, University of Virginia, PO Box 400400, Charlottesville, VA 22904, USA
e-mail: ebh7z@virginia.edu

Survival for any organism, including people, is a matter of resource management. To ensure survival, people necessarily budget their resources. Spatial perceptions contribute to resource budgeting by scaling the environment to an individual's available resources. Effective budgeting requires setting a balance of income and expenditures around some baseline value. For social resources, this baseline assumes that the individuals are embedded in their social network. A review of the literature supports the proposal that our visual perceptions vary based on the implicit budgeting of physical and social resources, where social resources, as they fluctuate relative to a baseline, can directly alter our visual perceptions.

Keywords: spatial perception, social baseline, social resources, visual perception, extraversion, attachment style

THE ECONOMY OF SOCIAL RESOURCES AND ITS INFLUENCE ON SPATIAL PERCEPTIONS

Conscious visual experience suggests that our perceptions simply mirror the environment as it is. However, visual perception varies with changes in our physical and social environments, which suggests that our visual system does not provide a geometrically accurate representation of our world, but rather one that is grounded in action capabilities and social influences (Proffitt and Linkenauger, 2013). Both physiological resources, for example, blood glucose, and social resources, such as supportive friends, can influence perceptions of the spatial environment (Proffitt, 2006; Schnall et al., 2008). Focusing on social influences on perception, the current paper offers the hypothesis that the magnitude of social resources is evaluated with respect to a baseline, and that visual perception will reflect variations around this baseline.

Research in behavioral ecology suggests cost-benefit resource analyses predict individual behavior. In order to determine costs and benefits in an economy of action, resources must be evaluated relative to a baseline, defined as a value around which costs and benefits are balanced. In the current paper, we review evidence that suggests our visual perceptions are scaled to both our physiological and social resources. Next, this paper introduces the concept of a social baseline as a reference value for evaluating social resources. Social Baseline Theory (SBT), introduced by Coan, Beckes, and colleagues (Beckes and Coan, 2011, 2012; Coan et al., in press), serves as a useful framework within which to construe social resources. We propose that a person's social baseline is set by the quality and breadth of their social network. Moreover, individual differences in attachment style and personality produce a unique social baseline for each individual. Variability in social environments will interact with the individual's baseline to produce fluctuations in social resources. This flux of social resources relative to a baseline will produce corresponding changes in

visual perception. Finally, we discuss how changes in the social environment also interact with individual baselines to produce changes in visual perception. This resource budgeting account derives, in part, from considerations of behavioral ecology and human physiology.

In order to evaluate and budget for the potential resource costs and benefits of an action, it is necessary to first determine a baseline, or the amount of resources the body will seek to maintain. Like a household budget, there is typically some desired positive value of savings around which income and expenditures are balanced. Rather than spending all of your income, some amount of monetary resources is protected. When the amount of savings dips below the baseline, resources are conserved by cutting unnecessary expenditures until the savings are restored. Alternatively, when the savings value is higher than the baseline, expenditures might increase.

The concept of a baseline in an economy of action is omnipresent in human physiology, for example in the maintenance of bodily glucose levels. There are multiple sources of energy in the body, but glucose, which exists as both blood glucose and as glycogen stores in the muscles and liver, is the main energy source for both our muscles and the brain (Benton et al., 1996). When blood glucose levels decline, glycogen stores are released into the bloodstream to restore glucose levels to baseline; likewise, when blood glucose levels rise above a baseline level, insulin is released and blood glucose is transported to and stored in the muscles and liver as glycogen (Benton et al., 1996). In other words, there exists a baseline level of blood glucose and, barring any medical disorders, the human body seeks to conserve this baseline, much like a thermostat.

For all animals, survival is a matter of budgeting physiological resources, where at the most basic level animals ultimately must not expend more calories than they consume. The field of

behavioral ecology elegantly demonstrates how animal behavior is predicted by models that optimize the cost-benefit ratio inherent in their actions. For example, eating larger prey means a higher caloric gain for a predator; however, eating bigger prey may also engender a higher cost. Shore crabs are sensitive to the cost-benefit ratio of prey size, and when shore crabs are given the opportunity to eat mussels of all sizes, their diet consists mostly of mussels affording the highest rate of caloric intake, not the largest mussels with the highest caloric value but hardest to crack open shells (Elner and Hughes, 1978). Such findings are abundant in the field of behavioral ecology, and they suggest that animals are sensitive to both the costs and benefits of their actions.

Cost-benefit analyses are also evident in visual perceptions. Research has shown that the visual system is sensitive to the costs and benefits of individuals' actions with respect to their bodily resources and social environments (Proffitt, 2006). One of the first studies to show a role of bioenergetics resources in spatial perception did so in the context of viewing hills (Proffitt et al., 1995). In virtually all circumstances, individuals overestimate the slants of hills. One striking anecdote of this phenomenon is to consider the streets in San Francisco. Even in pictures, these streets appear to be astronomically steep, but the steepest street in San Francisco is reportedly 17.5 degrees (Naylor and McBeath, 2008). The general overestimation of geographical slant was originally reported in the literature by Kammann (1967), but more recently it has been systematically studied by Proffitt and colleagues (Proffitt et al., 1995). They found that participants overestimated the slant of a 10 degree hill to be approximately 30 degrees when standing at the bottom of the hill (Proffitt et al., 1995), and the effect persists even when participants are allowed to view a cross-section of the hill (Proffitt et al., 2001). More importantly, participants who are physically fatigued, elderly, not physically fit, or encumbered with a heavy backpack¹ estimate the slant of a hill to be steeper than their counterparts, suggesting that perception varies with the effort and ability required to perform an action (Bhalla and Proffitt, 1999; Proffitt, 2006).

Research has found a direct physiological basis for overestimations in visual perception. Schnall et al. (2010) found that participants, who consumed a caloric drink that restored blood glucose levels following a cognitively depleting task, estimated the slant of a hill to be significantly less steep than those who had consumed a no-calorie drink after the depletion task. Additionally, Zadra et al. (2010) demonstrated that direct physiological measures of individual fitness predicted distance perception, in particular maximal aerobic capacity (VO₂) max at blood lactate threshold (the gold standard measure of physical fitness). Those who were more fit perceived targets as being closer than those who were less fit. These findings suggest that perception is influenced by the bioenergetic costs of acting on an extent relative to the amount of physical resources available in the body.

Moreover, research suggests visual perception is sensitive to *anticipated* resources and costs. Thirsty participants perceived a

bottle of water to be closer than non-thirsty participants (Balcetis and Dunning, 2010), and participants engaged in dieting perceived muffins to be larger in size than non-dieters (Van Koningsgruggen et al., 2011). Additionally, participants report that threatening objects, such as spiders, appear to be closer, larger (Vasey et al., 2012), and moving faster (Witt and Sugovic, 2013) than non-threatening objects. These findings suggest that perception also varies with motivations to acquire physiological resources and avoid threatening objects (Dunning and Balcetis, 2013; Riccio et al., 2013), presumably to facilitate acting on the environment (Witt and Sugovic, 2013). Collectively, the above studies demonstrate that the visual system also includes potential environmental benefits and costs in a cost-benefit analysis of resources.

Physical resources are not the only resources that people have at their disposal. As humans, we do not behave in isolation; rather, we function embedded in a social environment. People's ability to act in the environment is augmented if they have a friend or family member who will act on their behalf. Given that physiological potential influences perception, then the availability of social support provided by others should also influence visual perception. Indeed, there is evidence to support this claim. Schnall and colleagues (Schnall et al., 2008) demonstrated that participants who were either walking with or imagining a supportive friend gave lower slant estimates than participants who were walking alone or imagining a non-supportive friend. This has recently been extended to online social networking, where participants who browsed the Facebook profile of a supportive friend estimated the slope of a hill to be less steep than those who browsed the profile of a non-supportive friend (Faulkner and Clore, 2012). In an attempt to understand the mechanisms by which friends are influencing visual perception, Oishi et al. (2013) manipulated felt understanding between strangers and found that the participants who believed that the other participant understood their personality perceived a hill to be less steep than those who believed they were not understood.

Potential social costs also influence perception. Participants perceive aggressive male students to be standing closer than non-aggressive males (Cole et al., 2013) and threatening out-group members are perceived to be closer than non-threatening out-group members (Xiao and Bavel, 2012). Additionally, social resources can attenuate the effect of social costs. Following social rejection, participants report the interpersonal distance to accepting others to be closer than rejecting others (Knowles et al., 2013), and Harber and colleagues (Harber et al., 2011) report that psychosocial resources, such as self-worth, reduced perceived distance to threatening objects. Collectively, these findings indicate that social resources can function in a similar fashion to physiological resources, where social costs and benefits work to influence visual perception.

Coan and colleagues propose that "load sharing" is the mechanism by which social resources alter cognitive processes (Coan et al., in press). To successfully act in the environment, individuals must identify and solve a set number of problems. The social network allows individuals to offload problems, effectively reducing the cost of acting. An example from behavioral ecology clearly illustrates this mechanism. When feeding, ostriches must

¹There is criticism that the increase in slant estimates while wearing a heavy backpack is due to experimental demand characteristics (Durgin et al., 2010). While this is certainly a possibility, we do not feel that the support for this claim is convincing (Proffitt and Zadra, 2011).

simultaneously hunt for food and avoid predators. Hunting in groups allows the ostrich to offload the work of scanning for predators, resulting in more time to consume food than when feeding alone (Bertram, 1980). For the ostrich, hunting in groups does not increase the amount of available food, in fact it reduces it; rather, it increases the time spent foraging which more than offsets the cost of competing with others in the group. Similarly, in humans it is not that the presence of social support indicates a greater quantity of tangible resources. Instead, social support signifies the ability to offload work to the social network, which reduces the overall cost of acting in the environment.

The aforementioned principles regarding costs and benefits relative to a baseline value are applicable to a variety of ecological environments, including our social environment. Again, there exists research that suggests that, much like physical resources, our visual perceptions vary with changes in the social environment (Schnall et al., 2008; Harber et al., 2011; Faulkner and Clore, 2012; Knowles et al., 2013; Oishi et al., 2013). While there is an extensive literature on how costs and benefits are evaluated and maximized in human physiology, there is considerably less research investigating how social resources are evaluated. However, there is evidence that the concept of a baseline is paramount to evaluating social resources. In the social support literature, not receiving social support is most detrimental when support was expected, and receiving unexpected social support is more beneficial than receiving expected social support (Bergeman et al., 2010). That is, the costs and benefits of social support are evaluated relative to baseline expectations. What remains, then, is to define and determine the components that set the expected social baseline with which we evaluate our social resources.

One idea in particular, aptly named SBT, addresses this issue (Beckes and Coan, 2011, 2012; Coan et al., in press). For much of psychology, the unit of analysis is focused solely on the individual; the assumption being that the presence of social support adds resources to an otherwise self-sufficient individual. SBT asserts that the individual's default state is to assume social support. In other words, an individual's social baseline, by which an environment is determined to be costly or beneficial, includes the individual and part of their social network (Beckes and Coan, 2011). As social animals, people assume the presence of social support, which decreases the cost of acting by load sharing (Coan et al., in press). A person's social baseline assumes the presence of social support, and thus, to study an individual in isolation is to study someone whose resources are taxed.

However, just as variability exists in physiology across individuals, there exist differences in the social baselines of individuals. While almost all people function embedded in a social network, individuals will differ in the amount and quality of anticipated social resources. For the remainder of the paper, our attention turns to a discussion of the possible individual and situational differences that will interact to influence an individual's sense of social support. Based on the existing literature, we propose that individual differences, such as attachment style and personality traits, can set an individual's social baseline. Additionally, the state of the social network itself can vary. Differences that exist outside of the individual, for example the action capabilities of the friends within the network, can cause variations that interact

with the baseline of social support. Ultimately, we propose that these individual differences in social resources and the social environment should be reflected in visual perception.

SBT proposes that the individual's baseline resources are composed of both their own resources and those in their social network. We propose that social baselines vary across individuals and are determined, in part, by our early life experiences. In biology, studies in life history theory show that, across a wide range of organisms, nutritional deficits early in life are followed by an initial compensation that results in costly deficits later in life (Metcalfe and Monaghan, 2001). Variability in early life changes the organism's baseline to be lower such that, over time, they will show nutritional and growth deficits.

Similarly, in attachment style theory, variability in early life experiences in caregiver relationships will affect an individual's relationship styles well into adulthood (Bowlby, 1969). Children whose caregivers were attentive and responsive to their needs will develop a secure attachment style; they are comfortable and confident in their current relationships. On the other hand, if a child's primary caregiver responded inconsistently, the child will often develop an insecure or anxious attachment style. Insecurely attached individuals are concerned about the reliability and dependability of their current relationships (Ainsworth et al., 1978; Bartholomew and Horowitz, 1991). Similar to findings in biology, variability in early life relationships will negatively affect an individual's relationships over their lifetime.

The impact of attachment style is far reaching; attachment style also moderates the benefits of social support such that insecurely and anxiously attached individuals report less perceived social support; anxiously attached participants perceive supportive messages from their romantic partners to be less supportive (Collins and Feeney, 2004), and securely attached individuals that spent time in the presence of their romantic partners before a social stress task reported lower state anxiety levels than insecurely attached individuals (Ditzen et al., 2008). In sum, individuals that are more anxious about their relationships perceive that they have fewer social resources, and they benefit less from received social support. Presumably, these individuals regard supportive others as less reliable, rendering them unable to invest wholeheartedly in their social network.

A social baseline indicates the degree to which an individual incorporates others in their network of social resources. Individuals with a lower social baseline are more autonomous, meaning they are less likely to incorporate others as part of their resource pool. This value is independent of whether or not the individuals in their social network engender resources or costs. We propose that insecure and anxiously attached individuals' social baselines are set to a lower value. As a consequence, if the individuals that comprise a social network are particularly supportive, then insecurely and anxiously attached individuals will be less likely to utilize available social resources, a claim that is supported by research discussed above (Collins and Feeney, 2004; Ditzen et al., 2008). However, social relationships are dynamic, and at times the social network requires individuals to return a favor. In the instances where the social network is imposing a burden on the individual, anxiously and insecurely attached participants should be less burdened. That is, with a lower social baseline (indicating

more autonomy), an individual is also less likely to include the burdens of their social network into the total calculation of their costs.

In addition to attachment style, we expect that social baselines will also vary with individual differences in extraversion. According to Eysenck's personality theory, differences in arousal levels lead extraverts to seek out social contact and introverts to avoid social contact (Matthews and Gilliland, 1999). As a result, extraverts tend to have a larger social support network (Stokes, 1985; Cohen et al., 1997; Swickert et al., 2002) and report interacting more often with their social support network (Swickert et al., 2002). Additionally, extraverts are more likely to seek out social support (Amirkhan et al., 1995; Halamandaris and Power, 1999) and report more perceived and enacted social support than introverts (Swickert et al., 2002, 2010). Overall, extraverts report having more social resources and benefit more from social support, suggesting they are more inter-dependent; they may be more likely to include others' resources into their implicit assessment of their own costs and benefits. As such, we propose extraverts have a higher social baseline than introverts. Because the default state is to expect the support of a social network, extraverts have more assumed resources at their baseline than introverts. Of importance to note is that, due to their higher social baseline, extraverts incur more of a cost than introverts when they are called upon to support their social network.

Thus far, we have discussed the individual differences expected to produce higher or lower social baselines, namely, attachment style and extraversion. Another source of variability in social resources arises from the social network itself. As previously mentioned, those in the social network could either be an available resource or, depending on their capabilities, an added burden. For illustrative purposes, consider moving into a new apartment with a friend. Typically, the friend would share the load of carrying heavy boxes, rendering her a potential resource. However, suppose the friend has recently broken her leg. Now you are responsible for moving all of your and her personal belongings; your friend is now an added cost. The social baseline has remained the same, it includes your friend, but situational factors have drastically changed the impact on expected costs and benefits. In fact, altering the action capabilities of friends has been shown to mediate the effect of social support in visual perception. In a study by Doerrfeld et al. (2012), participants estimated the weight of boxes to be less heavy if a friend was helping, but not when the friend was present but physically impaired. In another study, participants playing pong estimated the speed of the ball to be traveling faster when it was more difficult for their partner to block the ball (Witt et al., 2012). As this research demonstrates, the capabilities of the social network are an important point to consider. With respect to SBT, it highlights that higher social baselines are not always better. Higher social baselines indicate that you are also more likely to incorporate the burdens of the network, resulting in times where a higher baseline results in an added cost. Therefore, the amount of total available social resources depends on both the social baseline in addition to the quality and capabilities of the social network itself.

The proposed conceptualization of social resources has several implications for visual perception. Our visual perceptions are

scaled to our physiological and social resources; as we accrue resources, distances appear closer and slants appear to be less steep, and vice versa (Proffitt, 2006; Schnall et al., 2008; Zadra et al., 2010). Social resources are evaluated relative to the individual's baseline, an indicator of the degree to which an individual includes others in their social network and the quality of the social relationship. When the social network is a resource, individuals with a higher baseline are more likely to include others as part of their evaluation of resources. In this case, individuals with a higher social baseline should perceive distances to be closer and slants to be less steep. Alternatively, when the social network is a burden, individuals with a higher social baseline will have an increase in their social costs, and their visual perceptions will reflect this increase such that distances appear to be farther and hills appear to be steeper. We propose that social baselines are determined, in part, by individual differences such as attachment style and extraversion. Extraverts and securely attached individuals have a higher social baseline compared to introverts and insecurely attached individuals. As a result, extraverts and securely attached individuals should perceive hills to be less steep and distances to appear closer relative to their peers, except when the social network is a burden. In that case, extraverts and securely attached individuals should perceive distances to be farther and hills to be less steep. In sum, the individual differences that reflect changes in the social baseline should also interact with the social network to produce changes in visual perceptions.

In conclusion, people adapt to and attempt to thrive in both social and physical environments, and studying individuals in isolation ignores a vital component of humans' ecological environment. Still, it is not simply that the presence of a friend is a guarantee of social resources. We propose social resources are evaluated in accordance with a baseline that varies with individual differences and with respect to the capabilities of the social network. Our visual perceptions reflect the implicit budgeting of physical and social resources. For social resources, fluctuations around the social baseline and variations in the state of the social network will cause corresponding changes in visual perception. Ultimately, this proposal prompts researchers to consider a more nuanced study of how social environments differentially impact visual perception.

REFERENCES

- Ainsworth, M. D. S., Blehar, M. C., Waters, E., and Wall, S. (1978). *Patterns of Attachment: A Psychological Study of the Strange Situation*. Oxford, England: Lawrence Erlbaum.
- Amirkhan, J. H., Risinger, R. T., and Swickert, R. J. (1995). Extraversion: a "hidden" personality factor in coping? *J. Pers.* 63, 189–212. doi: 10.1111/j.1467-6494.1995.tb00807.x
- Balcetis, E., and Dunning, D. (2010). Wishful seeing more desired objects are seen as closer. *Psychol. Sci.* 21, 147–152. doi: 10.1177/0956797609356283
- Bartholomew, K., and Horowitz, L. M. (1991). Attachment styles among young adults: a test of a four-category model. *J. Pers. Soc. Psychol.* 61, 226–244. doi: 10.1037/0022-3514.61.2.226
- Beckes, L., and Coan, J. A. (2011). Social baseline theory: the role of social proximity in emotion and economy of action. *Soc. Personal. Psychol. Compass* 5, 976–988. doi: 10.1111/j.1751-9004.2011.00400.x
- Beckes, L., and Coan, J. A. (2012). "Social baseline theory and the social regulation of emotion," in *The Science of the Couple*, eds L. Campbell, J. La Guardia, J. M. Olson and M. P. Zanna (Philadelphia: Psychology Press), 79–91.

- Benton, D., Parker, P. Y., and Donohoe, R. T. (1996). The supply of glucose to the brain and cognitive functioning. *J. Biosoc. Sci.* 28, 463–479. doi: 10.1017/S0021932000022537
- Bergeman, C. S., Braun, M. E., Scott, S. B., Baird, B. M., Montpetit, M. A., and Ong, A. D. (2010). Effects of daily stress on negative affect: do social interactions help or hinder? *Manuscript of the University of Notre Dame*. Retrieved at: http://www.nd.edu/~adalab/Documents/Bergeman_et_al_2010.pdf.
- Bertram, B. C. R. (1980). Vigilance and group size in ostriches. *Anim. Behav.* 28, 278–286. doi: 10.1016/S0003-3472(80)80030-3
- Bhalla, M., and Proffitt, D. R. (1999). Visual-Motor recalibration in geographical slant perception. *J. Exp. Psychol. Hum. Percept. Perform.* 25, 1076–1096. doi: 10.1037/0096-1523.25.4.1076
- Bowlby, J. (1969). *Attachment and Loss* (Vols. 1). New York, NY: Basic Books.
- Coan, J. A., Brown, C., and Beckes, L. (in press). “Our social baseline: the role of social proximity in economy of action,” Chapter to appear in *Nature and Formation of Social Connections: From Brain to Group*, eds M. Mikulincer and P. R. Shaver (Washington, DC: American Psychological Association Press).
- Cohen, S., Doyle, W. J., Skoner, D. P., Rabine, B. S., and Gwaltney, J. M. (1997). Social ties and susceptibility to the common cold. *JAMA* 277, 1940–1944. doi: 10.1001/jama.1997.03540480040036
- Cole, S., Balci, E., and Dunning, D. (2013). Affective signals of threat increase perceived proximity. *Psychol. Sci.* 24, 34–40. doi: 10.1177/0956797612446953
- Collins, N. L., and Feeney, B. C. (2004). Working models of attachment shape perceptions of social support: evidence from experimental and observational studies. *J. Pers. Soc. Psychol.* 87, 363–383. doi: 10.1037/0022-3514.87.3.363
- Ditzen, B., Schmidt, S., Strauss, B., Nater, U. M., Ehler, U., and Heinrichs, M. (2008). Adult attachment and social support interact to reduce psychological but not cortisol response to stress. *J. Psychosom. Res.* 64, 479–486. doi: 10.1016/j.jpsychores.2007.11.011
- Doerrfeld, A., Sebanz, N., and Shiffrar, M. (2012). Expecting to lift a box together makes the load look lighter. *Psychol. Res.* 76, 467–475. doi: 10.1007/s00426-011-0398-4
- Dunning, D., and Balci, E. (2013). Wishful seeing how preferences shape visual perception. *Curr. Dir. Psychol. Sci.* 22, 33–37. doi: 10.1177/0963721412463693
- Durgin, F. H., Hajnal, A., Li, Z., Tonge, N., and Stigliani, A. (2010). Palm boards are not action measures: an alternative to the two-systems theory of geographical slant perception. *Acta Psychol. (Amst)* 134, 182–197. doi: 10.1016/j.actpsy.2010.01.009
- Elnor, R. W., and Hughes, R. N. (1978). Energy maximization in the diet of the shore crab, *Carcinus maenas*. *J. Anim. Ecol.* 47, 103–116. doi: 10.2307/3925
- Faulkner, L. M., and Clore, G. (2012). Facebook friends with benefits: online social support and slant perception. (Honor’s Thesis) Retrieved from <http://www.virginia.edu/psychology/downloads/DMP%20Papers/DMP%202012/Faulkner.pdf>
- Halamandaris, K. F., and Power, K. G. (1999). Individual differences, social support and coping with the examination stress: a study of the psychosocial and academic adjustment of first year home students. *Pers. Individ. Dif.* 26, 665–685. doi: 10.1016/S0191-8869(98)00172-X
- Harber, K. D., Yeung, D., and Iacovelli, A. (2011). Psychosocial resources, threat, and the perception of distance and height: support for the resources and perception model. *Emotion* 11, 1080–1090. doi: 10.1037/a0023995
- Kammann, R. (1967). The overestimation of vertical distance and slope and its role in the moon illusion. *Percept. Psychophys.* 2, 585–589. doi: 10.3758/bf03210273
- Knowles, M. L., Green, A., and Weidel, A. (2013). Social rejection biases estimates of interpersonal distance. *Soc. Psychol. Personal. Sci.* doi: 10.1177/1948550613491972
- Matthews, G., and Gilliland, K. (1999). The personality theories of H. J. Eysenck and J. A. Gray: a comparative review. *Pers. Individ. Dif.* 26, 583–626. doi: 10.1016/S0191-8869(98)00158-5
- Metcalfe, N. B., and Monaghan, P. (2001). Compensation for a bad start: grow now, pay later? *Trends Ecol. Evol.* 16, 254–260. doi: 10.1016/S0169-5347(01)02124-3
- Naylor, Y. K., and McBeath, M. K. (2008). Gender differences in spatial perception of body tilt. *Percept. Psychophys.* 70, 199–207. doi: 10.3758/pp.70.2.199
- Oishi, S., Schiller, J., and Gross, E. B. (2013). Felt understanding and misunderstanding affect the perception of pain, slant, and distance. *Soc. Psychol. Personal. Sci.* 4, 259–266. doi: 10.1177/1948550612453469
- Proffitt, D. R. (2006). Embodied perception and the economy of action. *Perspect. Psychol. Sci.* 1, 110–122. doi: 10.1111/j.1745-6916.2006.00008.x
- Proffitt, D. R., Bhalla, M., Gossweiler, R., and Midgett, J. (1995). Perceiving geographical slant. *Psychon. Bull. Rev.* 2, 409–428. doi: 10.3758/bf03210980
- Proffitt, D. R., Creem, S. H., and Zosh, W. D. (2001). Seeing mountains in mole hills: geographical-slant perception. *Psychol. Sci.* 12, 418–423. doi: 10.1111/1467-9280.00377
- Proffitt, D. R., and Linkenauger, S. A. (2013). “Perception viewed as a phenotypic expression,” in *Action Science: Foundations of an Emerging Discipline*, eds W. Prinz, M. Beisert and A. Herwig (United States: MIT Press), 171–198.
- Proffitt, D. R., and Zadra, J. R. (2011). Explicit and motoric dependent measures of geographical slant are dissociable: a reassessment of the findings of Durgin, Hajnal, Li, Tonge, and Stigliani. (2010). *Acta Psychol. (Amst)* 138, 285–288. doi: 10.1016/j.actpsy.2010.06.003
- Riccio, M., Cole, S., and Balci, E. (2013). Seeing the expected, the desired, and the feared: influences on perceptual interpretation and directed attention. *Soc. Personal. Psychol. Compass* 7, 401–414. doi: 10.1111/spc3.12028
- Schnall, S., Harber, K. D., Stefanucci, J. K., and Proffitt, D. R. (2008). Social support and the perception of geographical slant. *J. Exp. Soc. Psychol.* 44, 1246–1255. doi: 10.1016/j.jesp.2008.04.011
- Schnall, S., Zadra, J. R., and Proffitt, D. R. (2010). Direct evidence for the economy of action: glucose and the perception of geographical slant. *Perception* 39, 464–482. NIHMSID: NIHMS359328 doi: 10.1068/p6445
- Stokes, J. P. (1985). The relation of social network and individual difference variables to loneliness. *J. Pers. Soc. Psychol.* 48, 981–990. doi: 10.1037/0022-3514.48.4.981
- Swickert, R. J., Rosentreter, C. J., Hittner, J. B., and Mushrush, J. E. (2002). Extraversion, social support processes, and stress. *Pers. Individ. Dif.* 32, 877–891. doi: 10.1016/S0191-8869(01)00093-9
- Swickert, R. J., Hittner, J. B., and Foster, A. (2010). Big five traits interact to predict perceived social support. *Pers. Individ. Dif.* 48, 736–741. doi: 10.1016/j.paid.2010.01.018
- Van Koningsgruggen, G. M., Stroebe, W., and Aarts, H. (2011). Through the eyes of dieters: biased size perception of food following tempting food primes. *J. Exp. Soc. Psychol.* 47, 293–299. doi: 10.1016/j.jesp.2010.10.012
- Vasey, M. W., Vilensky, M. R., Heath, J. H., Harbaugh, C. N., Buffington, A. G., and Fazio, R. H. (2012). It was as big as my head, I swear! biased spider size estimation in spider phobia. *J. Anxiety Disord.* 26, 20–24. doi: 10.1016/j.janxdis.2011.08.009
- Witt, J., and Sugovic, M. (2013). Spiders appear to move faster than non-threatening objects regardless of one’s ability to block them. *Acta Psychol. (Amst)* 143, 284–291. doi: 10.1016/j.actpsy.2013.04.011
- Witt, J. K., Sugovic, M., and Taylor, J. E. T. (2012). Action-specific effects in a social context: others’ abilities influence perceived speed. *J. Exp. Psychol. Hum. Percept. Perform.* 38, 715–725. doi: 10.1037/a0026261
- Xiao, Y. J., and Van Bavel, J. J. (2012). See your friends close and your enemies closer: social identity and identity threat shape the representation of physical distance. *Pers. Soc. Psychol. Bull.* 38, 959–972. doi: 10.1177/0146167212442228
- Zadra, J., Schnall, S., Weltman, A., and Proffitt, D. R. (2010). Direct physiological evidence for an economy of action: bioenergetics and the perception of spatial layout. *J. Vis.* 10:54. doi: 10.1167/10.7.54

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 01 July 2013; accepted: 27 October 2013; published online: 19 November 2013.

Citation: Gross EB and Proffitt D (2013) The economy of social resources and its influence on spatial perceptions. *Front. Hum. Neurosci.* 7:772. doi: 10.3389/fnhum.2013.00772

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2013 Gross and Proffitt. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Insight into others' minds: spatio-temporal representations by intrinsic frame of reference

Yanlong Sun^{1*} and Hongbin Wang^{2*}

¹ The University of Texas Health Science Center at Houston, Houston, TX, USA

² Center for Biomedical Informatics, Texas A&M University Health Science University, Houston, TX, USA

Edited by:

Klaus Kessler, Aston University, UK

Reviewed by:

Klaus Kessler, Aston University, UK
Christian Sorg, Klinikum rechts der
Isar der Technischen Universität
München, Germany

*Correspondence:

Yanlong Sun, The University of Texas
Health Science Center at Houston,
7000 Fannin Suite 600, Houston,
TX 77030, USA
e-mail: Yanlong.Sun@uth.tmc.edu;
Hongbin Wang, Center for Biomedical
Informatics, Texas A&M University
Health Science University, 2121 West
Holcombe Blvd., Suite 1109, Houston,
TX 77030, USA
e-mail: hwang@tamhsc.edu

Recent research has seen a growing interest in connections between domains of spatial and social cognition. Much evidence indicates that processes of representing space in distinct frames of reference (FOR) contribute to basic spatial abilities as well as sophisticated social abilities such as tracking other's intention and belief. Argument remains, however, that belief reasoning in social domain requires an innately dedicated system and cannot be reduced to low-level encoding of spatial relationships. Here we offer an integrated account advocating the critical roles of spatial representations in intrinsic frame of reference. By re-examining the results from a spatial task (Tamborello et al., 2012) and a false-belief task (Onishi and Baillargeon, 2005), we argue that spatial and social abilities share a common origin at the level of spatio-temporal association and predictive learning, where multiple FOR-based representations provide the basic building blocks for efficient and flexible partitioning of the environmental statistics. We also discuss neuroscience evidence supporting these mechanisms. We conclude that FOR-based representations may bridge the conceptual as well as the implementation gaps between the burgeoning fields of social and spatial cognition.

Keywords: theory of mind, false belief, spatial cognition, frame of reference, predictive learning

INTRODUCTION

Recent research has seen a growing interest in the connections between two disparate lines of investigations: spatial cognition that focuses on spatial and bodily representations, and, social cognition that examines the abilities of attributing other's intentions and beliefs, namely, theory of mind (TOM). Although researchers have learned much about the underlying mechanisms in each domain, there are still opposing perspectives and considerable conceptual gaps between the two domains. In particular, much contest revolves around the contribution of domain-specific spatial processing to domain-general TOM abilities.

At the center of the debate, is an apparent contradiction between the findings that human infants can pass false-belief tasks (e.g., holding an agent's belief about the original location of an object, which has been changed in the absence of the agent) and the general claim that children first understand false-beliefs at around 4 years of age (for reviews, see, Apperly and Butterfill, 2009; Perner et al., 2011; Frith and Frith, 2012). Some have suggested that sophisticated TOM inferences, as indicated by successfully performing the false-belief tasks, may evolve from a set of low-level encoding processes, for example, agent-object-location associations (Perner and Ruffman, 2005; Ruffman and Perner, 2005), identification of "external referent" (Perner et al., 2011), and, spatial perspective taking (Kessler and Rutherford, 2010; Kessler and Thomson, 2010). Yet other theorists have posited that beliefs are "invisible abstract entities" (Saxe, 2006), and that making inferences about other's beliefs requires a dedicated or innate system that cannot be accounted for by mere associations (Leslie, 2005;

Saxe and Wexler, 2005; Csibra and Southgate, 2006; Baillargeon et al., 2010).

In the present paper, we attempt to bridge the conceptual gaps between different perspectives by advocating an integrated account. We argue that a fundamental spatio-temporal association process, which is fraught in the domain of spatial cognition, is also essential in the domain of social cognition. At the computational level, spatio-temporal association is to extract statistical regularities from the task environment by detecting the correlations between representations of events over space and time. However, spatio-temporal association is not merely about matrices of associative weights that connect different representations in a static manner. Instead, it takes place over space and time through the lens of *predictive learning*. Recent advances in neuroscience suggest that – at both the algorithmic and neural architectural levels – it is not reward that drives learning *per se*, but the temporal discrepancy between actual and expected outcomes (Gerstner et al., 2012; O'Reilly et al., 2012). That is, the task environment constantly changes. At any moment, environmental statistics present themselves as multimodal inputs to the mind. By constantly comparing the observed and expected outcomes, the mind selectively re-encodes the raw environmental statistics and transforms them into a hierarchy of representations at different levels of abstraction, which eventually produce complex behaviors such as thought, language, and, intelligence (Hawkins and Blakeslee, 2004).

Our approach to understanding the process of spatio-temporal association utilizes frames of reference (FOR) as the building blocks of both spatial and social cognition. A growing body of

research has shown that FOR-based representations are not only behaviorally plausible but are also supported by the neurological structures in both human and animal brains. As spatio-temporal association re-encodes the environmental statistics by removing task-irrelevant variances (e.g., instability, noise), FOR-based representations provide a straightforward way of partitioning spatio-temporal variances. In addition, it has been a central contention that theory-of-mind abilities are subject to competing demands for efficient and flexible processing and require two distinct systems, “one that is efficient and inflexible and one that is flexible but cognitively demanding” (Apperly and Butterfill, 2009, p. 957). Instead of focusing on the distinction between different systems, we emphasize the common representations shared by different sets of abilities and mechanisms. We argue that when people perform spatial and social tasks, both efficiency and flexibility can emerge from the expectation-driven competition among multiple FOR-based representations.

INTRINSIC FRAME OF REFERENCE (IFOR) IN SPATIAL COGNITION

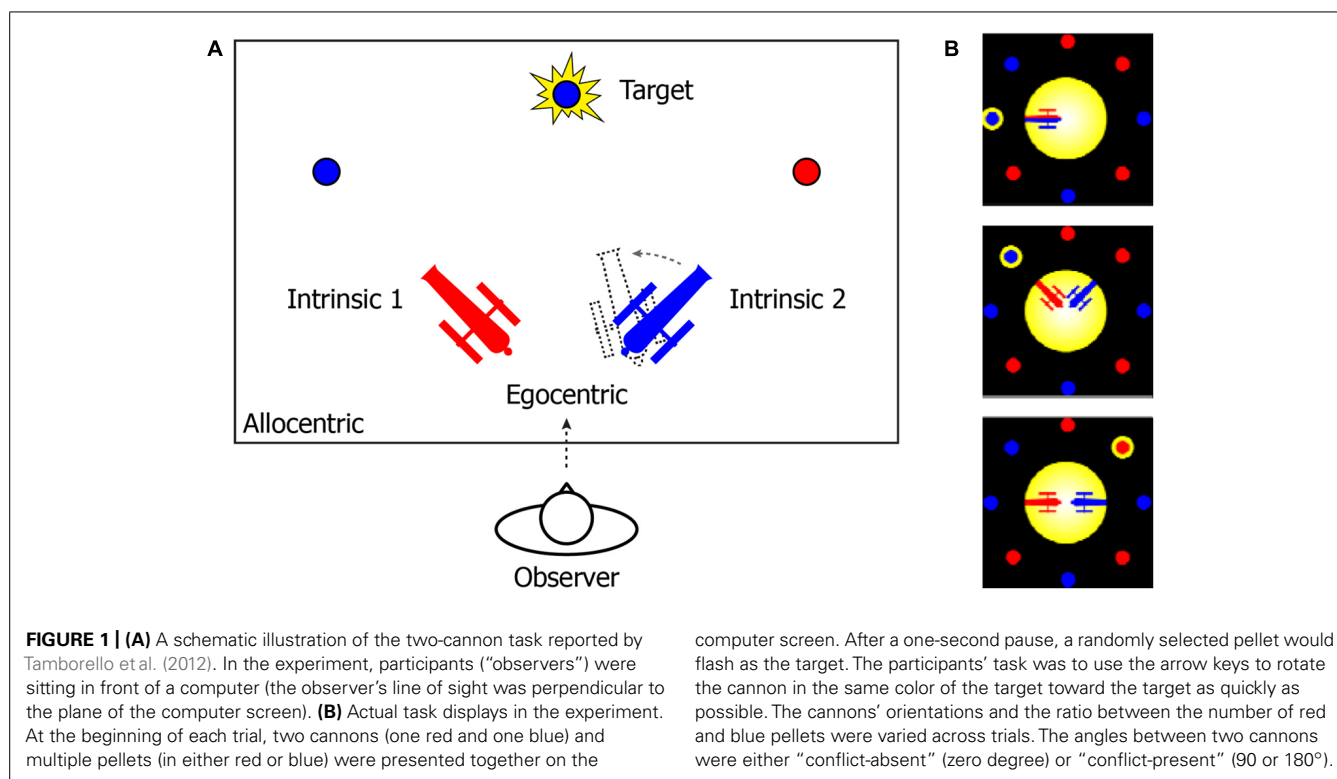
The notion of “FOR” has been crucial to all the disciplines that study spatial relationships and relies on a diverse terminology for its classification (Levinson, 2004). For example, a conventional approach is to classify a reference system by its origin: whether it is anchored to the observer self (e.g., “egocentric”) or the environment (e.g., “allocentric”; Andersen et al., 1997; Wang and Spelke, 2002; Burgess, 2008). However, we adopt a classification system that – besides the self-centered egocentric frame of reference (EFOR) – further differentiates the environment-centric frames into two categories: *allocentric* (AFOR, with an absolute and fixed anchor), and, *intrinsic* (IFOR, with a relative and flexible anchor). With roots in psycholinguistic research, the advantage of this classification scheme is that it reduces ambiguity in spatial descriptions of the world (Miller and Johnson-Laird, 1976; Carlson-Radvansky and Irwin, 1993; Levelt, 1996; Levinson, 2004; Carlson and Van Deman, 2008). For example, when describing the location of a coffee cup, one may say, “the cup is in front of me (observer self)” (in EFOR); “the cup is on the desk” (in AFOR); or “the cup is in front of John” (in IFOR). Note that, while both AFOR and IFOR use an external anchor, the anchor in AFOR (the desk in this case) is more stable than IFOR (John in this case, who can freely change his location or orientation). Our interest in IFOR is motivated by vision and spatial memory research that emphasizes the dynamic updating of object-centered representations (Marr, 1982; Wang et al., 2005a; Mou et al., 2008; Sun and Wang, 2010; Chen and McNamara, 2011). In this respect, the interactions between EFOR and IFOR (e.g., the intertwined representations of self-other-object relationship) are ubiquitous in everyday tasks, where the “other” can be either an anchoring object (Wang et al., 2005b; Tamborello et al., 2012), or another agent or human being as in social situations (Mitchell, 2006; Kessler and Rutherford, 2010; Kessler and Thomson, 2010; Perner et al., 2011).

One fundamental distinction among different FOR-based representations is the manner in which each representation handles *temporal instability* during the interactions between the mind and the environment. Temporal instability manifests itself as both spatial and temporal variances during the encoding of

spatio-temporal relationships between various entities in the environment (e.g., self, agents, objects, locations, and events). Different reference systems partition these variances in different manners and therefore afford structures at different levels of instability. In the “coffee cup” example, the spatial relations among relevant entities can change over time. To locate the coffee cup, an EFOR representation from the observer’s perspective is relatively stable, to the extent that the anchor is always the “observer self.” In contrast, an IFOR representation of the coffee cup anchored to John is unstable because John can freely move around and the observer is therefore required to track both the coffee cup and John in order to maintain an IFOR representation.

Critically, temporal instability evokes *predictive learning*. Simply put, whereas temporal instability means that the current input is expected to change at the next time point, predictive learning is a process of spatio-temporal integration in which the internal representation is constructed by remapping attention toward the expected outcomes (Hawkins and Blakeslee, 2004; O’Reilly et al., 2012). It has been suggested that predictive learning is a driving force in learning structured abstractions of the environment (Hawkins and Blakeslee, 2004; Krauzlis and Nummela, 2011; Rolfs et al., 2011; Gerstner et al., 2012; O’Reilly et al., 2012). Consider the coffee cup example again: predictive learning takes the anticipated movements into consideration and produces a dynamic representation of the relevant spatial relations. When an observer is reaching for a coffee cup, predictive learning occurs within EFORs, such that the coffee cup’s location is updated relative to the observer’s hand or body. By making constant predictions, the observer would know when to grab even before her hand touches the cup. When the observer watches John reaching for the coffee cup, predictive learning involves IFORs, such that the coffee cup’s location is updated relative to John. Yet, should John suddenly change his course and pick up another object (e.g., a stapler), the observer would be surprised as John’s initial movements led to an expectation that he would pick up the coffee cup instead of the stapler.

That the mind uses different FOR to manage temporal instability and drive spatio-temporal association is consistent with an accumulating body of neurological and behavioral studies (Marr, 1982; Krauzlis and Nummela, 2011; Pertzov et al., 2011; Rolfs et al., 2011; Van Der Werf et al., 2013). To further illustrate this notion, consider an example from the two-cannon experiment reported by Tamborello et al. (2012). In their experiment (**Figure 1**), participants were instructed to use the arrow keys to rotate the cannon in the same color of a to-be-revealed target as quickly as possible, so that the cannon could point to (and shoot at) the target. Three different types of reference systems can be used to describe the target location (**Figure 1A**). In an EFOR representation (relative to the observer), the target is at the front-top of the observer’s visual field (the observer’s line of sight was perpendicular to the plane of the computer screen). In an AFOR representation, the target can be described in reference to the computer screen frames. In an IFOR representation (relative to a cannon), the target has a counterclockwise bearing relative to the orientation of the blue cannon (or a clockwise bearing relative to the red cannon). Mathematically, all of these representations are equivalent, to the extent that one representation can be transformed into another without



losing any information. However, in terms of efficient and flexible removal of task-irrelevant variance, different representations are unique in the way they are updated and maintained.

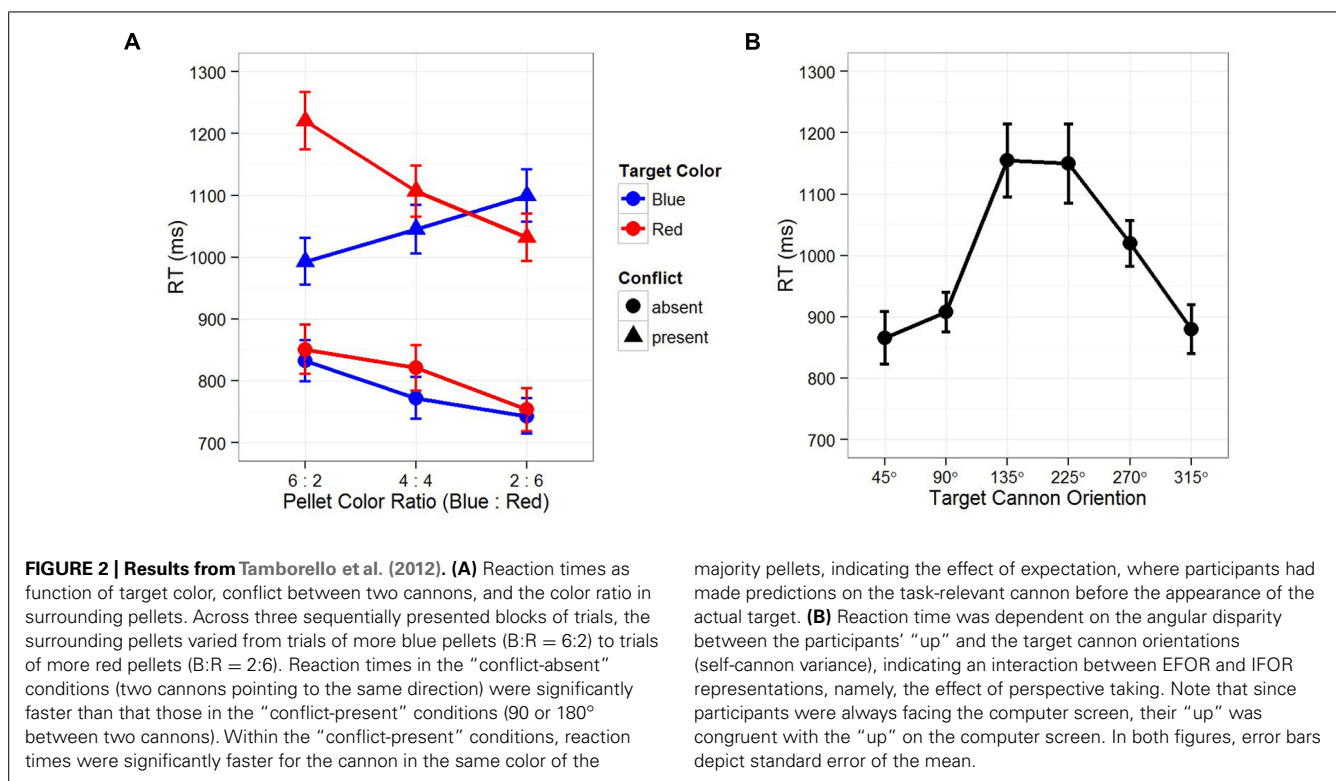
Let us first examine temporal instability. It is clear that both EFOR and AFOR representations have relatively fixed anchors (e.g., the observer and the computer monitor frames, respectively). In contrast, IFOR is only *tentatively* anchored to one of the two cannons: the color and location of the target is initially unknown, thus, which cannon is task-relevant depends on the visual input at the next time point. Recall that temporal instability evokes predictive learning, in which internal representations of the environment are constructed based on the current observations toward the expected future outcomes. In this case, the color ratio of the pellets provides a reliable cue for predicting the relevancy between two competing cannons. **Figure 2A** shows that reaction times in the conflict-present condition (cannons pointing to different directions) were significantly slower than those in the conflict-absent condition (cannons pointing to the same direction). Within the conflict-present conditions, the cannon in the same color of the majority pellets resulted in faster reaction times. These results indicate that in resolving the conflict between different IFOR representations, participants planned their responses by predicting the task-relevant cannon based on the pellet color ratio. That is, prediction occurs before the appearance of an actual target, leading to a stronger IFOR representation anchored to the task-relevant cannon, thus resulting in faster reaction times.

Second, in order to achieve computational efficiency and flexibility, multiple IFOR representations may coexist and interact with each other. **Figure 2A** shows that even when participants made correct predictions on the task-relevant cannon in the conflict-present

condition, their reaction times were still significantly slower than that in the conflict-absent condition. This indicates that, while anticipating the upcoming target, the competition between two conflicting IFOR representations resulted in a partial dissociation. That is, as the IFOR representation anchored to the predicted task-relevant cannon was the focus of attention, the other one was only partially disengaged – a strategy of prioritizing but still preparing for the unexpected. As a result, even when the prediction was correct, the partially disengaged IFOR representation interfered with performance and produce longer reaction times.

Third, an interaction may also occur between EFOR and IFOR representations. **Figure 2B** shows that reaction times were significantly dependent on the angular disparity between the self and cannon orientations, indicating a strategy of combining EFOR and IFOR representations, or *perspective taking*. Perspective taking has been considered as an important stepping stone from automatic and unaware perception toward a conscious and deliberate process in which people mentally perform a movement simulation of other people or objects (Kessler and Rutherford, 2010; Kessler and Thomson, 2010; Zwicker et al., 2011). Here, we consider perspective taking in terms of partitioning the statistical variances in the task environment.

Specifically, for a given cannon, we consider three parts of the spatial variances (angular disparities) that could be mentally encoded: self-cannon, self-target, and cannon-target. Since the correct response is determined by the cannon-target variance, it requires either a complete or a partial disengagement of the EFOR representation. If the EFOR representation is to be completely disengaged (i.e., removing self-target and self-cannon variances), the task could be accomplished by *object rotation* based



only on an IFOR representation. However, the reaction time pattern in **Figure 1B** suggests a case of partial EFOR disengagement: the task was accomplished by *self rotation with perspective taking*, in which the self-cannon variance was first removed so that the self-target variance became exactly the same as the cannon-target variance. Similar to the interaction between multiple IFOR representations, the interaction between EFOR and IFOR representations also serves the purpose of both computational efficiency and flexibility. On the one hand, an IFOR representation is parsimonious in encoding only task-specific variances (e.g., encoding only the target-cannon but not the self-cannon, the self-target relations). On the other hand, an EFOR representation tend to be automatic and effortless (Wang and Spelke, 2002; Frith and Frith, 2007; Kessler and Thomson, 2010). Therefore, an efficient and flexible solution would be to combine EFOR and IFOR representations into one representation. That is, instead of utilizing a purely IFOR-based strategy in which the cannon is mentally rotated toward the target (i.e., object rotation), participants might superimpose their egocentric perspective onto the cannon – that is, take the perspective of the cannon – then mentally self-rotate toward the target.

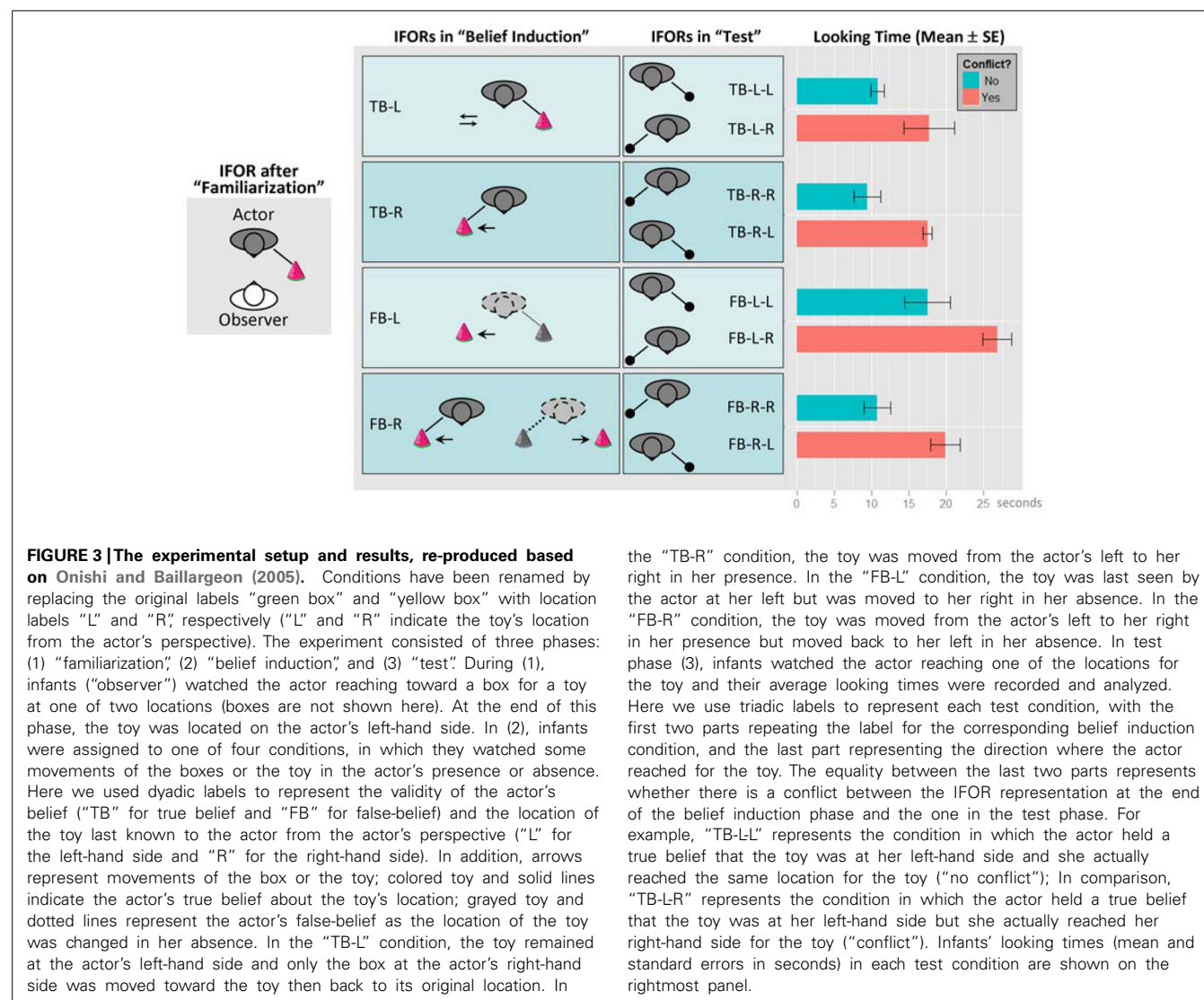
Overall, this new interpretation of the two-cannon experiment results suggests that expectation-driven competitions can take place not only between different IFOR representations (**Figure 2A**), but also between EFOR and IFOR representations (**Figure 2B**). By this account, the internal spatial representation of the environment is always dynamically constructed and updated toward the anticipated outcomes, rather than static associations of the current spatial configuration. Depending on whether there are conflicts between representations and whether the actual

outcome meets the expectation, competition takes place at different levels and results in the engagement and disengagement of different FOR-based representations. In the following section, we demonstrate that the same mechanisms may well lay the foundation for more complex representations in the domain of social cognition.

INTRINSIC FRAME OF REFERENCE IN BELIEF ATTRIBUTION

A landmark finding in belief attribution is that fifteen-month-old infants appear to be able to appeal to other's beliefs, that is, they were able to keep track of an actor's perception about the location of a toy, and, using this perception rather their own, to predict the actor's searching behavior (Onishi and Baillargeon, 2005). This finding has triggered a substantial debate over the question whether the theory-of-mind abilities evolved from “actor-object-location associations” (Perner and Ruffman, 2005, p. 215), or are due to an innate mechanism specialized for belief attribution (Leslie, 2005; Baillargeon et al., 2010). Here we offer a reinterpretation of the original findings based on the same spatio-temporal association account outlined above.

Figure 3 re-produces the experimental setup and results from Onishi and Baillargeon (2005). Note that we have re-labeled the experimental conditions by replacing the original object labels with location labels from the actor's perspectives: “green box” replaced by “L” (actor's left-hand side), and, “yellow box” replaced by “R” (actor's right-hand side). Hence, our new labels are essentially placeholders for representing different locations. However, the new labels also highlight the spatial component of the task environment and potential interference between the different FOR. Similar to the two-cannon experiment, this task involves the



the "TB-R" condition, the toy was moved from the actor's left to her right in her presence. In the "FB-L" condition, the toy was last seen by the actor at her left but was moved to her right in her absence. In the "FB-R" condition, the toy was moved from the actor's left to her right in her presence but moved back to her left in her absence. In test phase (3), infants watched the actor reaching one of the locations for the toy and their average looking times were recorded and analyzed. Here we use triadic labels to represent each test condition, with the first two parts repeating the label for the corresponding belief induction condition, and the last part representing the direction where the actor reached for the toy. The equality between the last two parts represents whether there is a conflict between the IFOR representation at the end of the belief induction phase and the one in the test phase. For example, "TB-L-L" represents the condition in which the actor held a true belief that the toy was at her left-hand side and she actually reached the same location for the toy ("no conflict"); In comparison, "TB-L-R" represents the condition in which the actor held a true belief that the toy was at her left-hand side but she actually reached her right-hand side for the toy ("conflict"). Infants' looking times (mean and standard errors in seconds) in each test condition are shown on the rightmost panel.

interplay of multiple representations. For example, the toy's location can be described in EFOR (relative to the observer, which is the infant in the experiment), AFOR (relative to the table or the room), or IFOR (relative to the actor). According to the original object labels, the toy's location was described by the color of the box, which was the same to both the infant and the actor. In contrast, as the infant was facing the actor, the "left" and "right" labels were completely opposite, depending on whether they were from the infant's perspective (EFOR) or from the actor's perspective (IFOR). Therefore, the new labels were more effective in distinguishing EFOR and IFOR representations.

COMPARISON WITHIN BELIEF INDUCTION CONDITIONS

The main finding by Onishi and Baillargeon (2005) involved comparing the infants' looking times between the two "test" conditions within each of the four "belief induction" conditions. They reported that looking times were shorter when the actor reached for the toy where she believed it was located ("no conflict" conditions in Figure 3) and longer when the actor reached the

opposite location ("conflict" conditions). Based on this comparison, the authors concluded that infants were able to use the actor's belief state instead of the actual toy location from infants' own perspective to predict the actor's reaching behavior.

Rather than resorting to an innately dedicated belief attribution mechanism, we would like to offer a different explanation based on fundamental spatial information processing mechanisms. Our interpretations is that belief attribution derives from the proper maintenance of and dissociation between multiple representations based on EFOR (for encoding self-toy or self-actor relations) and IFOR (for encoding actor-toy relations). In particular, it has been suggested that infants' looking time provides a measurement of surprise, such that longer looking times indicate greater violation of infants' expectations relative to their prior knowledge or greater novelty relative to their interpretation of habituation stimuli (Baillargeon, 1986; Onishi and Baillargeon, 2005; Téglás et al., 2011). Here we argue that for the false-belief task by Onishi and Baillargeon (2005) surprise might have resulted from the violation of infant's expected spatial configuration relative to the actual one.

Our earlier argument suggests that, among all possible FOR-based representations, those leading to task-relevant predictions tend to be actively updated and maintained. Since the looking times were about the actor's reaching for the toy, both the expected and actual spatial configurations would be encoded in the form of IFOR representations (actor-toy), rather than irrelevant EFOR representations (infant-toy). In other words, the IFOR-based expectation reflects a simple behavioral rule by means of spatial association – people (the actor) look for objects at their last known location (Ruffman and Perner, 2005). Consequently, the difference in looking times between “conflict” and “no conflict” conditions may be explained by the effort of resolving the discrepancy between the IFOR representation at the end of the belief induction phase, relative to the actual IFOR representation in the test phase. Results in **Figure 3** support this explanation by showing that, in each of the four belief conditions, looking times were reliably longer (with a mean difference always around 7–9 s) when there was a conflict between the IFOR representations at the end of the induction phase (the same as the expectation) and in the test phase (the actual outcome). For example, looking times for “x-L-R” conditions were consistently longer than those for “x-L-L” conditions (“x” stands for either “TB” or “FB”, and, a conflict is present if the last two alphabets are different).

COMPARISON BETWEEN BELIEF INDUCTION CONDITIONS

It is apparent from **Figure 3** that there were differences in looking times among the four belief induction conditions. For example, whereas the FB-L condition had the longest looking times, the FB-R condition had similar looking times as those in TB conditions. It is surprising that these differences were not mentioned nor accounted for by Onishi and Baillargeon (2005). Using the same argument in the two-cannon task, we speculate that the looking time difference between belief induction conditions might also be due to the interference from a partially disengaged representation. In this case, there could be different levels of the dissociation between EFOR and IFOR representations due to the different sequences of temporal events during the belief induction phase. Based on the comparison between “test” conditions above, it appears that the surprise effect (i.e., “conflict” versus “no conflict”) in all belief induction conditions remained approximately constant (7–9 s). This implies that the variance in looking times, less the surprise effect, would be independent of the predictions by the actor-toy IFOR representation. Accordingly, the remaining variance in looking times could be due solely to the interference from the infant-toy EFOR representation.

In the following, we use the conditional means and standard errors reported in the original study to make three sets of *post hoc* comparisons across different belief induction conditions but within the same “conflict” or “no conflict” test conditions (e.g., x-L-L compared with x-R-R, x-L-R compared with x-R-L, and etc.).

First, the mean looking times were about the same in the TB-L and TB-R conditions (i.e., TB-L-L \approx TB-R-R, and, TB-L-R \approx TB-R-L), despite different manipulation sequences in the belief induction phase – the former (TB-L) only involved the movement of an empty container (the “yellow box” on the actor's left hand side) and the latter (TB-R) involved the change of the toy's location (see **Figure 3**). This indicates that the looking times were primarily

determined by the active maintenance of the IFOR representation of the actor-toy relationship. If there was any interference from the EFOR representation of the infant-toy relationship, the effect remained constant between these two conditions.

Second, the mean looking times were significantly longer in the FB-L condition than in the TB-R condition (i.e., FB-L-L > TB-R-R, mean difference \approx 8 s; FB-L-R > TB-R-L, mean difference \approx 9 s; two tailed $p < 0.05$ in both comparisons). Such differences could be accounted for by stronger interference from the EFOR representation in the FB-L condition than in the TB-R condition. Specifically, the change of the toy's location was visible only to the infant in the FB-L condition but visible to both the infant and the actor in the TB-R condition. Thus, the infant-toy EFOR representation in the FB-L condition would be relatively stronger (more engaged). Being task-irrelevant (e.g., irrelevant to the actor's fetching action), the stronger EFOR representation in the FB-L condition would lead to greater interference, resulting in longer looking times during the test phase.

Third, the mean looking times were significantly shorter in the FB-R condition than in the FB-L condition (i.e., FB-L-L > FB-R-R, mean difference \approx 7 s; FB-L-R > FB-R-L, mean difference \approx 7 s; one tailed $p < 0.05$ in both comparisons). Interestingly, despite the more complicated manipulation sequences in the FB-R condition, looking times were about the same as those in the true belief conditions (TB-L and TB-R). Consistent with the aforementioned explanation, it is likely that the IFOR representation in the FB-R condition became stronger when it was reinforced in the presence of the actor (the actor last saw the toy moving to her right-hand side). By competition, a stronger IFOR representation led to a weaker EFOR representation. Although both were false-belief conditions, the weaker EFOR representation in the FB-R condition resulted in less interference and, therefore, shorter looking times than the FB-L condition.

In summary, it appears that FOR-based representations may provide a more transparent and detailed explanation to the findings reported by Onishi and Baillargeon (2005). In contrast to the two-cannon experiment by Tamborello et al. (2012), this false-belief task was not explicitly designed to detect the EFOR–IFOR interaction (e.g., infants were always facing the actor with the same bearing). Therefore, the interpretation of our *post hoc* comparisons between belief induction conditions could be limited. Nevertheless, our interpretation remained consistent across all comparisons and across both tasks. That is, in order to track and predict other agent's behavior, the internal process would involve at least a partial disengagement of EFOR representations, an active engagement of IFOR representations, and, potential interference between EFOR and IFOR representations.

Note that our interpretation is in the same vein as the “actor-object-location association” account (Perner and Ruffman, 2005). In addition, we identify the role of EFOR–IFOR dissociation. This interpretation is along the same line as the proposals that belief attribution may evolve from low-level spatial encoding processes, including the identification of “external referent” (Perner et al., 2011) and perspective taking (Kessler and Rutherford, 2010; Kessler and Thomson, 2010). Similar to the original interpretation by Onishi and Baillargeon (2005), here we also emphasize the role of expectation. However, expectation in our account is not

the end product of belief attribution. Rather, it starts early at the level of FOR-based spatial representations. In this respect, belief representation emerges as the mind integrates different spatial representations at different time points by reducing the discrepancy between the actual and the expected outcomes.

FROM SPATIAL TO SOCIAL: THE COMMON NON-COGNITIVE ORIGINS

Although we have demonstrated that the same language from spatial cognition may be used to interpret infants' performance in the false-belief task, we do not claim that social cognitive abilities can be completely accounted for by those in spatial cognition. Moreover, we do not claim a parallel between an explicit spatial orientation task and 15-month-old infants' preferential looking task. Rather, we focus on the common representations underlying these two seemingly different tasks. We argue that abilities from both spatial and social domains share common non-cognitive origins at the level of spatio-temporal association in extracting the environmental statistics. Ergo, these abilities, even if they appear different from each other, may not be domain-specific *per se*, but reflect the different requirements in computational efficiency and flexibility.

In bridging the conceptual gaps between spatial and social cognitive abilities, it is critical to understand the common dynamic nature of spatio-temporal association in both domains. In the present paper, we have shown that, in terms of FOR-based representations, the two-cannon task and the false-belief task share at least three computational properties. First, both tasks require encoding multiple spatial relations with different reference points (spatial association); Second, both involve comparisons of representations at different time points (temporal association); Third, the internal representations for both tasks are not static spatial encodings at isolated time points, rather, they are constructed and maintained through competitions toward the expected outcomes (predictive learning). We argue that all these three properties are governed by the same principle, whether one's goal is to learn a spatial configuration or infer other's intentions and beliefs. That is, the internal representations are developed in the direction of reducing spatio-temporal instability (variances) in order to extract statistical regularities at different levels of abstraction from the task environment.

Commonly shared computational processes could well be supported by commonly shared neural implementations. A growing body of research suggests that brain mechanisms supporting sophisticated social abilities may derive from low-level processes such as spatial tracking, predictive encoding, and attention shifting (for reviews, see, Mitchell, 2006; Corbetta et al., 2008; Frith and Frith, 2012). In the same vein, we argue that the key ingredient in both spatial and social cognition is the expectation-driven competition between multiple FOR-based representations, that are supported by a set of intrinsically distributed neural networks, rather than separately dedicated brain mechanisms. In the following, we discuss the neural evidence that supports this view.

Even a simple task could demand multiple representations of the task environment at different temporal points. Then, the need for selection arises at different levels of processing due to the limitation of resources. On the basis of functional and anatomical

distinctions, a model of attention selection has been proposed, suggesting that the attentional operations are carried out by the interactions between two fronto-parietal systems – a dorsal attention system (also referred to as top-down attention network, or, canonical sensory-motor pathway) and a ventral attention system (or, bottom-up attention network; Corbetta and Shulman, 2002; Corbetta et al., 2008; Yeo et al., 2011). The dorsal system is bilateral and mainly composed of the frontal eye field (FEF) and the intraparietal sulcus (IPS). It is specialized for selecting and linking stimuli and responses by sending top-down “filtering” signals to visual areas and via the middle frontal gyrus (MFG) to the ventral network. The ventral system is right-lateralized and includes the right temporal-parietal junction (TPJ), the right ventral frontal cortex (VFC), parts of the MFG, and the inferior frontal gyrus (IFG). Coordinated by the dorsal system, the ventral system sends bottom-up “reorienting” signals that interrupt and reset ongoing activity upon detection of salient targets, especially when there is a violation of expectation (for reviews, see, Corbetta et al., 2008).

The filtering and reorienting functionality in the dorsal–ventral attention networks is particularly useful for implementing the computation of multiple FOR-based representations, particularly when multiple FORs compete. We consider two levels of competition: (1) competition within the dorsal pathway (filtering), and (2), competition carried out by the interaction between the dorsal and ventral pathway (reorienting). Some evidence suggest that, along the dorsal pathway, multiple representations in different FOR can coexist – from lower-level retinotopic representations to higher-level self-centered (EFOR) and world-centered representations (IFOR and AFOR), and that the parietal cortex, particularly the IPS, is central to the construction of these representations (Marr, 1982; Andersen et al., 1997; Colby and Goldberg, 1999; Burgess, 2008; Pertzov et al., 2011; Van Der Werf et al., 2013). Recent rest-state data indicate that the dorsal attention network follows a serial and hierarchical organization, whereas the functional connectivity of parietal and prefrontal association cortices appears to be embedded with largely parallel and interdigitated circuits (Yeo et al., 2011). We argue that such an organization would allow a hierarchical abstraction of the task environment based on flexible selections among multiple representations. That is, in terms of FOR-based representations, it is possible that the invariance extracted at early cortical stages (e.g., visual areas and the parietal cortex) is incomplete, causing different representations to overlap with one another. In order to support higher-level abstractions, a more complete dissociation is required at the level of the prefrontal areas. For instance, it has been suggested that the FEF region plays a crucial role in the construction of intrinsic reference frames among multiple objects in spatial tasks (Wallentin, 2012). Likewise, studies with neural network simulations have shown that, although partial dissociation between different types of spatial information can occur by re-encoding visual information in the parietal cortex, dorsal control from the prefrontal cortex is necessary to achieve a more explicit dissociation (Sun and Wang, 2013); Moreover, efficient and flexible representations of the changing environment requires the maintenance of both latent representations (through altered firing thresholds in non-frontal regions) and active representations (through sustained firing in the prefrontal cortex) (Morton and Munakata, 2002). It is suggested

that such a maintenance mechanism is involved when the infants created actor-object-location associations in the false-belief task (Perner and Ruffman, 2005).

More dramatic competition between multiple representations would likely occur when expectations derived from actual sensory input have been violated. In such instances, the ventral attention network sends out reorienting signals and the dorsal attention network is reconfigured (Corbetta et al., 2008). Evidence for dorsal–ventral interaction comes from studies that use perspective taking tasks, which typically involve conflicting perspectives in EFOR and IFOR representations. For example, it has been reported that the transformation from participants' own perspective to another agent's body axis was associated with activations in posterior parietal cortical regions, such as the left inferior parietal lobe (IPL) and parietal–temporal–occipital junction as well as the right superior parietal lobe (Vogeley et al., 2004; David et al., 2006). Additionally, it has been found that TPJ shows enhanced activities in voluntary orienting of attention when participants are cued about the future location of a target stimulus (Corbetta et al., 2000), and when they need to distinguish between self-produced actions and actions generated by others (Blakemore and Frith, 2003; Jackson and Decety, 2004). Recently, Mazzarella et al. (2013) reported that responses in right IFG are sensitive to another person's orientation when participants perform the task from their own egocentric perspective. Thus, these studies are consistent with the suggestion that taking another person's perspective requires extra effort as compared with using one's own perspective (Kessler and Thomson, 2010).

It should be pointed out that among different brain areas, the TPJ region has been a major topic of debate regarding the neural mechanisms of belief attribution abilities in social interactions. Some researchers argue that this region is specifically involved in the theory-of-mind functions (Saxe and Kanwisher, 2003; Apperly et al., 2004; Saxe and Wexler, 2005; Saxe and Powell, 2006; Saxe et al., 2009; Young et al., 2010). However, the studies mentioned above suggest that the TPJ's function is not unique in the social context. In fact, many theorists consider the TPJ the key hub of the ventral attention network, which essentially supports attention reorienting for resolving conflicts between different visual perspectives, especially when there is a violation of the expected outcomes (Posner et al., 2006; Decety and Lamm, 2007; Mitchell, 2008; Perner and Aichhorn, 2008). Similarly, it has been suggested that the dorsal part of the TPJ region is involved in representing different perspectives and making behavioral predictions, whereas the more ventral part of TPJ and the medial prefrontal cortex region (MPFC) are responsible for predicting behavioral consequences (Aichhorn et al., 2006). Along the same line, Corbetta et al. (2008, p. 317) posited that, "Similar environmental and bodily representations and their comparison may be co-opted for ToM interactions and that attention signals in TPJ may be important to switch between internal, bodily, or self-perspective and external, environmental, or other's viewpoint, a key ingredient of ToM."

In sum, we argue that by supporting different levels of competition between multiple representations, the functions of dorsal–ventral attention networks play a major role in both spatial and social cognitive abilities. Whereas the filtering function manages competition among representations required for the

ongoing activity, the reorienting function facilitates competition and reconfiguration when the new sensory input violates the expectation from the current representations. Crucially, different levels of competition allow partial engagement (or disengagement) of certain representations, which facilitate the integration of potentially conflicting representations. As mentioned earlier, maintaining multiple IFOR representations is essential for prioritizing while being prepared for the unexpected. Combining EFOR and IFOR representations (perspective taking) takes advantage of both the efficient removal of task-irrelevant variance and fast mental simulation. When infants start to learn by copying others' actions (Meltzoff, 1995; Tomasello et al., 2005; Nielsen, 2006), it is important for them to hold both EFOR and IFOR representations so that imitation and emulation are possible.

SUMMARY

The central theme in our proposal is that the complex achievements in either spatial cognition or social cognition may rely on the fundamental processes of spatio-temporal integration and, moreover, that there is a set of distributed brain regions shared by both types of cognition. In our framework, both spatial and social abilities arise in the form of spatio-temporal association in which the mind constantly deals with the temporal instability in the environment by predictive learning. In the effort of extracting statistical regularities, the internal representations evolve by first partitioning the environmental variances – namely, developing FOR-based representations – then, encoding statistical invariance at different levels of abstractions. Since the statistical regularities include not only the spatial relations of static configurations but also the temporal relations between sequential events, predictive learning links various representations with different anchors (spatial integration) at different time points (temporal integration). Together, abstract knowledge of the environment (including those about other's beliefs and intentions) emerges from the expectation-driven competitions among multiple FOR-based representations.

In our view, different abilities are not domain-specific *per se*, rather, they are subject to the competing demands of computational efficiency and flexibility, yet are bounded by the statistical structures in the environment. By reinterpreting the results from the two-cannon experiment (Tamborello et al., 2012) and the false-belief task (Onishi and Baillargeon, 2005) and reviewing recent neurocognitive findings, we advocate an integrated approach that connects low-level perceptual processes, such as spatial representations, with high-level functions such as belief reasoning. The advantage of this approach is that, rather than singling out a certain brain system for a certain set of cognitive abilities (e.g., the TPJ for belief reasoning), we can pursue a better understanding of the mind–environment interaction over a developmental continuum. For example, the FOR-based account proposed here largely relies on the mechanisms of attentional network in spatial cognition, which have been extensively studied on from non-human animals to human infants and adults (for reviews, see, Corbetta and Shulman, 2002; Posner et al., 2006; Corbetta et al., 2008; Kavšek, 2013). Thus, this account may provide not only a transparent partitioning of the environmental statistics, but also potential explanations for the relationship between different abilities and

the development of specific attentional networks. For instance, it has been suggested that “rudimentary executive attention capacities may emerge during the first year of life but that more advanced conflict resolution capacities are not present until 2 years of age” (Posner et al., 2006, p. 1425). This line of reasoning could explain why young infants suddenly appear to comprehend the complex world and pass various spatial tasks (McCrink and Wynn, 2007; Surian et al., 2007; Kovács et al., 2010; Gweon and Schulz, 2011; Téglás et al., 2011).

Legend has it that in formulating his theory of gravitation, Newton was inspired by observing the acceleration of an apple falling from a tree. Subsequently, he inferred the existence of gravity and extended the effect from the top of the tree to the Moon (White, 1991). Perhaps more interestingly, Newton also first stated the principle of relativity (later modified by Einstein), which essentially claims that observations of the physical world depend on the particular “frame of reference” (Feynman et al., 1963, p. 162). Although we may never know the exact details of his revelation, the “apple incident” exemplifies how early perceptual analyses are triggered by temporal instability in the environment and the resulting extraction of statistical regularities with various reference points. In addition, it illuminates recent proposals that complex achievements such as mathematics and geometry, which are uniquely human in their full linguistic and symbolic realization, rest nevertheless on a set of core knowledge systems that are driven by the representations of object, space, time and number (Spelke and Kinzler, 2007; Spelke et al., 2010), and, knowledge structures emerge from non-cognitive processes by dynamic associations (McClelland et al., 2010). While controversies still exist between seemingly diverging perspectives, we take the primary theme of the debates to be the converging efforts of seeking for the cognitive or non-cognitive origins of human thinking and reasoning abilities. If we subscribe to the notion of “bounded rationality” (Simon, 1982), both spatial and social abilities are bounded by the learning agent’s computation capacity and the structure of the environment. In order to bridge the conceptual gaps between spatial and social cognition, the key is to understand the interactions between “genetic endowment and the environment” (Ruffman and Perner, 2005, p. 462).

ACKNOWLEDGMENTS

This work was supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of the Interior (DOI) Contract no. D10PC20021 and the Office of Naval Research (ONR) Grant no. N00014-08-1-0042. The US Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DOI, or the US Government. We would like to thank Dr. Paul J. Schroeder for helpful comments.

REFERENCES

Aichhorn, M., Perner, J., Kronbichler, M., Staffen, W., and Ladurner, G. (2006). Do visual perspective tasks need theory of mind? *Neuroimage* 30, 1059–1068. doi: 10.1016/j.neuroimage.2005.10.026

- Andersen, R. A., Snyder, L. H., Bradley, D. C., and Xing, J. (1997). Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annu. Rev. Neurosci.* 20, 303–330. doi: 10.1146/annurev.neuro.20.1.303
- Apperly, I. A., and Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychol. Rev.* 116, 953–970. doi: 10.1037/a0016923
- Apperly, I. A., Samson, D., Chiavarino, C., and Humphreys, G. W. (2004). Frontal and temporo-parietal lobe contributions to theory of mind: neuropsychological evidence from a false-belief task with reduced language and executive demands. *J. Cogn. Neurosci.* 16, 1773–1784. doi: 10.1162/0899829042947928
- Baillargeon, R. (1986). Representing the existence and the location of hidden objects: object permanence in 6- and 8-month-old infants. *Cognition* 23, 21–41. doi: 10.1016/0010-0277(86)90052-1
- Baillargeon, R., Scott, R. M., and He, Z. (2010). False-belief understanding in infants. *Trends Cogn. Sci.* 14, 110–118. doi: 10.1016/j.tics.2009.12.006
- Blakemore, S.-J., and Frith, C. (2003). Self-awareness and action. *Curr. Opin. Neurobiol.* 13, 219–224. doi: 10.1016/S0959-4388(03)00043-6
- Burgess, N. (2008). Spatial cognition and the brain. *Ann. N. Y. Acad. Sci.* 1124, 77–97. doi: 10.1196/annals.1440.002
- Carlson, L. A., and Van Deman, S. R. (2008). Inhibition within a reference frame during the interpretation of spatial language. *Cognition* 106, 384–407. doi: 10.1016/j.cognition.2007.03.009
- Carlson-Radvansky, L. A., and Irwin, D. E. (1993). Frames of reference in vision and language: where is above? *Cognition* 46, 223–244. doi: 10.1016/0010-0277(93)90011-J
- Chen, X., and McNamara, T. (2011). Object-centered reference systems and human spatial memory. *Psychon. Bull. Rev.* 18, 985–991. doi: 10.3758/s13423-011-0134-5
- Colby, C. L., and Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annu. Rev. Neurosci.* 22, 319–349. doi: 10.1146/annurev.neuro.22.1.319
- Corbetta, M., Kincade, J. M., Ollinger, J. M., McAvoy, M. P., and Shulman, G. L. (2000). Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nat. Neurosci.* 3, 292–297. doi: 10.1038/73009
- Corbetta, M., Patel, G., and Shulman, G. L. (2008). The reorienting system of the human brain: from environment to theory of mind. *Neuron* 58, 306–324. doi: 10.1016/j.neuron.2008.04.017
- Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215. doi: 10.1038/nrn755
- Csibra, G., and Southgate, V. (2006). Evidence for infants’ understanding of false beliefs should not be dismissed. *Trends Cogn. Sci.* 10, 4–5. doi: 10.1016/j.tics.2005.11.011
- David, N., Bewernick, B. H., Cohen, M. X., Newen, A., Lux, S., Fink, G. R., et al. (2006). Neural representations of self versus other: visual-spatial perspective taking and agency in a virtual ball-tossing game. *J. Cogn. Neurosci.* 18, 898–910. doi: 10.1162/jocn.2006.18.6.898
- Decety, J., and Lamm, C. (2007). The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. *Neuroscientist* 13, 580–593. doi: 10.1177/1073858407304654
- Feynman, R. P., Leighton, R. B., and Sands, M. (1963). *The Feynman Lectures on Physics*, Vol. 1. Reading, MA: Addison-Wesley.
- Frith, C. D., and Frith, U. (2007). Social cognition in humans. *Curr. Biol.* 17, R724–R732. doi: 10.1016/j.cub.2007.05.068
- Frith, C. D., and Frith, U. (2012). Mechanisms of social cognition. *Annu. Rev. Psychol.* 63, 287–313. doi: 10.1146/annurev-psych-120710-100449
- Gerstner, W., Sprekeler, H., and Deco, G. (2012). Theory and simulation in neuroscience. *Science* 338, 60–65. doi: 10.1126/science.1227356
- Gweon, H., and Schulz, L. (2011). 16-month-olds rationally infer causes of failed actions. *Science* 332, 1524. doi: 10.1126/science.1204493
- Hawkins, J., and Blakeslee, S. (2004). *On Intelligence*. New York: Henry Holt.
- Jackson, P. L., and Decety, J. (2004). Motor cognition: a new paradigm to study self-other interactions. *Curr. Opin. Neurobiol.* 14, 259–263. doi: 10.1016/j.conb.2004.01.020
- Kavšek, M. (2013). The comparator model of infant visual habituation and dishabituation: recent insights. *Dev. Psychobiol.* 55, 793–808. doi: 10.1002/dev.21081
- Kessler, K., and Rutherford, H. (2010). The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions. *Front. Psychol.* 1:213. doi: 10.3389/fpsy.2010.00213

- Kessler, K., and Thomson, L. A. (2010). The embodied nature of spatial perspective taking: embodied transformation versus sensorimotor interference. *Cognition* 114, 72–88. doi: 10.1016/j.cognition.2009.08.015
- Kovács, Á. M., Téglás, E., and Endress, A. D. (2010). The social sense: susceptibility to others' beliefs in human infants and adults. *Science* 330, 1830–1834. doi: 10.1126/science.1190792
- Krauzlis, R. J., and Nummela, S. U. (2011). Attention points to the future. *Nat. Neurosci.* 14, 130–131. doi: 10.1038/nn0211-130
- Leslie, A. M. (2005). Developmental parallels in understanding minds and bodies. *Trends Cogn. Sci.* 9, 459–462. doi: 10.1016/j.tics.2005.08.002
- Levelt, W. J. M. (1996). "Perspective taking and ellipsis in spatial descriptions," in *Language and Space*, eds P. Bloom, M. A. Peterson, L. Nadel, and M. Garrett (Cambridge, MA: MIT Press), 77–108.
- Levinson, S. C. (2004). *Space in Language and Cognition: Explorations in Cognitive Diversity*. New York: Cambridge University Press.
- Marr, D. (1982). *Vision*. San Francisco, CA: W. H. Freeman.
- Mazzarella, E., Ramsey, R., Conson, M., and Hamilton, A. (2013). Brain systems for visual perspective taking and action perception. *Soc. Neurosci.* 8, 248–267. doi: 10.1080/17470919.2012.761160
- McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., et al. (2010). Letting structure emerge: connectionist and dynamical systems approaches to cognition. *Trends Cogn. Sci.* 14, 348–356. doi: 10.1016/j.tics.2010.06.002
- McCrink, K., and Wynn, K. (2007). Ratio abstraction by 6-month-old infants. *Psychol. Sci.* 18, 740–745. doi: 10.1111/j.1467-9280.2007.01969.x
- Meltzoff, A. N. (1995). Understanding the intentions of others: re-enactment of intended acts by 18-month-old children. *Dev. Psychol.* 31, 838–850. doi: 10.1037/0012-1649.31.5.838
- Miller, G. A., and Johnson-Laird, P. N. (1976). *Language and Perception*. Cambridge, MA: Harvard University Press.
- Mitchell, J. P. (2006). Mentalizing and Marr: an information processing approach to the study of social cognition. *Brain Res.* 1079, 66–75. doi: 10.1016/j.brainres.2005.12.113
- Mitchell, J. P. (2008). Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cereb. Cortex* 18, 262–271. doi: 10.1093/cercor/bhm051
- Morton, J. B., and Munakata, Y. (2002). Active versus latent representations: a neural network model of perseveration, dissociation, and decalage. *Dev. Psychobiol.* 40, 255–265. doi: 10.1002/dev.10033
- Mou, W., Fan, Y., McNamara, T. P., and Owen, C. B. (2008). Intrinsic frames of reference and egocentric viewpoints in scene recognition. *Cognition* 106, 750–769. doi: 10.1016/j.cognition.2007.04.009
- Nielsen, M. (2006). Copying actions and copying outcomes: social learning through the second year. *Dev. Psychol.* 42, 555–565. doi: 10.1037/0012-1649.42.3.555
- Onishi, K. H., and Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science* 308, 255–258. doi: 10.1126/science.1107621
- O'Reilly, R. C., Munakata, Y., Frank, M. J., Hazy, T. E., and Contributors. (2012). *Computational Cognitive Neuroscience*, Wiki Book, 1st Edn, Available at URL: <http://ccnbook.colorado.edu> (accessed January 31, 2014).
- Perner, J., and Aichhorn, M. (2008). Theory of mind, language and the temporoparietal junction mystery. *Trends Cogn. Sci.* 12, 123–126. doi: 10.1016/j.tics.2008.02.001
- Perner, J., Mauer, M. C., and Hildenbrand, M. (2011). Identity: key to children's understanding of belief. *Science* 333, 474–477. doi: 10.1126/science.1201216
- Perner, J., and Ruffman, T. (2005). Infants' insight into the mind: how deep? *Science* 308, 214–216. doi: 10.1126/science.1111656
- Pertsov, Y., Avidan, G., and Zohary, E. (2011). Multiple reference frames for saccadic planning in the human parietal cortex. *J. Neurosci.* 31, 1059–1068. doi: 10.1523/JNEUROSCI.3721-10.2011
- Posner, M. I., Sheese, B. E., Odludaz, Y., and Tang, Y.-Y. (2006). Analyzing and shaping human attentional networks. *Neural Netw.* 19, 1422–1429. doi: 10.1016/j.neunet.2006.08.004
- Röls, M., Jonikaitis, D., Deubel, H., and Cavanagh, P. (2011). Predictive remapping of attention across eye movements. *Nat. Neurosci.* 14, 252–256. doi: 10.1038/nn.2711
- Ruffman, T., and Perner, J. (2005). Do infants really understand false belief? Response to Leslie. *Trends Cogn. Sci.* 9, 462–463. doi: 10.1016/j.tics.2005.08.001
- Saxe, R. R. (2006). Why and how to study Theory of Mind with fMRI. *Brain Res.* 1079, 57–65. doi: 10.1016/j.brainres.2006.01.001
- Saxe, R. R., and Kanwisher, N. (2003). People thinking about thinking people: the role of the temporo-parietal junction in "theory of mind." *Neuroimage* 19, 1835–1842. doi: 10.1016/S1053-8119(03)00230-1
- Saxe, R. R., and Powell, L. J. (2006). It's the thought that counts: specific brain regions for one component of theory of mind. *Psychol. Sci.* 17, 692–699. doi: 10.1111/j.1467-9280.2006.01768.x
- Saxe, R. R., and Wexler, A. (2005). Making sense of another mind: the role of the right temporo-parietal junction. *Neuropsychologia* 43, 1391–1399. doi: 10.1016/j.neuropsychologia.2005.02.013
- Saxe, R. R., Whitfield-Gabrieli, S., Scholz, J., and Pelphrey, K. A. (2009). Brain regions for perceiving and reasoning about other people in school-aged children. *Child Dev.* 80, 1197–1209. doi: 10.1111/j.1467-8624.2009.01325.x
- Simon, H. A. (1982). *Models of Bounded Rationality*. Cambridge, MA: MIT Press.
- Spelke, E. S., and Kinzler, K. D. (2007). Core knowledge. *Dev. Sci.* 10, 89–96. doi: 10.1111/j.1467-7687.2007.00569.x
- Spelke, E. S., Lee, S. A., and Izard, V. (2010). Beyond core knowledge: natural geometry. *Cogn. Sci.* 34, 863–884. doi: 10.1111/j.1551-6709.2010.01110.x
- Sun, Y., and Wang, H. (2010). Perception of space by multiple intrinsic frames of reference. *PLoS ONE* 5:e10442. doi: 10.1371/journal.pone.0010442
- Sun, Y., and Wang, H. (2013). The parietal cortex in sensemaking: the dissociation of multiple types of spatial information. *Comput. Intell. Neurosci.* 2013:152073. doi: 10.1155/2013/152073
- Surian, L., Caldi, S., and Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychol. Sci.* 18, 580–586. doi: 10.1111/j.1467-9280.2007.01943.x
- Tamborello, F. P., Sun, Y., and Wang, H. (2012). Spatial reasoning with multiple intrinsic frames of reference. *Exp. Psychol.* 59, 3–10. doi: 10.1027/1618-3169/a000119
- Téglás, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., and Bonatti, L. L. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. *Science* 332, 1054–1059. doi: 10.1126/science.1196404
- Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* 28, 675–691. doi: 10.1017/S0140525X05000129
- Van Der Werf, J., Buchholz, V. N., Jensen, O., and Medendorp, W. P. (2013). Reorganization of oscillatory activity in human parietal cortex during spatial updating. *Cereb. Cortex* 23, 508–519. doi: 10.1093/cercor/bhr387
- Vogeley, K., May, M., Ritzl, A., Falkai, P., Zilles, K., and Fink, G. R. (2004). Neural correlates of first-person perspective as one constituent of human self-consciousness. *J. Cogn. Neurosci.* 16, 817–827. doi: 10.1162/0898929049707099
- Wallentin, M. (2012). The role of the brain's frontal eye fields in constructing frame of reference. *Cogn. Process.* 13, 359–363. doi: 10.1007/s10339-012-0461-0
- Wang, H., Johnson, T. R., Sun, Y., and Zhang, J. (2005a). Object location memory: the interplay of multiple representations. *Mem. Cogn.* 33, 1147–1159. doi: 10.3758/BF03193219
- Wang, H., Sun, Y., Johnson, T. R., and Yuan, Y. (2005b). Prioritized spatial updating in the intrinsic frame of reference. *Spat. Cogn. Comput.* 5, 89–113. doi: 10.1207/s15427633scc0501_4
- Wang, R. F., and Spelke, E. S. (2002). Human spatial representation: insights from animals. *Trends Cogn. Sci.* 6, 376–382. doi: 10.1016/S1364-6613(02)01961-7
- White, M. (1991). *Isaac Newton: The Story of a Great Mathematician Who Changed Our Perception of the Universe*. Watford, UK: Exley.
- Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* 106, 1125–1165. doi: 10.1152/jn.00338.2011
- Young, L., Dodel-Feder, D., and Saxe, R. (2010). What gets the attention of the temporo-parietal junction? An fMRI investigation of attention and theory of mind. *Neuropsychologia* 48, 2658–2664. doi: 10.1016/j.neuropsychologia.2010.05.012
- Zwiesel, J., White, S. J., Coniston, D., Senju, A., and Frith, U. (2011). Exploring the building blocks of social cognition: spontaneous agency perception and visual perspective taking in autism. *Soc. Cogn. Affect. Neurosci.* 6, 564–571. doi: 10.1093/scan/nsq088

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 28 June 2013; accepted: 24 January 2014; published online: 14 February 2014.

Citation: Sun Y and Wang H (2014) Insight into others' minds: spatio-temporal representations by intrinsic frame of reference. *Front. Hum. Neurosci.* 8:60. doi: 10.3389/fnhum.2014.00058

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2014 Sun and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Visual perspective taking and laterality decisions: problems and possible solutions

Mark May^{1*} and Mike Wendt²

¹Spatial Cognition Research Unit, Helmut-Schmidt-University, Hamburg, Germany

²Experimental Psychology Unit, Helmut-Schmidt-University, Hamburg, Germany

Edited by:

Sarah H. Creem-Regehr, University of Utah, USA

Reviewed by:

Mark Gardner, University of Westminster, UK

Alfred Brian Yu, Washington University in St. Louis, USA

***Correspondence:**

Mark May, Spatial Cognition Research Unit, Helmut-Schmidt-University, Holstenhofweg 85, D-22043 Hamburg, Germany
e-mail: mm@hsu-hh.de

Perspective taking plays an important role in different areas of psychological and neuroscientific research. Visual perspective taking is an especially prominent approach generally using one of two experimental tasks: in the own-body-transformation task observers are asked to judge the laterality of a salient feature of a human figure (e.g., is the glove on the left or right hand?) from the figure's perspective. In the avatar-in-scene task they decide about the laterality of objects in a scene (e.g., is the flower on the left or right?) from the avatar's point of view. Increases in latencies and/or errors are interpreted as originating from additional cognitive processes predominately described as observer-based perspective transformations. A closer look reveals that such an account is disputable on grounds related to the use of laterality judgments. Other transformation accounts, i.e., object or array transformations, as well as non-transformational accounts, i.e., extra processing due to spatial conflicts, have not been adequately considered, tested, or ruled out by existing research. Our review examines visual perspective tasks in detail, identifies problems and makes recommendations for future research.

Keywords: spatial cognition, embodiment, visual perspective taking, mental transformation, own-body-transformation task, laterality tasks, spatial S-R compatibility, agency

INTRODUCTION

Research on human perspective taking is gaining momentum as can be seen by the increasing number of experimental studies in different research areas, such as spatial reasoning, mental imagery, life-span cognitive development, theory of mind, empathy, aviation research, and teleoperations. The different fields have in common that they want to come up with accounts of the cognitive mechanisms underlying the ability to mentally switch into and spatially act from perspectives that are not our own, and sometimes those of others.

Two fundamentally different lines of research on spatial perspective taking can be distinguished: Research on mental perspective taking uses memory-based testing methods. In one line of work, participants first learn a layout of objects and are then asked to point to the previously learned objects without being able to look at the scene while bodily or imaginably switching into various perspectives. Measures of geometric differences between learned, body-defined and to-be-imagined perspectives have been found to be good predictors of pointing latencies and errors, and results are used to test competing processing accounts (e.g., Rieser, 1989; Easton and Sholl, 1995; Shelton and McNamara, 1997; Creem-Regehr, 2003; May, 2004; Avraamides and Kelly, 2008). Other studies examine mental perspective taking by using language, graphics, or maps as learning input and with other testing procedures (e.g., De Vega et al., 1996; Bryant and Tversky, 1999; Avraamides, 2003; Sohn and Carlson, 2003).

Research on visual perspective taking, on the other hand, uses perception-based testing methods. Participants usually look at a visual display including a human figure, and have to decide about the side of a critical feature of the figure while adopting the figure's

perspective (OBT or own-body transformation task; e.g., Parsons, 1987), or about relative object positions from the figure's point of view (AIS or avatar-in-scene task; e.g., Amorim, 2003). Although both tasks are usually treated separately in the literature, the majority of OBT- and AIS-studies have in common that they use laterality decisions, i.e., observers have to make left or right judgments about absolute or relative object locations from the figure's point of view. Recently, the number of behavioral and neurophysiological studies on visual perspective taking has been growing (Blanke et al., 2005; Creem-Regehr et al., 2007; Kessler and Thomson, 2010; Yu and Zacks, 2010; Dalecki et al., 2012 and others reviewed here). Note, that studies on viewpoint-dependent object (Tarr and Bülthoff, 1998) or scene recognition (Diwadkar and McNamara, 1997) are not considered, as their focus is on memory-based identification processes, and not on perspective taking.

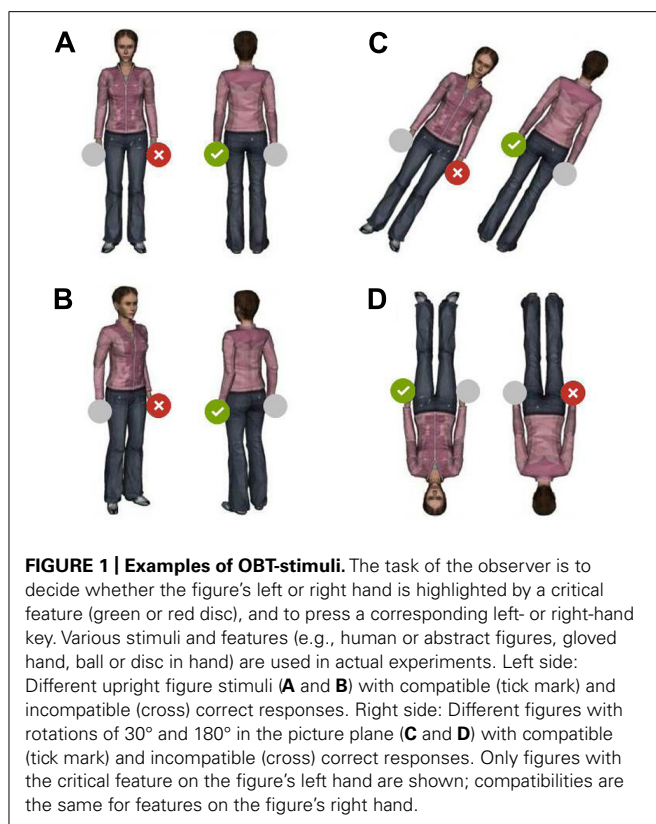
The overall picture of findings on visual perspective taking is complex. In general, one finds increases in response times and errors the larger the spatial difference between the observer's and the figure's spatial perspective. This is taken to reflect additional cognitive processes described as observer-based perspective transformations (PT). In contrast to this widely held view, our review will argue that alternative accounts, e.g., object transformations (OT) of the figure in the OBT-task, or array transformations (AT) in the AIS-task, have been brought forth, but so far have not been systematically evaluated and pursued. Furthermore, the review will show that combining visual perspective taking tasks with laterality judgments leads to spatially compatible and incompatible responses, with consequences that have not been adequately addressed up to now.

TASKS AND BASIC FINDINGS

OWN-BODY TRANSFORMATION TASK

Experiments using the OBT task show an isolated human figure with a salient body feature (e.g., a glove, a hand-held ball or disk). The observer's task is to decide whether the salient feature is on the left or right as seen from the figure's point of view and to respond by pressing a left- or right-hand key (or using another response indicating left and right). **Figure 1** shows examples of OBT-stimuli.

Consistent with the notion that observers mentally transform their own perspective until it matches the figure's perspective before deciding about laterality, responses usually are faster for back-facing figures, i.e., when figures look in the same direction as the observer, compared to front-facing figures, i.e., when observer and figure look in opposite directions (Parsons, 1987; Zacks et al., 1999; Blanke et al., 2005; Jola and Mast, 2005; Arzy et al., 2006, 2007; Mohr et al., 2006, 2010; Gardner and Potts, 2010, 2011; Thakkar and Park, 2010; Braithwaite et al., 2011; Steggemann et al., 2011; Gardner et al., 2012; Gronholm et al., 2012; May and Wendt, 2012). No such performance differences are found when observers have to decide about the laterality of the critical feature from their own perspective (referred to as which-side-task), rather than from the avatar's perspective (Blanke et al., 2005; Gardner and Potts, 2010, 2011; Braithwaite et al., 2011; Gardner et al., 2012). In support of a PT account of these findings, more than half of the participants report to switch into the avatar's perspective when solving the task (Parsons, 1987; Zacks and Tversky, 2005; Gronholm et al., 2012).



AVATAR-IN-SCENE TASK

Experiments using the AIS-task show an avatar (or a different symbol indicating the relevant perspective) looking at a spatial scene from varying angles of rotation in the horizontal plane. The observer's task is to decide whether a critical object in the scene (e.g., flower) is on the left or right side from the avatar's point of view. **Figure 2** provides examples of AIS-stimuli.

Response times for laterality judgments grow monotonically with the disparity of the avatar's and the participant's perspectives (e.g., Keehner et al., 2006; Michelon and Zacks, 2006; Kessler and Rutherford, 2010; Kessler and Thomson, 2010; Kockler et al., 2010). Similar to the back-facing advantage in the OBT task, these findings have been interpreted in terms of time to transform one's own perspective into the avatar's perspective.

THEORETICAL ACCOUNTS OF VISUAL PERSPECTIVE TAKING

The above interpretations of OBT- and AIS-studies have been used to identify brain regions mediating visual perspective taking (e.g., Zacks et al., 1999, 2002; Blanke et al., 2005), and also to look into processing strategies used with human and non-human stimuli (e.g., Yu and Zacks, 2010). It turns out, however, that observed performances in laterality judgment tasks lead to difficulties when researchers try to interpret them as indicators of PTs. In the following, we look at existing evidence from the perspective of a PT-account, and at arguments used to defend it against competing OT/AT-accounts, or spatial compatibility explanations.

CONFOUNDING SPATIAL TRANSFORMATIONS AND RESPONSE CONFLICTS

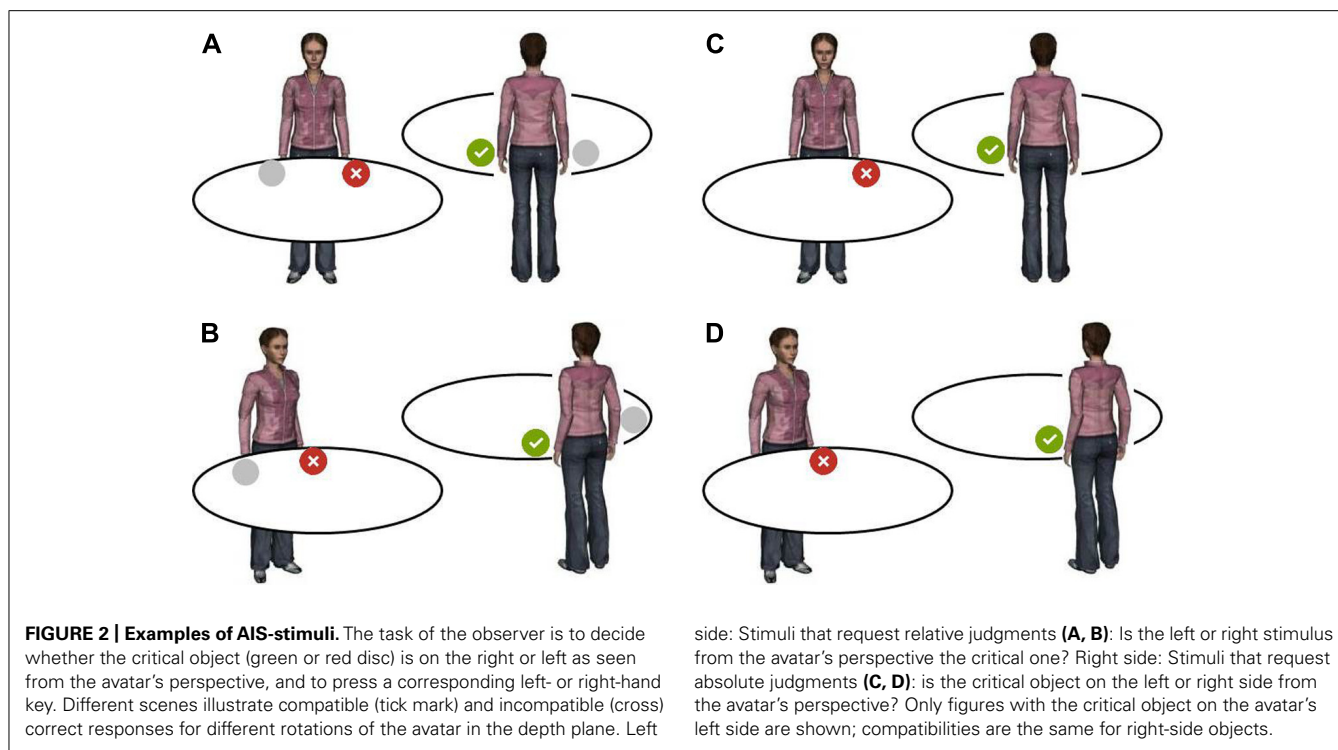
Under a variety of conditions, responses are faster and less error-prone when a target is presented at a location that spatially corresponds with the location of the requested response as compared to situations where the target location spatially corresponds with an incorrect response (Proctor and Vu, 2006). Dual-route models attribute such S-R compatibility effects to automatic activation of the spatially corresponding response along a processing route largely independent of intention-based S-R translation processes (Hommel, 1993; De Jong et al., 1994).

Spatial compatibility in OBT-tasks

Figure 1 shows that the location of the target feature spatially corresponds with the correct response for the back-facing upright figure whereas it corresponds with the incorrect response when the figure is shown front-facing. Spatial compatibility should facilitate responses to back-facing compared to front-facing figures. Intermediate orientations of the OBT-figure in the depth plane can be presumed to lead to graded compatibility effects.

Spatial compatibility in AIS-tasks

The AIS-task is in most aspects similar to the OBT-task, and similar problems arise (see **Figure 2**). On the one hand, and different from the centered presentation of OBT-stimuli, the positioning of the avatar-object-ensemble on the screen, can shift to the left or right from the screen's center, potentially producing independent spatial (i.e., Simon-type) compatibility effects. On the other hand, and similar to the OBT-task, the relative position of the target object (left/right) within the ensemble as seen from the observer's perspective, corresponds to the laterality of the correct response up



to rotation angles of 90° . In contrast, rotations larger than 90° lead to a reversal of this situation, yielding spatial S-R-correspondence in the former case, and spatial non-correspondence in the latter case. Again, spatial compatibility and incompatibility should be maximal for avatars in the 0° (back-facing) and 180° (front-facing) positions, and might produce gradual effects for intermediate rotations. Both problems are rarely addressed in the literature.

Empirical evidence for spatial compatibility effects

Several findings are consistent with the assumption that OBT-performances are influenced by spatial compatibility. For instance, Gardner and Potts (2011, Exp. 1; Parsons, 1987, Exp. 2a) used vocal “left” and “right” responses, which are known to produce smaller compatibility effects than manual responses, and found a reduced back-facing advantage. Moreover, some manipulations which reversed the assigned correspondence values yielded a back-facing disadvantage. For instance, Arzy et al. (2006) asked participants to treat the depicted figure as a mirror reflection of their own body, and obtained slower responses for back-facing as compared to front-facing figures, while at the same time observing the well-known back-facing advantage with standard OBT task instructions.

Other studies presented the figure in different orientations in the picture plane, including upside-down versions, for which front-facing figures come with spatially corresponding, and back-facing figures with spatially conflicting responses (Figure 1). Upside-down presentation of figures either reduced (Steggemann et al., 2011), or even changed the back-facing advantage into a disadvantage (Parsons, 1987; Zacks et al., 1999; Jola and Mast, 2005; May and Wendt, 2012). Furthermore, Gardner and Potts

(2011, Exp. 2) obtained a back-facing disadvantage by instructing participants to cross their hands and decide about laterality by key-presses with their corresponding hand, thereby reversing laterality and response locations (see, however, May and Wendt, 2012, for a back-facing advantage with uncrossed arms when left and right keys were labeled “right” and “left,” respectively).

Although this review focuses on perceptual laterality judgment tasks for which the confound of facing direction and compatibility is most obvious it should be noted that Simon-like spatial interference effects have also been found with respect to a remembered previous location of a current stimulus (e.g., Zhang and Johnson, 2004). More generally, the problem of spatial compatibility is also present in memory-based perspective taking tasks and has been subject of thorough discussion (e.g., May, 2004). Furthermore, perception- and memory-based tasks not asking for laterality decisions (e.g., color judgments) may induce spatial compatibility effects if the location of the response varies with respect to the same spatial dimension as the target stimulus feature (e.g., indicating red and green with left- and right-side key presses, respectively). Other tasks with non-spatial decision criteria (e.g., same-different, visibility, or numerosity judgments) could also induce spatial conflicts. In such cases it must be ensured that compatibility levels are balanced across facing directions.

Controlling for spatial compatibility

Attempts to control for compatibility have used figures with an outstretched arm across the body midline, where observers make laterality decisions regarding the outstretched arm. Although a back-facing advantage was also found for such figures (Parsons, 1987, Exp. 2b), this evidence is not conclusive, as it is possible that

participants respond on the basis of a non-switching body feature such as the shoulder (Jola and Mast, 2005). Avoiding this problem, May and Wendt (2012) controlled for spatial compatibility by using horizontal figures (i.e., 90°-rotated) with hands equidistant to the figure's upper and lower end; in spite of this arguably neutral conditions, a clear back-facing advantage was found (see also Parsons, 1987; Jola and Mast, 2005; Stegmann et al., 2011).

LITTLE INDUBITABLE EVIDENCE FOR PERSPECTIVE TRANSFORMATIONS

Since the spatial end-state of PTs can principally also be reached by spatially equivalent OTs of the figure in the OBT-task, or ATs of the avatar-object-ensemble in the AIS-task, both constitute potential alternative explanations for the typical facing direction effects found in both tasks. OTs have been extensively studied in mental rotation research, by presenting a stimulus that differs in orientation from a second version of the same stimulus, or that is moved away from its canonical orientation, while asking participants to make a same/mirror-reversed judgment. Such studies show monotonic increasing reaction time (RT)-slopes for increasing rotation angles in both the picture and the depth plane (Shepard and Cooper, 1982).

Slope differences as evidence for PT

Slope differences play an important role in studies using OBT-tasks which try to distinguish between PT- and OT-accounts. These studies include rotations in the picture plane, and find slope differences for back- and front-facing figures. While RTs increase with rotation angle for back-facing figures, slopes are strongly reduced, absent, or even reversed for front-facing figures (Parsons, 1987; Jola and Mast, 2005; Zacks and Tversky, 2005; Yu and Zacks, 2010; Stegmann et al., 2011; May and Wendt, 2012, Exp. 2; Zacks et al., 2000, 2002). Thus, performances for back-facing (but not for front-facing) figures are consistent with findings from research on mental rotation with same vs. mirror-reversed objects. The missing slopes in laterality decisions for front-facing figures have been repeatedly taken as evidence for PT-accounts (e.g., Yu and Zacks, 2010). For this argument to work, minimal costs for transformations in the picture plane have to be postulated. This constraint can be met by assuming that PTs are realized as shortest path spatial transformations; i.e., all rotation trajectories of observer-based switches into front-facing figures have the same rotation angle (i.e., 180°), irrespective of the figure's orientation in the picture plane (see Parsons, 1987, p. 190).

Alternative explanations for slope differences

The observed slope differences for rotations in the picture plane can also be accounted for by compatibility assumptions (May and Wendt, 2012). Specifically, figures presented upside-down reverse S-R-compatibility values; i.e., compatible responses become incompatible, and vice versa. Applied to upside-down figures this means, that back-facing figures produce spatial conflicts, while front-facing figures do not (see **Figure 1**). Intermediate rotations of the upright figure in the picture plane, should lead to graded effects of compatibility.

PT- vs. OT-instructions

Independent support for PT-assumptions comes from experiments that use particular transformation instructions. Specifically, Zacks and Tversky (2005) observed positive RT-slopes for front-facing figures in a laterality judgment task when participants were asked to use object rotation strategies on the figures. However, near-zero slopes were found when participants received explicit PT-instructions or unspecific task instructions. Furthermore, averaged across all orientations of the figures in the picture plane substantial RT-increases for object-based instructions as compared to both observer-based transformation or unspecific instructions were found.

Although the findings of Zacks and Tversky (2005) can be interpreted to reveal that PTs are naturally used for human figures (if not instructed otherwise), in our opinion this does not provide indisputable evidence for PTs, as the following considerations show: explicit object rotation instructions (e.g., "imagine the figure rotating until it is upright," p. 281) may induce OTs (i.e., picture-plane rotations of the front-facing figure) that are not the same OTs that can be assumed to be at work with unconstrained task instructions (i.e., shortest path object rotations of front-facing to back-facing figures). In other words, finding positive RT-slopes with explicit instructions to rotate the object in the picture plane speaks against the use of such OTs with non-specific instructions (i.e., flat slopes), but not against other types of OTs as a strategy spontaneously adapted by observers. Further doubt concerning a PT-interpretation of near-zero slopes comes from findings that reveal flatter or missing RT-slopes with figure stimuli in a standard mental rotation task (Amorim et al., 2006), as well as in a hand laterality identification task when a palm view of the human hand is presented (Ionta and Blanke, 2009). Without going into the particular nature of the underlying mechanisms (e.g., embodiment), such findings suggest that the absence of RT-slopes should not be regarded as positive evidence to dismiss OT-accounts.

Perspective transformations vs. OT/AT-instructions can also be manipulated by using stimuli rotated in the depth plane (i.e., the plane for which PTs in OBT- and AIS-tasks have been postulated). In memory-based AIS-tasks this has consistently yielded different RT-profiles (Wraga et al., 2000). Using a visual AIS-task with laterality decisions, Keehner et al. (2006) obtained comparable results, finding, in addition, differential brain activation for PT- vs. OT-instructions, supporting the assumption of processing differences between both. Although the experimental setup in Keehner et al. (2006) confounds rotations in the depth plane with incompatibility, this confound was, on average, equal for the PT- and OT-instructions. This seems a promising approach to gain further insight into the processes invoked by different transformation instructions (for other examples see Zacks et al., 2003; Tadi et al., 2009; Wraga et al., 2010).

EVIDENCE FOR SPONTANEOUS PERSPECTIVE TRANSFORMATIONS

Whereas the evidence reported so far does not seem compelling in ruling out alternative transformation accounts of OBT- and AIS-performances, more convincing evidence for PTs comes from research in which observers make laterality decisions regarding their current perspective on a visual scene, showing that task performance suffers interference from the depicted avatar's

perspective. More specifically, Zwickel (2009; also Zwickel and Müller, 2013) presented animations of simple geometrical shapes and asked participants to make left/right decisions – from their own perspective – about briefly presented dots. Performance in this task was impaired when the laterality of the dot mismatched its laterality regarding the perspective ascribed to the animated figure. Obviously, such a finding could not result from a confounding with spatial compatibility, because responding always corresponded to the laterality of the critical feature from the observer's perspective. It also does not seem reasonable to assume that OTs operated on the avatar-stimulus itself. The fact that no similar interference effects for laterality decisions about OBT-stimuli from their own perspective (i.e., which-side-task) were found, suggests that ascriptions of agency and/or embodied processing of the stimuli may be a prerequisite for spontaneous perspective taking (for discussions e.g., Kessler and Thomson, 2010; Kockler et al., 2010; Surtees and Apperly, 2012). This line of research seems interesting to pursue, as it could build a bridge to research on perspective conflicts and interference effects in cognitive (May, 2004, 2007; Wang, 2004; Kelly et al., 2007; Keehner and Fischer, 2012), as well as emotional and social perspective taking (Vogeley et al., 2004; Decety and Jackson, 2006; Duran et al., 2011; Mazzarella et al., 2013).

CONCLUSION

Our review reveals that there is less support for the assumption that visual perspective taking is based on observer-based PTs than one would believe when looking at the literature. The foregoing analysis of OBT- and AIS-studies using laterality judgments (and these are the majority of studies) reveals a quite complicated research situation with different problems standing in the way of a PT-account of visual perspective taking. On the one hand, OBT- or

AIS-studies using laterality judgments have problems to separate spatial incompatibility costs from transformation costs, making compatibility a potential alternative explanation for some of the findings. On the other, there is at least some evidence that spatial transformations play a role in visual perspective taking, but little evidence that PT-accounts of this role are more convincing than OT-accounts in case of OBT-performances, or AT-accounts in case of AIS-performances.

RECOMMENDATIONS

In order for future research to further close in on the mechanisms underlying visual perspective taking the following methodological recommendations might be helpful:

- (1) When using OBT- or AIS-tasks in combination with laterality decisions, take effective measures to control for spatial compatibility.
- (2) The measure taken should allow disentangling the independent contributions of spatial transformation and spatial conflict costs (for steps in this direction see Gardner and Potts, 2011, or May and Wendt, 2012, similar measures are conceivable for AIS-tasks).
- (3) To exclude compatibility influences altogether, non-lateralized spatial judgments should be preferred; for example, same/different decisions in OBT-tasks, force-choice decisions, object naming or object counting in AIS-tasks. When using such tasks, look out and control for possible hidden laterality influences (e.g., uneven spatial distribution of features/objects in both tasks).

ACKNOWLEDGMENT

This research was partially supported by the German Research Foundation (DFG MA-1515-3-1, and DFG WE 4105/1-2).

REFERENCES

- Amorim, M. A. (2003). What is my avatar seeing?: the coordination of “out-of-body” and “embodied” perspectives for scene recognition across views. *Vis. Cogn.* 10, 157–199. doi: 10.1080/713756678
- Amorim, M. A., Isableu, B., and Jaraya, M. (2006). Embodied spatial transformations: “body analogy” for the mental rotation of object. *J. Exp. Psychol. Gen.* 135, 327–347. doi: 10.1037/0096-3445.135.3.327
- Arzy, S., Mohr, C., Michel, C., and Blanke, O. (2007). Duration and not strength of activation in temporoparietal cortex positively correlates with schizotypy. *Neuroimage* 35, 326–333. doi: 10.1016/j.neuroimage.2006.11.027
- Arzy, S., Thut, G., Mohr, C., Michel, C., and Blanke, O. (2006). Neural basis of embodiment: distinct contributions of temporoparietal junction and extrastriate body area. *J. Neurosci.* 26, 8074–8081. doi: 10.1523/jneurosci.0745-06.2006
- Avraamides, M. N. (2003). Spatial updating of environments described in texts. *Cogn. Psychol.* 47, 402–431. doi: 10.1016/s0010-0285(03)00098-7
- Avraamides, M. N., and Kelly, J. W. (2008). Multiple systems of spatial memory and action. *Cogn. Proc.* 9, 93–106. doi: 10.1007/s10339-007-0188-5
- Blanke, O., Mohr, C., Michel, C. M., Pascual-Leone, A., Brugger, P., Seeck, M., et al. (2005). Linking out-of-body experience and self processing to mental own-body imagery at the temporoparietal junction. *J. Neurosci.* 25, 550–557. doi: 10.1523/jneurosci.2612-04.2005
- Braithwaite, J. J., Samson, D., Apperly, I., Brogna, E., and Hulleman, J. (2011). Cognitive correlates of the spontaneous out-of-body experience (OBE) in the psychologically normal population: evidence for an increased role of temporal-lobe instability, body-distortion processing, and impairments in own-body transformations. *Cortex* 47, 839–854. doi: 10.1016/j.cortex.2010.05.002
- Bryant, D. J., and Tversky, B. (1999). Mental representations of perspective and spatial relations from diagrams and models. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 137–156. doi: 10.1037/0278-7393.25.1.137
- Creem-Regehr, S. H. (2003). Updating space during imagined self- and array translations. *Mem. Cogn.* 31, 941–952. doi: 10.3758/bf03196447
- Creem-Regehr, S. H., Neil, J. A., and Jeh, H. J. (2007). Neural correlates of two imagined egocentric transformations. *Neuroimage* 35, 916–927. doi: 10.1016/j.neuroimage.2006.11.057
- Dalecki, M., Hoffmann, U., and Bock, O. (2012). Mental rotation of letters, body parts and complex scenes: separate or common mechanisms? *Hum. Mov. Sci.* 31, 1151–1160. doi: 10.1016/j.humov.2011.12.001
- Decety, J., and Jackson, P. L. (2006). A social-neuroscience perspective on empathy. *Curr. Dir. Psychol. Sci.* 15, 54–58. doi: 10.1111/j.0963-7214.2006.00406.x
- De Jong, R., Liang, C.-C., and Lauber, E. (1994). Conditional and unconditional automaticity: a dual-process model of effects of spatial stimulus-response correspondence. *J. Exp. Psychol. Hum. Percept. Perform.* 20, 731–750. doi: 10.1037/0096-1523.20.4.731
- De Vega, M., Rodrigo, M. J., and Zimmer, H. (1996). Pointing and labelling directions in egocentric frameworks. *J. Mem. Lang.* 35, 821–839. doi: 10.1006/jmla.1996.0042
- Diwadkar, V. A., and McNamara, T. P. (1997). Viewpoint dependence in scene recognition. *Psychol. Sci.* 8, 302–307. doi: 10.1111/j.1467-9280.1997.tb00442.x
- Duran, N. D., Dale, R., and Kreuz, R. J. (2011). Listeners invest in an assumed other's perspective despite cognitive costs. *Cognition* 121, 22–40. doi: 10.1016/j.cognition.2011.06.009

- Easton, R. D., and Sholl, M. J. (1995). Object-array structure, frames of reference, and retrieval of spatial knowledge. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 483–500. doi: 10.1037/0278-7393.21.2.483
- Gardner, M. R., and Potts, R. (2010). Hand dominance influences the processing of observed bodies. *Brain Cogn.* 73, 35–40. doi: 10.1016/j.bandc.2010.02.002
- Gardner, M. R., and Potts, R. (2011). Domain general mechanisms account for imagined transformations of whole body perspective. *Acta Psychol.* 137, 371–381. doi: 10.1016/j.actpsy.2011.04.006
- Gardner, M. R., Sorhus, I., Edmonds, C. J., and Potts, R. (2012). Sex differences in components of imagined perspective transformation. *Acta Psychol.* 140, 1–6. doi: 10.1016/j.actpsy.2012.02.002
- Gronholm, P. C., Flynn, M., Edmonds, C. J., and Gardner, M. R. (2012). Empathic and non-empathic routes to visuospatial perspective-taking. *Conscious. Cogn.* 21, 494–500. doi: 10.1016/j.concog.2011.12.004
- Hommel, B. (1993). The relationship between stimulus processing and response selection in the Simon task: evidence for a temporal overlap. *Psychol. Res.* 55, 280–290. doi: 10.1007/bf00419688
- Ionta, S., and Blanke, O. (2009). Differential influence of hands posture on mental rotation of hands and feet in left and right handers. *Exp. Brain Res.* 195, 207–217. doi: 10.1007/s00221-009-1770-0
- Jola, C., and Mast, F. W. (2005). Mental object rotation and egocentric body transformation: two dissociable processes? *Spat. Cogn. Comput.* 5, 217–237. doi: 10.1080/13875868.2005.9683804
- Keehner, M., and Fischer, M. H. (2012). Unusual bodies, uncommon behaviors: individual and group differences in embodied cognition in spatial tasks. *Spat. Cogn. Comput.* 12, 71–82. doi: 10.1080/13875868.2012.659303
- Keehner, M., Guerin, S. A., Miller, M. B., Turk, D. J., and Hegarty, M. (2006). Modulation of neural activity by angle rotation during imagined spatial transformations. *Neuroimage* 33, 391–398. doi: 10.1016/j.neuroimage.2006.06.043
- Kelly, J. W., Avraamides, M. N., and Loomis, J. M. (2007). Sensorimotor alignment effects in the learning environment and in novel environments. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 1092–1107. doi: 10.1037/0278-7393.33.6.1092
- Kessler, K., and Rutherford, H. (2010). The two forms of visuospatial perspective taking are differently embodied and subserve different spatial prepositions. *Front. Psychol.* 1:213. doi: 10.3389/fpsyg.2010.00213.
- Kessler, K., and Thomson, L. A. (2010). The embodied nature of spatial perspective taking: embodied transformation versus sensorimotor interference. *Cognition* 114, 72–88. doi: 10.1016/j.cognition.2009.08.015
- Kockler, H., Scheef, L., Tepest, R., David, N., Bewernick, B. H., Newen, A., et al. (2010). Visuospatial perspective taking in a dynamic environment: perceiving moving objects from a 1st-person-perspective induces a disposition to act. *Conscious. Cogn.* 19, 690–701. doi: 10.1016/j.concog.2010.03.003
- May, M. (2004). Imaginal perspective switches in remembered environments: transformation vs. interference accounts. *Cogn. Psychol.* 48, 163–206. doi: 10.1016/s0010-0285(03)00127-0
- May, M. (2007). Imaginal repositioning in everyday environments: effects of testing method and setting. *Psychol. Res.* 71, 277–287. doi: 10.1007/s00426-006-0083-1
- May, M., and Wendt, M. (2012). Separating mental transformations and spatial compatibility effects in the own body transformation task. *Cogn. Process.* 13, S257–S260. doi: 10.1007/s10339-012-0455-y
- Mazzarella, E., Ramsey, R., Conson, M., and Hamilton, A. (2013). Brain systems for visual perspective taking and action perception. *Soc. Neurosci.* 8, 248–267. doi: 10.1080/17470919.2012.761160
- Michelon, P., and Zacks, J. M. (2006). Two kinds of visual perspective taking. *Percept. Psychophys.* 68, 327–337. doi: 10.3758/bf03193680
- Mohr, C., Blanke, O., and Brugger, P. (2006). Perceptual aberrations impair mental own-body transformations. *Behav. Sci.* 120, 528–534. doi: 10.1037/0735-7044.120.3.528
- Mohr, C., Rowe, A. C., and Blanke, O. (2010). The influence of sex and empathy on putting oneself in the shoes of others. *Br. J. Psychol.* 101, 277–291. doi: 10.1348/000712609x457450
- Parsons, L. M. (1987). Imagined spatial transformations of one's body. *J. Exp. Psychol. Gen.* 116, 172–191. doi: 10.1037/0096-3445.116.2.172
- Proctor, R. W., and Vu, K. L. (2006). *Stimulus-Response Compatibility Principles*. Boca Raton, FL: Taylor and Francis.
- Rieser, J. J. (1989). Access to knowledge of spatial structure at novel points of observation. *J. Exp. Psychol. Learn. Mem. Cogn.* 15, 1157–1165. doi: 10.1037/0278-7393.15.6.1157
- Shelton, A. L., and McNamara, T. P. (1997). Multiple views of spatial memory. *Psychon. Bull. Rev.* 4, 102–106. doi: 10.3758/bf03210780
- Shepard, R. N., and Cooper, L. A. (1982). *Mental Images and Their Transformations*. Cambridge, MA: MIT Press.
- Sohn, M. H., and Carlson, R. A. (2003). Viewpoint alignment and response conflict during spatial judgment. *Psychon. Bull. Rev.* 10, 907–916. doi: 10.3758/bf03196551
- Stegemann, Y., Engbert, K., and Weigelt, M. (2011). Selective effects of motor expertise in mental body rotation tasks: comparing object-based and perspective transformations. *Brain Cogn.* 76, 97–105. doi: 10.1016/j.bandc.2011.02.013
- Surtees, A. D. R., and Apperly, I. A. (2012). Egocentrism and automatic perspective taking in children and adults. *Child Dev.* 83, 452–460. doi: 10.1111/j.1467-8624.2011.01730.x
- Tadi, T., Overney, L. S., and Blanke, O. (2009). Three sequential brain activations encode mental transformations of upright and inverted human bodies: a high resolution evoked potential study. *Neuroscience* 159, 1316–1325. doi: 10.1016/j.neuroscience.2009.02.012
- Tarr, M. J., and Bülhoff, H. H. (1998). Image-based object recognition in man, monkey and machine. *Cognition* 67, 1–20. doi: 10.1016/s0010-0277(98)00026-2
- Thakkar, K. N., and Park, S. (2010). Empathy, schizotypy, and visuospatial transformations. *Cogn. Neuropsychiatry* 15, 477–500. doi: 10.1080/13546801003711350
- Vogele, K., May, M., Ritzl, A., Falkai, P., Zilles, K., and Fink, G. R. (2004). Neural correlates of first-person-perspective as one constituent of human self-consciousness. *J. Cogn. Neurosci.* 16, 817–827. doi: 10.1162/08992904970799
- Wang, R. F. (2004). Action, verbal response and spatial reasoning. *Cognition* 94, 185–192. doi: 10.1016/j.cognition.2004.05.001
- Wraga, M., Boyle, H. K., and Flynn, C. M. (2010). Role of motor processes in extrinsically encoding mental transformations. *Brain Cogn.* 74, 193–202. doi: 10.1016/j.bandc.2010.07.005
- Wraga, M., Creem, S. H., and Proffitt, D. R. (2000). Updating displays after imagined object and viewer rotations. *J. Exp. Psychol. Learn. Mem. Cogn.* 26, 151–168. doi: 10.1037/0278-7393.26.1.151
- Yu, A. B., and Zacks, J. M. (2010). The role of animacy in spatial transformations. *Mem. Cognit.* 38, 982–993. doi: 10.3758/mc.38.7.982
- Zacks, J. M., Mires, J., Tversky, B., and Hazeltine, E. (2000). Mental spatial transformations of objects and perspective. *Spat. Cogn. Comput.* 2, 315–332. doi: 10.1023/a:1015584100204
- Zacks, J. M., Ollinger, J. M., Sheridan, M. A., and Tversky, B. (2002). A parametric study of mental spatial transformations of bodies. *Neuroimage* 16, 857–872. doi: 10.1006/nimg.2002.1129
- Zacks, J. M., Rypma, B., Gabrieli, J. D. E., Tversky, B., and Glover, G. H. (1999). Imagined transformation of bodies: an fMRI investigation. *Neuropsychologia* 37, 1029–1040. doi: 10.1016/s0028-3932(99)00012-3
- Zacks, J. M., and Tversky, B. (2005). Multiple systems for spatial imagery: transformations of objects and bodies. *Spat. Cogn. Comput.* 5, 271–306. doi: 10.1207/s15427633scc0504_1
- Zacks, J. M., Vettel, J. M., and Michelon, P. (2003). Imagined viewer and object rotations dissociated with event-related fMRI. *J. Cognit. Neurosci.* 15, 1002–1018. doi: 10.1162/08992903770007399
- Zhang, J. X., and Johnson, M. K. (2004). A memory-based, Simon-like, spatial congruence effect: evidence for persisting spatial codes. *Q. J. Exp. Psychol.* A 57, 419–436. doi: 10.1080/02724980343000404
- Zwicker, J. (2009). Agency attribution and visuospatial perspective taking. *Psychon. Bull. Rev.* 16, 1089–1093. doi: 10.3758/pbr.16.6.1089
- Zwicker, J., and Müller, H. (2013). On the relation between spontaneous perspective taking and other visuospatial processes. *Mem. Cogn.* 41, 558–570. doi: 10.3758/s13421-012-0283-4

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 26 June 2013; accepted: 20 August 2013; published online: 06 September 2013.

Citation: May M and Wendt M (2013) Visual perspective taking and laterality decisions: problems and possible

solutions. *Front. Hum. Neurosci.* 7:549. doi: 10.3389/fnhum.2013.00549

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2013 May and Wendt. This is an open-access article

distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the

original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Minimal self-models and the free energy principle

Jakub Limanowski^{1*} and Felix Blankenburg^{1,2,3}

¹ Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, Berlin, Germany

² Dahlem Institute for Neuroimaging of Emotion, Freie Universität Berlin, Berlin, Germany

³ Center for Adaptive Rationality (ARC), Max Planck Institute for Human Development, Berlin, Germany

Edited by:

Antonia Hamilton, University of Nottingham, UK

Reviewed by:

Jakob Hohwy, Monash University, Australia

Matthew Apps, University of Oxford, UK

*Correspondence:

Jakub Limanowski, Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, Luisenstraße 56, Haus 1, 10117 Berlin, Germany
e-mail: jakub.limanowski@hu-berlin.de

The term “minimal phenomenal selfhood” (MPS) describes the basic, pre-reflective experience of being a self (Blanke and Metzinger, 2009). Theoretical accounts of the minimal self have long recognized the importance and the ambivalence of the body as both part of the physical world, and the enabling condition for being in this world (Gallagher, 2005a; Grafton, 2009). A recent account of MPS (Metzinger, 2004a) centers on the consideration that minimal selfhood emerges as the result of basic self-modeling mechanisms, thereby being founded on pre-reflective bodily processes. The free energy principle (FEP; Friston, 2010) is a novel unified theory of cortical function built upon the imperative that self-organizing systems entail hierarchical generative models of the causes of their sensory input, which are optimized by minimizing free energy as an approximation of the log-likelihood of the model. The implementation of the FEP via predictive coding mechanisms and in particular the active inference principle emphasizes the role of embodiment for predictive self-modeling, which has been appreciated in recent publications. In this review, we provide an overview of these conceptions and illustrate thereby the potential power of the FEP in explaining the mechanisms underlying minimal selfhood and its key constituents, multisensory integration, interoception, agency, perspective, and the experience of mineness. We conclude that the conceptualization of MPS can be well mapped onto a hierarchical generative model furnished by the FEP and may constitute the basis for higher-level, cognitive forms of self-referral, as well as the understanding of other minds.

Keywords: free energy principle, predictive coding, active inference, self, minimal phenomenal selfhood, ownership, agency, self-model

INTRODUCTION

What lets an organism be a self? Throughout philosophical attempts to understand the enabling conditions of minimal self-awareness (Zahavi, 1999), or *minimal phenomenal selfhood* (MPS)¹ (Blanke and Metzinger, 2009), the special status of the body among all other physical things has long been apparent (Merleau-Ponty, 1962; Bermúdez et al., 1998; Anderson and Perlis, 2005; Legrand, 2006; Blanke, 2012). Recently, the role of the human body for cognition has been re-emphasized in the field of embodied cognition (Varela et al., 1994; Clark, 1999; Gallagher, 2005a; Grafton, 2009; Gallese and Sinigaglia, 2011).

¹In general, this approach is concerned with “global aspects of bodily self-consciousness” (Blanke and Metzinger, 2009), where a *global* property is something that can only be ascribed to a system as a whole, and *self-consciousness* refers to “the ability to become aware of one’s own mental and bodily states . . . as one’s own mental and bodily states” (Vogeley and Fink, 2003). The kind of self-consciousness meant here is not cognitive but “immediate, pre-reflective and non-observational” (see also Zahavi, 1999; Gallagher, 2005a; Legrand, 2006; Hohwy, 2007), where the term pre-reflective is referring to levels of self-awareness that are independent of explicit cognition and linguistic abilities (Blanke and Metzinger, 2009). In its simplest form, this is the *minimal phenomenal self*, the “fundamental conscious experience of being someone” (Blanke and Metzinger, 2009).

The body lets us interact with the world via perception and action (Legrand, 2006; Friston, 2011; Farmer and Tsakiris, 2012), leading to a whole new form of intelligence that is different from, for example, mere computation (Frith, 2007; Grafton, 2009). One’s everyday experience is enabled and structured through a body that is “always there” (James, 1890), and hence the body—*my* body—is not just part of the physical world, but also the “vehicle” that enables being a self in this world (Merleau-Ponty, 1962; Varela et al., 1994; Gallagher, 2005a). Minimal, or pre-reflective selfhood emerges from this experience of a unified, situated living body as a “sensorimotor unity anchored to its world” (Bermúdez et al., 1998; Anderson and Perlis, 2005; Gallagher, 2005a; Legrand, 2006; Hohwy, 2010; Blanke, 2012; Apps and Tsakiris, 2013).

In this review, we will particularly consider an account of the mechanisms giving rise to minimal selfhood that has recently been proposed by Metzinger (2003, 2004a,b, 2005). Central to the theory is the premise that minimal selfhood emerges as the result of pre-reflective self-modeling, i.e., through an organism’s model of the world that is phenomenologically centered onto the self. Thereby, Metzinger’s account builds on the proposition that the brain is a representational system that needs to interpret the

world (Gallese and Metzinger, 2003), and thus constructs and simulates a model in order to reduce ambiguity originating from the external world (Metzinger, 2005). For this system-model to be successful, i.e., of adaptive value, “the self needs to be embedded into the causal network of the physical world” (Knoblich et al., 2003; Metzinger, 2004a, 2005). The model thus also has to include as part of itself the physical body—“the part of the simulation that represents the system itself” (Edelman, 2008, p. 419). Metzinger (2004a) emphasizes that this self-representation of the system is special in that it (i.e., the body) is the only representational structure that constantly generates and receives internal input via its different intero- and proprioceptive systems. Notably, a resulting structural property of the system-model is the spatiotemporal centeredness of the model onto a coherent phenomenal subject, described by Metzinger with the term *perspectivalness* (Metzinger, 2004a, 2005; Blanke and Metzinger, 2009). Throughout this review, we will return to this, and propose to understand it as an instance of “perspective taking”, whereby the brain assigns the subjective, first-person perspective (1PP) to its self-model.

Following their emphasis of self-modeling mechanisms for minimal selfhood, Metzinger and colleagues (Knoblich et al., 2003) have argued that an analysis of selfhood should focus on the underlying *functional* properties of the system, i.e., the brain. In this review, we will examine one promising candidate brain theory for this analysis: over the last years, a general theoretical account of cortical function based on the “free energy principle” (FEP) has been put forth by Friston (Friston et al., 2006; Friston, 2009, 2010; Clark, 2013), based on the conclusive assumption that the brain entails hierarchical dynamical models to predict the causes of its sensory data (Hohwy, 2007; Frith, 2007; Friston and Kiebel, 2009; Bubic et al., 2010).

The key premise of the FEP is that self-organizing organisms have to resist the natural tendency to disorder that is implied by the second law of thermodynamics, i.e., they have to “maintain their states and form in the face of a constantly changing environment” (Friston, 2010). Organisms do so by avoiding *surprise* associated with their sensory states (Friston et al., 2011, 2012; Friston, 2012a,b), which in turn will result in a (desired) state where the world is highly predictable. The FEP proposes that the brain infers the hidden causes of the environment via the inversion of hierarchical generative models that predict their sensory consequences (Friston, 2010; Bastos et al., 2012), with higher levels encoding increasingly abstract and information-integrating conceptions of the world (Fotopoulou, 2012; Clark, 2013). Importantly, as biological organisms are embodied in the environment, the “world-model” of a self-organizing system also has to include the sensory apparatus (the body) of the organism (Friston, 2012b; Friston et al., 2012; Clark, 2013). In agreement with the Good Regulator theorem (Conant and Ashby, 1970; Edelman, 2008; Friston et al., 2012), which states that every good regulator of a system will ultimately become a model of that system, the FEP thus proposes as a consequence of hierarchical predictive modeling that “I model myself as existing” (Friston, 2011, 2013b). We will later highlight that this conforms nicely to accounts of minimal selfhood, whereby the self is perceived as a result of dynamic self-modeling mechanisms (Metzinger, 2004a; Hohwy, 2007).

Conceptually, the FEP is based on the evaluation of the improbability of some sensory data under a hierarchical generative model, where the (model-conditional) improbability of the data is commonly referred to as *surprise* (Friston et al., 2006; Friston, 2010, 2011). The theory builds on *free energy* as an information-theoretical quantity on the upper bound of surprise that can be formally assessed (Friston et al., 2006, 2012; Friston, 2010, 2011). By minimizing free energy within a model, biological agents thus always also minimize surprise. In principle, this can be done in two ways: By changing the *predictions of the model* by means of perception, or by changing *what is predicted* by selectively sampling those sensations that confirm the model’s predictions by means of action (a “systematic bias in input sampling”, Verschure et al., 2003; Friston, 2011).

Minimizing surprise associated with sensory data by the inversion of the hierarchical generative model (and the dynamic optimization of its parameters) has been established as *predictive coding* (Srinivasan et al., 1982; Mumford, 1992; Rao and Ballard, 1999; Friston, 2005a; Friston and Stephan, 2007; Kilner et al., 2007; Friston and Kiebel, 2009). Thereby, the predictive coding scheme infers the hidden causes of its sensory input by minimizing the difference between the predictions about sensory data and the actual sensory data at any level of the model’s hierarchy, which is encoded by the *prediction error* (Friston and Kiebel, 2009; Bubic et al., 2010; Friston, 2010; Brown and Brüne, 2012; Friston, 2012a). Thus the *feedforward* signal is not the sensory information *per se*, but the associated prediction error that is passed up the hierarchy (Hohwy, 2012; Clark, 2013), while the generative model’s predictions are the *feedback* signal (Friston, 2010; Bastos et al., 2012; Edwards et al., 2012). The second form of prediction error minimization via interaction with the environment is described under the *active inference* principle (Friston, 2012a, 2013a). Reminiscent of “affordances”, Gibson’s (1977) famous description of the fact that the environment is “co-perceived” depending on the perceiver’s bodily endowment, active inference thus emphasizes the bi-directional role of embodiment such that “not only does the agent embody the environment but the environment embodies the agent” (Friston, 2011). Interestingly, the computational assumptions of predictive coding are surprisingly well reflected by neuroanatomical organization of the cortex (Bastos et al., 2012; Friston, 2012a), suggesting that neuronal populations indeed encode probabilities, i.e., uncertainty (Clark, 2013). In sum, predictive coding and active inference are neurobiologically plausible, “action-oriented” (Bastos et al., 2012; Clark, 2013) implementations of free energy minimization (Friston, 2011; Bastos et al., 2012; Friston, 2012a; Clark, 2013).

In this review, we summarize recently formulated free energy accounts of key aspects of minimal selfhood: multisensory integration, interoception, agency, ownership or “mineness” of experience, the perspectivity of self-models and models of other selves. Common to these FEP applications is the focus on “self modeling” (Friston, 2012a). We hence consider these approaches in the light of the proposal that the minimal self is the result of an ongoing predictive process within a generative model that is centered onto the organism (Metzinger, 2004a; Hohwy, 2007; Friston, 2011).

ASPECTS OF THE MINIMAL SELF IN THE FREE ENERGY FRAMEWORK

A number of publications have recently put forward the idea that (minimal) selfhood is based on the neurobiological implementation of hierarchical generative models in the brain (Hohwy, 2007, 2010; Seth et al., 2011; Fotopoulou, 2012; Friston, 2012a,b; Apps and Tsakiris, 2013; Clark, 2013). In one sentence, these accounts propose to “understand the elusive sense of minimal self in terms of having internal models that successfully predict or match the sensory consequences of our own movement, our intentions in action, and our sensory input” (Hohwy, 2007). In accordance with Friston (2011, 2012b, 2013b), who has already emphasized the fundamental, bi-directional role of embodiment in the FEP, these accounts also embrace the body as a central part of the self-model. The aspects of the minimal self that these approaches formalize in the FEP all follow as consequences from this embodied self-modeling (Metzinger, 2004a; Hohwy, 2007; Friston, 2011): The body predicts and integrates multisensory information in a way that no other physical object does (Hohwy, 2007, 2010; Apps and Tsakiris, 2013), it is the only source of internally generated input (Seth et al., 2011; Critchley and Seth, 2012), it is crucial for interaction with the environment and a sense of agency (Kilner et al., 2007; Frith, 2007; Friston et al., 2011). From the phenomenological and spatiotemporal centeredness of experience onto the body (Friston, 2011) emerges the 1PP, and ultimately, the “mineness” of experience (Hohwy, 2007; Apps and Tsakiris, 2013).

MULTISENSORY INTEGRATION

A very important implication of the free energy framework is that sensory information is processed probabilistically, and thus it follows that the representation of the self is also probabilistic (Friston, 2011). This conceptualization fits comfortably with Metzinger’s (2004b) theory, where the content of the self-model is probabilistic, i.e., it is “simply the best hypothesis about the current state of the system, given all constraints and information resources currently available” (see also Hohwy, 2010; Clark, 2013; Friston, 2013b). However, sensory information is not *per se* specific to the self, which implies that there must be additional levels of information processing in which information is related to the self (Apps and Tsakiris, 2013).

Previous accounts of bodily self-awareness, inspired by work on illusions of body ownership and related paradigms, have emphasized the role of multimodal, hierarchical cortical networks in processing self-related information (Hohwy, 2007, 2010; Tsakiris, 2010; Petkova et al., 2011a; Blanke, 2012). In a recent paper, Apps and Tsakiris (2013) propose that hierarchical prediction error minimization can explain processes of self-recognition and self-representation: for the processing of information relating to the self, free energy minimization happens via the integration of various streams of surprise from unimodal sensory information in hierarchically higher multimodal areas, where information from any system can be used to “explain away” surprise in any other system (Hohwy, 2010; Apps and Tsakiris, 2013; Clark, 2013). This corresponds to the basic claim of predictive coding about crossmodal information processing, according to which hierarchically higher levels form amodal concepts that

generate multimodal predictions and prediction errors (Friston, 2012a). Following this logic, higher-level multisensory areas must predict input in multiple sensory modalities, which according to Apps and Tsakiris (2013) implies “a high level representation (of self) that elaborates descending predictions to multiple unimodal systems” (see also Clark, 2013; Friston, 2013b). This self-model can thus be seen as the most accurate, immediately available explanation of the bottom-up surprise from incoming multisensory information (Apps and Tsakiris, 2013; thereby the model need not be “true”, just a *sufficient* explanation of the sensory input, Schwabe and Blanke, 2008; Hohwy and Paton, 2010; Hohwy, 2012). The predictive coding account suggests that, at the hierarchically highest level, such a self-model will encode, as model evidence, the evidence for the existence of the agent in the present form (Hohwy, 2010; Friston, 2011).

A particularly intriguing example of how self-representation is constructed in a probabilistic way is the rubber hand illusion (RHI; Botvinick and Cohen, 1998): observing a dummy hand being touched, while receiving synchronous tactile stimulation at the anatomically congruent location of one’s real, hidden hand typically leads to an illusory experience of feeling the touch on the dummy hand (Botvinick and Cohen, 1998; Ehrsson et al., 2004, 2005; Makin et al., 2008). This usually results in a self-attribution, or “incorporation” (Holmes and Spence, 2004) of the fake hand as a part of one’s own body (Tsakiris and Haggard, 2005; Hohwy and Paton, 2010; Tsakiris, 2010; Petkova et al., 2011a). A number of behavioral measures such as a fear response to the dummy hand being threatened (Armel and Ramachandran, 2003; Ehrsson et al., 2007), or the mislocalization of one’s real hand towards the location where the dummy hand is seen (Botvinick and Cohen, 1998; Tsakiris and Haggard, 2005), suggest that the brain indeed seems to treat the dummy hand as part of the body as a result of the multisensory stimulation (see Tsakiris, 2010, or Blanke, 2012, for detailed reviews). Using virtual reality techniques, the RHI paradigm has been extended to induce an illusory self-identification with a whole dummy body located at a different position in space (Ehrsson, 2007; Lenggenhager et al., 2007). In those cases, participants exhibited a bias in judging their own spatial location towards the location where the dummy body was positioned in space, just as the mislocalization of the own hand during the RHI (see Blanke, 2012, for a review). These findings thus impressively demonstrate that perceived self-location can be manipulated with appropriate stimulation.

Generally, illusory percepts are well explained as a result of Bayes-optimal inference, i.e., arising from an interpretation of ambiguous sensory input under strong prior hypotheses (Friston, 2005b; Brown and Friston, 2012; Apps and Tsakiris, 2013; Clark, 2013). Correspondingly, a combination of bottom-up input and modulatory top-down factors has been suggested to drive illusory ownership of body parts as experienced during the RHI (de Vignemont et al., 2005; Tsakiris and Haggard, 2005; de Preester and Tsakiris, 2009; Hohwy and Paton, 2010; Tsakiris, 2010). While congruent multisensory input seems crucial for the RHI (Botvinick and Cohen, 1998; Armel and Ramachandran, 2003; Ehrsson et al., 2004, 2005; Hohwy and Paton, 2010; Petkova et al., 2011a), there have been strong arguments for top-down “body representations” that define which objects

(namely, only anatomically plausible hand-shaped objects, see e.g., Tsakiris and Haggard, 2005) can be incorporated during the RHI (de Vignemont et al., 2005; IJsselstein et al., 2006; Costantini and Haggard, 2007; Tsakiris et al., 2007; de Preester and Tsakiris, 2009). However, various inconsistent definitions of body representations may have lead to some confusion and thus prevented the emergence of a unifying theoretical account (de Vignemont, 2007; Longo et al., 2008; Apps and Tsakiris, 2013).

As a solution to this problem, several authors have endorsed a predictive coding approach (Hohwy, 2007, 2010; Apps and Tsakiris, 2013). Consider that, under normal circumstances, observed touch on our skin is accompanied by a corresponding, temporally congruent tactile sensation—in predictive coding terms, the underlying generative model of our physical self predicts a somatosensory sensation when touch is about to occur on the body, because associations between events that have a high probability of predicting events in another system lead to the formation of beliefs, or priors on a hierarchically higher level (Apps and Tsakiris, 2013). Note that it are not *per se* the associations between different kinds of sensory input that are of importance here, but the parallel predictions of the generative model. Among all physical objects in the world, it is only our body that will evoke (i.e., predicts) this kind of multisensory sensation—congruence of multisensory input has (not surprisingly) been called “self-specifying” (Botvinick, 2004) and has been ascribed a crucial role in self-representation (Botvinick and Cohen, 1998; Armel and Ramachandran, 2003; Ehrsson et al., 2005; Hohwy and Paton, 2010). Following this logic, during the RHI, surprise² or prediction error is evoked by the simultaneous occurrence of observed touch on an external object (the dummy hand) together with a somatosensory sensation, because such congruence is not predicted by the brain’s initial generative model.

The predictive coding account suggests that, as stimuli can usually be caused “in an infinite number of ways” (Brown and Friston, 2012), there are several competing explanations of the sensory input between which the brain needs to decide. In the case of the RHI, these are coded by the probabilities of the actual hand, or the dummy hand being “me” (Apps and Tsakiris, 2013). One explanation, or model, of the sensory input is that vision and touch occur at different locations (the “true” model, Hohwy, 2010). However, during the RHI, spatially distributed observed and felt touch are “bound together” by causal inference (Hohwy, 2012): this “false” model (that observed and felt touch occur at the same location, namely, one’s own hand) is selected because it more successfully explains the incoming prediction error in favor of a unified self (see also Schwabe and Blanke, 2008; Hohwy, 2010; Hohwy and Paton, 2010). This is a crucial point, because

²Although the illusory experience of feeling the touch on the dummy hand is certainly surprising, one has to distinguish this cognitive surprise of the agent from “surprise” on a neurobiological level (“surprisal”, see Clark, 2013), as defined by prediction error. In fact, here these two notions may be somewhat opposing: the dummy hand is accepted as a part of one’s body as a result of successfully *explaining away* the surprise evoked by the ambiguous multisensory stimulation (Hohwy, 2010; Hohwy and Paton, 2010). However, the agent experiences exactly this state—owning a lifeless dummy hand—as surprising.

predictive coding is a “winner takes all” strategy (Hohwy, 2007, 2010): there is always one model that has the lowest amount of free energy (the highest model evidence) among all possible models of the sensory input (Friston et al., 2012; Apps and Tsakiris, 2013; Clark, 2013), and this model is selected as the explanation for the world. This model does not have to be “true”, just a better explanation of the sensory input than competing models (Friston et al., 2012). As minimizing surprise is the same as maximizing model-evidence (where model-evidence is evidence for the agent’s existence), the agent, or self, in its present form will cease to exist if another model has to be chosen as a better explanation of sensory input (Hohwy, 2010; Friston, 2011): “I” (i.e., the embodied model of the world) will only exist “iff (sic) I am a veridical model of my environment” (Friston, 2011).

Applied to the RHI example, this means that if prediction error could not be explained away in this way, the system might have to dismiss its current self-model in favor of a better explanation of the input—which would result in the representation of a “disunified self” (Hohwy, 2010). The FEP states that, if prediction error can be explained away at lower levels, there is no need to adjust higher-level representations (Friston, 2012a). Apps and Tsakiris (2013) propose that, as the prediction error is passed up the hierarchy during the RHI, it can be explained away at multimodal cortical nodes. Thereby “explaining away” means an updating of the generative model’s predictions about the physical features of the self to minimize the overall level of surprise in the system. This results in a different posterior probabilistic representation of certain *features* of the self (Hohwy and Paton, 2010; Apps and Tsakiris, 2013), however, without any necessity to change the actual generative self-model (Hohwy, 2010). Specifically, the dummy hand is now probabilistically more likely to be represented as part of one’s body, which in turn is accompanied by a decrease in the probability that one’s actual hand will be represented as “self”. This manifests as a self-attribution of the dummy hand, and a partial rejection of the real limb (de Preester and Tsakiris, 2009; Tsakiris, 2010).

Indeed, there is compelling experimental evidence in support of such a probabilistic integration process underlying the RHI. For example, the mislocalization of one’s real hand towards the location of the dummy hand is never absolute, but relative; participants usually judge the location of their hand several centimeters closer to the dummy, but not at the same location (Tsakiris and Haggard, 2005). Lloyd (2007) showed that the RHI gradually decreases with increasing distance between the own and the dummy hand. Furthermore, a drop in skin temperature of the stimulated real hand was found to accompany the RHI (Moseley et al., 2008), which has been interpreted as evidence for top-down regulations of autonomic control and interoceptive prediction error minimization during the RHI (Moseley et al., 2008; Seth et al., 2011; Suzuki et al., 2013). Also, after the illusion, the dummy hand is frequently perceived as more similar to one’s real hand (Longo et al., 2009). These findings suggest that in fact, explaining away prediction error from ambiguous multisensory stimulation may lead to changes in the encoded features of the self (Hohwy and Paton, 2010).

The idea of a probabilistic self-representation in the brain benefits from the fact that the free energy account is relatively

unconstrained and thus not as heavily dependent on conceptual assumptions as other theories (Hohwy, 2007, 2010; Friston, 2008; Friston and Kiebel, 2009; Friston et al., 2012). Thus the FEP does not need to treat information relating to the self as a distinct class of information (Apps and Tsakiris, 2013), because it is concerned with information flow and system structure. For example, the matching of sensory predictions based on corollary discharge with actual sensory input has been previously proposed as a basis for self-awareness (see Gallagher, 2000; Brown et al., 2013). In the free energy account, however, self-awareness is not restricted to the integration of sensorimotor efference and re-efference. Rather, *any* type of sensory information can be integrated within a multimodal, abstract representation of the self, and explain away surprise in another system (Apps and Tsakiris, 2013). The RHI example demonstrates that, as claimed by the FEP (Friston, 2012a), if prediction error can be explained away in the periphery (e.g., adjusting the encoded location of one's real hand), there is no need to adjust higher-level representations (the unified self-model). The FEP is thus a parsimonious, and hence inherently flexible, formal description of how multisensory information integration underpins minimal forms of self-awareness (Hohwy, 2010; Blanke, 2012).

INTEROCEPTION

A special case of information that the self-model receives is input from interoceptive senses: within the world-model, the (own) body is special among all physical objects in that it constantly receives a “background buzz” of somatosensory input, including input from somato-visceral and mechanoreceptors, and higher-level feeling states (Metzinger, 2004a, 2005; see Friston, 2011). Acknowledging the importance of interoception, recent work by Seth (Critchley and Seth, 2012; Seth et al., 2011; Suzuki et al., 2013) has promoted interoceptive prediction error minimization as a mechanism for self-representation. Specifically, Seth et al. provide a predictive coding account of “presence”, where presence means the subjective experience of being in the here and now (see Metzinger, 2004a). Presence is hence a structural property of conscious experience (Seth, 2009) that is transparent in the sense that Metzinger (2003) uses the term (Seth et al., 2011). According to Seth et al. (2011), interoceptive predictions arise from autonomic control signals and sensory inputs evoked by motor control signals. The generative model of the causes of interoceptive input gives rise to “interoceptive self-representations” and “emotional feeling states” (Suzuki et al., 2013). Presence results as the successful suppression of the associated prediction error (Seth et al., 2011), more specifically, “self-consciousness is grounded on the feeling states that emerge from interaction of interoceptive predictions and prediction errors” (Critchley and Seth, 2012). The emphasis on subjective feeling states (Critchley et al., 2004; Seth et al., 2011) as a key component of interoceptive predictive coding links this account to emotion frameworks like the somatic marker hypothesis (Damasio, 1999; Bechara et al., 2000).

Half a century ago, Schachter and Singer (1962) showed that people seek explanations for their bodily sensations after having become aware of them. Reversing this argument, Pennebaker and Skelton (1981) showed that the perception of bodily sensations

depended on the hypotheses held by the participants, and was thus not different from the processing of any other ambiguous information. More recently, Moseley et al. (2008) found that the RHI led to a cooling of participants' real hand (and only the hand affected by the illusion), and concluded that there is a causal link between self-awareness and homeostatic regulation, where bodily self-awareness regulates physiological processing in a top-down manner. In accordance with these results, the FEP indicates that interoceptive predictions are “one—among many—of multimodal predictions that emanate from high-level hypotheses about our embodied state.” (Friston, 2013b; Suzuki et al., 2013). Interestingly, as we will see later (see *Modeling Others*), these predictions can also be used to model others' internal states (Bernhardt and Singer, 2012). In sum, although predictive coding accounts of interoception still need detailed work, the corresponding emphasis of interoceptive signals by predictive coding (Seth et al., 2011) and philosophical (Metzinger, 2004a) accounts of the self promises many insightful studies to come.

ACTION AND AGENCY

Agency as a “sense of initiative” (Edelman, 2008) has been emphasized as a key component of MPS (Gallagher, 2000; Metzinger, 2004a; Frith, 2007). Distinguishing between self-initiated actions and actions of other organisms is crucial for being a self. The importance of the motor system in the brain's ontology (interpretation) of the world (Gallese and Metzinger, 2003) has been promoted by forward models of agency based on corollary discharge (Blakemore et al., 2002; Gallagher, 2005a; Frith, 2012), which have also been applied to describe disturbances of agency resulting from a failure of these mechanisms (Gallagher, 2000). Advancing on these accounts, action and the phenomenology of agency have both been accounted for in terms of hierarchical generative models (Hohwy, 2007).

The active inference principle is of central importance in the FEP (Friston and Stephan, 2007; Hohwy, 2007, 2010; Kilner et al., 2007; Brown et al., 2013; Friston, 2013a): action changes the sensory input of an organism so that it better corresponds to the current generative model, without having to revise the model parameters (Friston and Stephan, 2007; Hohwy, 2010). This validation of the current generative system-model is a confirmation of the agent's existence (Friston, 2011). However, for active inference to be feasible, the agent has to be able to predict which actions will lead to a better confirmation of its predictions. Friston (2012b) thus states that “implicit in a model of sampling is a representation or *sense of agency*”, since the effects of selective sampling of sensations as through active inference have to be known—modeled—as well. Thus, by selectively sampling sensations so that they confirm the model's predictions, action is a form of “reality testing” (Hohwy, 2007). For instance, consider that the induction of illusory limb or body ownership via multisensory stimulation (like in the RHI) only works because this kind of active inference is suppressed.³ If allowed, participants would

³But, as pointed out by Hohwy (2007, 2010), active inference is still happening at a more subtle level, as participants focus their attention on the rubber hand to detect potential mismatches of observed and felt touch.

probably instantaneously move their hand to *test* whether the rubber hand moves as well. The illusion will be immediately abolished once participants see that the rubber hand does not move according to their intentions (IJsselstein et al., 2006; Slater et al., 2009; Maselli and Slater, 2013), because now there is a clear mismatch between predicted and actual sensory outcome, which cannot be explained away.

It is noteworthy that failures in basic inference mechanisms are a likely cause of many symptoms connected to a disturbed sense of agency (Gallagher, 2000; Frith, 2007). As stated by the FEP, probabilistic inference under uncertainty underlies all perception, and it thus seems reasonable to explain abnormal experiences in the same framework (Fletcher and Frith, 2008; Hohwy, 2013). Predictive coding schemes and Bayesian inference have been successfully applied to explain symptoms like delusion formation (Fletcher and Frith, 2008; Hohwy, 2013) or failures in sensory attenuation occurring in schizophrenia (Brown et al., 2013), hysteria or functional symptoms (Edwards et al., 2012), out-of-body experiences (Schwabe and Blanke, 2008), and depersonalization (Seth et al., 2011). In many of these cases, basic mechanisms of active inference fail (Brown et al., 2013), but it is not yet clear whether these symptoms can be explained by failures at low levels alone, or rather by a failure of mechanisms across the hierarchy (Fletcher and Frith, 2008). For instance, a noisy prediction error signal has been suggested as the cause for positive symptoms in schizophrenia (Fletcher and Frith, 2008), while delusions are seen as the result of false inference “at a conceptual level” (Brown et al., 2013), which may be characterized by a “lack of independent sources of evidence for reality testing” (Hohwy, 2013).

In conclusion, action and agency are of fundamental importance for the experience of normal minimal selfhood. However, although a sense of agency (Gallagher, 2000) is sufficient for MPS, it may not be the most basal constituent (Blanke and Metzinger, 2009). What matters is that I experience the action as *mine* (Gallagher, 2000), which brings us to the most important aspect of the generative self-model: the experience of “mineness” (Hohwy, 2007).

MINENESS

The phenomenal experience of “mineness” is a key property of MPS (Metzinger, 2004a). The idea that the living body is experienced as mine (“owned”) can be traced back to early phenomenologists like Merleau-Ponty or Husserl (see Gallagher, 1986, 2009). It has been claimed that this “self-ownership” (Gallagher, 2000) is the most fundamental sense of phenomenal selfhood (Aspell et al., 2009; Blanke and Metzinger, 2009). Similarly, Hohwy (2007) equates experienced mineness of actions and perceptions with the experience of a minimal self.

In Hohwy’s (2007) FEP account of the self, mineness is a general phenomenon, resulting from successful predictions of actions and perceptions. It is hereby important to keep in mind that prediction is more than mere anticipation (Hohwy, 2007; Bubic et al., 2010), but describes predictive *modeling* as a fundamental principle of the brain, and that what is informative in predictive coding is the prediction *error*. Following Hohwy’s (2007) logic, phenomenal selfhood thus arises as a consequence of successfully

having predicted incoming sensory input across the hierarchy of the self-model. Within predictive coding, prediction error is not explained away post-hoc, but constantly, and across all levels of the model (Friston, 2012a). Thus mineness is always *implicit* in the flow of information within the hierarchical generative self-model, and can correspondingly be experienced for actions and perceptions in the same way (note how once again the FEP is simple in its assumptions). Crucially, this means that the minimal self is the result of an ongoing, dynamic process, not a static representation. In this account, mineness is thus situated in a spatiotemporal reference frame (see Metzinger, 2004a; Hohwy, 2007), where prediction introduces the temporal component of “being already familiar” with the predicted input (Hohwy, 2007; see Kiebel et al., 2008; Bubic et al., 2010).

Perhaps a good example for this construction of temporally extended phenomenal experience from predictive processes is the classical concept of a *body schema* (Head and Holmes, 1911–1912; Merleau-Ponty, 1962). The body schema describes the dynamic organization of sensorimotor processes subserving motor and postural functions in a form of “embodied memory” that ultimately presents the body for action (Gallagher, 2009). These processes are pre-reflective, operating “below the level of self-referential intentionality” (Gallagher and Cole, 1995), and thus the body schema is not a static representation (Gallagher, 2005a). But note that the body schema defines the range of possible actions that my body can perform, while being “charged” with what has happened before (see Gallagher, 2009, for a nice review). In the hierarchical generative self-model, the body schema might thus be pictured as encoded by a structure of predictions (e.g., of self-location and proprioception).

In conclusion, the following picture seems to emerge from the reviewed literature: the FEP is capable of describing the functional regularities of the brain’s “ontology” (Gallese and Metzinger, 2003), such as the prediction and integration of intero- and exteroceptive signals (Hohwy, 2010; Seth et al., 2011; Apps and Tsakiris, 2013), the importance of action and agency (Gallagher, 2000; Hohwy, 2007; Friston, 2012a), and the mineness of experience (Hohwy, 2007, 2010). In agreement with the Good Regulator theorem (Conant and Ashby, 1970; Edelman, 2008; Friston et al., 2012), which states that every good regulator of a system will ultimately become a model of that system, both the FEP and the philosophical account of minimal selfhood agree that the agent is the current embodied model of the world (Metzinger, 2004a; Hohwy, 2007; Friston, 2011).

THE PERSPECTIVITY OF THE SELF-MODEL

In accordance with the FEP, the phenomenal self-model (PSM) theory views selves as processes, not objects. Accordingly, the self is perceived *because* systems with a PSM constantly assume, or model, their own existence as a coherent entity (Metzinger, 2004a; Blanke and Metzinger, 2009). However, to assume that there is a perceiver is a fallacy (“no such things as selves exist in the world”, Metzinger, 2005). Rather, a conscious self is a result of the system’s identification with its self-model (“you *are* the content of your PSM”, Metzinger, 2005).

This self-identification is possible because the “attentional unavailability of earlier processing stages in the brain for intro-

spection" (Metzinger, 2003, 2005) leads to a gradually increasing *transparency* of higher-level phenomenal states. Transparency thus describes the fact that only the contents of phenomenal states, not their underlying mechanisms, are introspectively accessible to the subject of experience (Metzinger, 2003, 2004a). Interestingly, it has been proposed that the cognitive impenetrability of predictive coding mechanisms can be explained by the fact that hierarchically higher levels predict on longer timescales, and more abstractly than lower levels (Hohwy, 2007, 2010; Kiebel et al., 2008). Failures in these mechanisms may result in severe symptoms that seem to be related to a loss of global experiential selfhood, as demonstrated by certain disorders of "presence" such as depersonalization disorder (Seth et al., 2011). These phenomena might also be described by a loss of transparency ("if ... the self-model of a conscious system would become fully opaque, then the phenomenal target property of experiential "selfhood" would disappear", Metzinger, 2004b).

Thus, the crucial implication of transparency is that the PSM "cannot be recognized as a model by the system using it" (Metzinger, 2004a), which greatly reduces computational load within the system by efficiently avoiding an infinite regression that would otherwise arise from the logical structure of self-modeling (Metzinger, 2004a, 2005): "I can never conceive of what it is like to be me, because that would require the number of recursions I can physically entertain, plus one" (Friston et al., 2012). Similarly, the FEP states that systems operating with a self-model will have an advantage because "a unified self-model is what best allows computation of the system's current state such that action can be undertaken" (Hohwy, 2010; see Friston et al., 2012, for a discussion).

Note how, by the transparent spatiotemporal centeredness of the model onto the self (Metzinger, 2003, 2004a; see also Hohwy, 2007; Friston, 2011, 2012b), the model takes on a 1PP (Vogeley and Fink, 2003). However, the centeredness of the model is *phenomenal*, and not just (but also) geometrical (a temporal centering on the subject happens through successful prediction, see previous section). This is well reflected by Blanke and Metzinger (2009), who distinguish between the phenomenally distinct *weak 1PP*, and *strong 1PP*: The weak 1PP means a purely geometric centering of the experiential space upon one's body, and thus corresponds most to the "egocentre" (Roelofs, 1959; Merker, 2007) or "cyclopean eye" (von Helmholtz, 1962), which can be traced back to Hering's (1942) projective geometry. Experimental work on extending the RHI paradigm has shown that the strength of illusory self-identification with a dummy or virtual body crucially depends on this kind of 1PP (Petkova and Ehrsson, 2008; Petkova et al., 2011b; Maselli and Slater, 2013), and that in addition to proprioceptive information, vestibular information is crucial for determining self-location in space (Schwabe and Blanke, 2008; Blanke, 2012).

As an attempt to summarize the reviewed accounts of the basic constituents of MPS, **Figure 1** shows a schematic depiction of a hierarchical generative model, predicting from the *minimal phenomenal self* to increasingly specific, unimodal lower levels on shorter timescales (Kiebel et al., 2008; Hohwy, 2010; Clark, 2013). For simplicity, we have only included one intermediate level in the hierarchy, consisting of the basic aspects of minimal selfhood as

discussed in the reviewed articles (see Figure caption for a detailed description).

In the generative self-model (**Figure 1**), the first-person perspective (1PP) node should be taken as a purely geometrical point of convergence of sensory information from a particular sensory modality (a "weak 1PP"), whereas the phenomenal centeredness of the model onto the experiencing subject would correspond to a "strong 1PP" (Blanke and Metzinger, 2009). Note that although the weak 1PP and self-location usually coincide, these two phenomena can be decoupled in neurological patients with autoscopic phenomena, while MPS still seems to be normal in these conditions (Blanke and Metzinger, 2009; Blanke, 2012). This seems to speak for a probabilistic processing of minimal selfhood, and also for a relative independence of 1PP and self-location (which are therefore also modeled as separate nodes on the intermediate level of the generative model in **Figure 1**).

In conclusion, the experienced 1PP presents itself as a key feature of "mineness", and thus as a basic constituent of, and a prerequisite for a minimal self (Gallagher, 2000; Vogeley and Fink, 2003; Metzinger, 2004a; Blanke and Metzinger, 2009). Some authors speak of a system's "ability" to take the 1PP, meaning the ability to integrate and represent experience, i.e., mental states, in a common egocentric reference frame centered upon the body (Vogeley and Fink, 2003). The FEP very comfortably complies with the assumption that a body model "defines a volume within a spatial frame of reference ... within which the origin of the weak 1PP is localized" (Blanke and Metzinger, 2009; Friston, 2011, 2012b). In this light, we now review the explanatory power of the FEP for mechanisms of modeling other agents.

MODELING OTHERS

In opposition to the 1PP, the third-person perspective (3PP) is the perspective of the observer, i.e., the perspective that is taken when states are ascribed to someone else (Vogeley and Fink, 2003; Blanke and Metzinger, 2009; Fuchs, 2012). This form of perspective taking is of essential importance, for how we make sense of ourselves in a social environment depends on the representation of, and distinction between, actions and states of the self and those of others (Decety and Sommerville, 2003; Frith, 2007; Bernhardt and Singer, 2012; Farmer and Tsakiris, 2012; Frith and Frith, 2012). Traditionally, at least two distinct mechanisms have been postulated to underlie our understanding of other's internal states: *experience sharing* and *mentalizing* (Brown and Brüne, 2012; Zaki and Ochsner, 2012). While experience sharing refers to a mere mirroring of others' action intentions, sensations, or emotions (Gallese and Sinigaglia, 2011), the term mentalizing describes explicitly reflecting others' internal states: in a recent review, Zaki and Ochsner (2012) define the mechanism behind mentalizing as "the ability to represent states outside of a perceiver's 'here and now'", thus having both a spatial 1PP and a temporal (present versus past and future) aspect. Crucially, this involves a representation of other agents as possessing a 1PP that differs from one's own (Farmer and Tsakiris, 2012). One can also describe these processes as simulating other PSMs (Metzinger, 2004a); in this way, a pre-reflective, phenomenally transparent self-model is necessary for the formation of higher-level cognitive

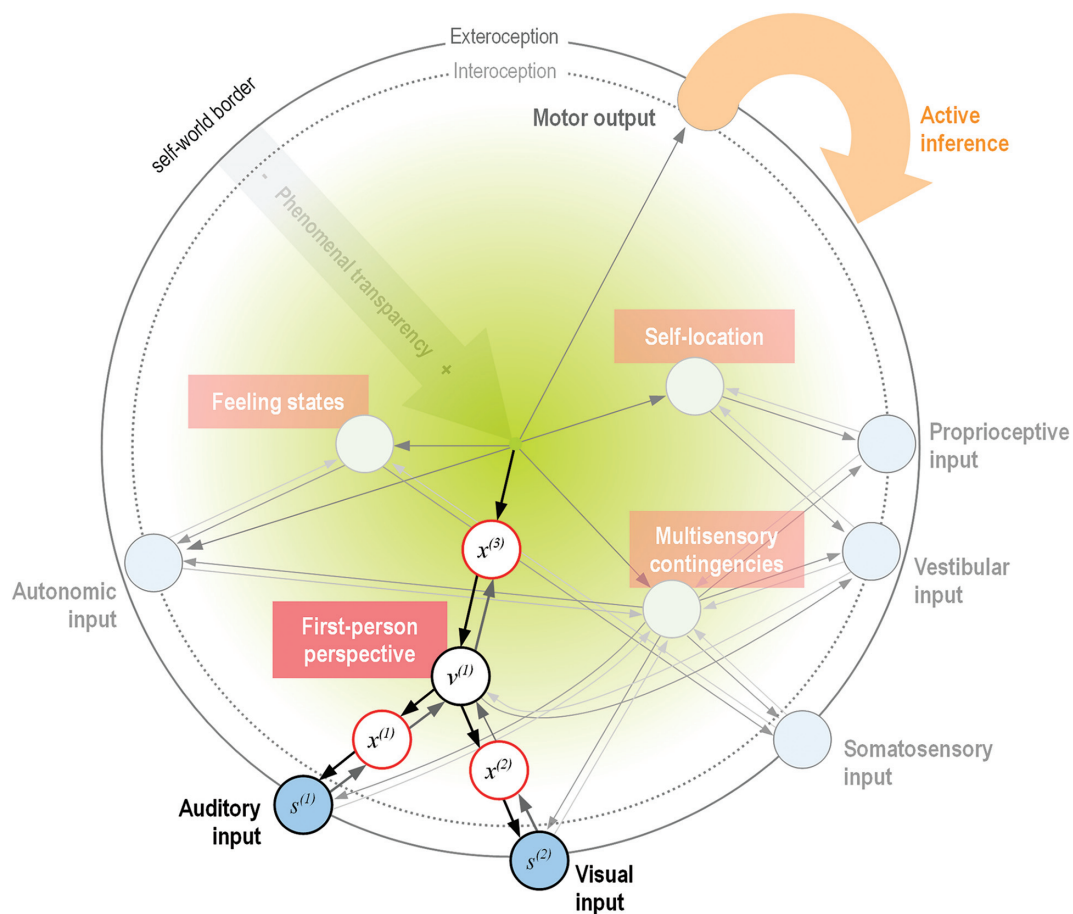


FIGURE 1 | Schematic proposal for a mapping of the *phenomenal self-model* onto a *hierarchical generative model* (format adapted from Bastos et al., 2012). Shown here is only the system's model of itself, where representational nodes at each level generate descending predictions to increasingly specialized lower levels (symbolized by darker arrows). In this special case, the single modeled cause of sensations is the *minimal phenomenal self* (Metzinger, 2004a), which generates predictions about the state of one or many sensory modalities (blue circles). The inversion of this generative model (a predictive coding scheme, lighter arrows) infers hidden causes—and thus ultimately, the self as the single cause—of sensory input via minimization of prediction error (Friston, 2011). For simplicity, only one intermediate level of nodes within the hierarchy is displayed, consisting of the basic properties of minimal selfhood as reviewed (white circles). As a (simplified) illustration of the hierarchical generative processing, the case of the 1PP is highlighted. Here, descending predictions of the unified self-model (black arrows) generate sensory data $s^{(i)}$ in the respective modalities (auditory and visual). This happens via a hierarchy of hidden states $x^{(i)}$ and hidden

causes $v^{(i)}$ (the 1PP), which generate predictions about data in the level below. The green gradient symbolizes increasing transparency of the accompanying phenomenal states with ascending hierarchy, where the final cause (the self) is completely transparent. Note that at this (highest) level, there is no further representational node; this acknowledges the fact that the perception of a unified minimal self is the result of a temporally extended predictive process, not a static representation (Metzinger, 2004a; Hohwy, 2007). The experience of “mineness” of the self (and of perception and action in general, Hohwy, 2007) is a result of the model's successful predictions and thus implicitly symbolized by the arrows. Input into this system-model comes from intero- and exteroception (blue circles), while active inference is a means of changing predicted input in all modalities through interaction with the environment. As the model-evidence is evidence for the agent's existence (Friston, 2011, 2013b), the model will necessarily be a veridical model of the agent: if there was too much unexplained prediction error, the model would be abandoned in favor of a model with a higher evidence; the self in the present form would cease to exist (Hohwy, 2010; Friston, 2011, 2012b).

and social mental concepts (Metzinger, 2003, 2004a, 2005; Edelman, 2008; Blanke and Metzinger, 2009).

Humans display first instances of experience sharing almost from birth onwards (Tomasello et al., 2005), for example, human infants as young as one hour after birth can already imitate facial gestures (Meltzoff and Moore, 1983). It hence seems that an “experiential connection” between self and others is already present in newborn infants (Gallagher and Meltzoff, 1996; Fuchs, 2012). Another example for such a pre-reflective self-other con-

nection is sensorimotor mirroring (“neural resonance”, Zaki and Ochsner, 2012). Many studies have reported vicarious activations of the motor system by observing others' actions (Rizzolatti and Craighero, 2004), or likewise of the somatosensory system by the observation of touch (Keysers et al., 2010) or pain to others (Bernhardt and Singer, 2012). These findings suggest a very basic, automatic activation of one's representations to another person's action intentions, or experience (Keysers et al., 2010; Zaki and Ochsner, 2012). There have been arguments for a link between

sensory mirroring mechanisms and higher-level perspective taking abilities (see Preston and de Waal, 2002, for a discussion), suggesting that although such vicarious responses are activated automatically, they are not purely sensory-driven (Singer and Lamm, 2009).

The FEP emphasizes models of the behavior and intentions of others as a crucial determinant of our own behavior (Frith, 2007; Friston, 2012a). It has accordingly been proposed that mechanisms of social cognition are based on predictive coding as well (Baker et al., 2011; Brown and Brüne, 2012; Frith and Frith, 2012), where perspective taking can be described as forming “second order representations” (Friston, 2013b). In other words, as agents, we also have to predict the behavior of other agents, by not only generating a model of the physical world (and our body) but also of the mental world-models of our conspecifics based on their behavior (Frith, 2007; Frith and Frith, 2012). Crucially, we have to continually update our models of others’ mental states via prediction errors, because these states are not stable but vary over time (Frith and Frith, 2012). This task is far from trivial, and involves many levels of differential self-other modeling ranging from a purely spatial differentiation (other agents occupy different positions in the world) to the abstract modeling of other minds like in Theory of Mind (Vogele and Fink, 2003; Baker et al., 2011).

Several recent accounts have proposed that associative learning updated through prediction errors is a common computational mechanism underlying both reward learning and social learning (Behrens et al., 2008; Hampton et al., 2008; Frith and Frith, 2012). Experimental evidence from these studies suggests that prediction errors code for false predictions about others’ mental states (Behrens et al., 2008; Hampton et al., 2008), and even for discrepancies between predictions of others and actual outcome of their choice (Apps et al., 2013). Interestingly, it seems that even low-level predictions can also be updated interactively. For example, dyads of individuals with similar perceptual sensitivity may benefit from interactive decision-making, as shown by an increased performance in a collective perceptual decision task during which levels of confidence were communicated (Bahrami et al., 2010). As mentioned before, if these basic predictive mechanisms fail, pathological behavior can emerge (Fletcher and Frith, 2008; Brown et al., 2013). For example, perspective taking abilities seem to be often impaired in individuals suffering from Autism Spectrum Disorder (ASD; Oberman and Ramachandran, 2007; but cf. Hamilton et al., 2007), while there is also evidence for impaired predictive coding mechanisms in ASD (Friston, 2012a).

An intriguing question is whether the brain uses the same models to generate predictions about own and other behavior. In a predictive coding account of action understanding, Kilner and colleagues (Kilner et al., 2007; Friston et al., 2011) have argued that the *mirror neuron system* is part of a generative model predicting the sensory consequences of actions, and that indeed, it seems that the brain applies the same model to predict one’s own, and others’ actions. Actions are thereby modeled on four hierarchical levels (Hamilton and Grafton, 2008): intentions, goals, kinematics, and muscles. By inversion of the model, the brain can thus infer the causes of own and others’ actions, via explaining away prediction error across these four levels. Thus the mirror neu-

ron system is active during action observation because the “own” generative model is inverted to infer the intention underlying the observed action. A similar argument is made by Gallese and Sinigaglia (2011) (see also Goldman and de Vignemont, 2009) to explain embodied simulation in general by the fact that representations of states of the self and others’ states have the same bodily format, and thus the same constraints. Correspondingly, there is evidence that the same neuronal structures may be involved in predicting own and others’ internal states (Bernhardt and Singer, 2012), for example, in predicting how pain will feel for others (Singer et al., 2004). In sum, there is strong evidence that others’ mental states are inferred via internal models. It seems that the use of generative models by the brain can explain many of these basic, as well as more elaborated social mechanisms. Thereby (at least partially) common predictive mechanisms for self and others strongly support the notion of perspective taking as an “embodied cognitive process” (Kessler and Thomson, 2010). This is a relatively young, but promising field of research; it is up to future studies to evaluate the explanatory power of the FEP in this domain.

CONCLUSION

In this review, we have summarized proposals from different authors, all emphasizing the concept of hierarchical generative models to explain processes underlying the bodily foundations of MPS, including its fundamental constituents such as multisensory integration, the sense of agency, the experience of mineness, perspectivity, and its phenomenal transparency. We have reviewed these free energy accounts of key aspects of minimal selfhood in the light of the premise that the self is the result of a generative process of self-modeling (Metzinger, 2004a; Hohwy, 2007). The approaches reviewed here show that the FEP complies with the claim that minimal selfhood emerges from physiological processes (Gallagher, 1986, 2000; Zahavi, 1999; Legrand, 2006; Blanke and Metzinger, 2009), and acknowledges both the phenomenal and spatiotemporal centeredness of the generative self-model as a key for minimal self-awareness. Albeit still schematic, these accounts demonstrate that the predictive coding account can inform theoretical and experimental approaches towards the normal and pathological self. The FEP is increasingly gaining influence as a “deeply unified account of perception, cognition, and action” (Friston, 2010; Hohwy, 2010; Apps and Tsakiris, 2013; Clark, 2013), up to recent accounts proposing it as a general mechanism underlying evolution and the “emergence of life” itself (Friston, 2013c). A particular strength of the approach seems to be that it makes relatively few conceptual assumptions (Hohwy, 2007, 2010; Friston, 2008; Friston and Kiebel, 2009; Friston et al., 2012), thus being capable of formalizing both spatial and social aspects of self-models. Of course, there are many outstanding issues, and the free energy formulation will have to withstand thorough empirical testing (for discussions, Friston et al., 2012; Apps and Tsakiris, 2013; see Clark, 2013). While it is well-established in the domains of action and perception, future work will have to show whether the FEP can be similarly influential in cognitive and social domains. Particularly, the social domain lacks models (Frith and Frith, 2012), and currently the FEP seems one of the most promising candidate theories to formally describing the mechanisms

underlying the experience of being a “self in relation to others” (Frith, 2007; Friston, 2012a). The FEP may thus provide a framework to address philosophical debates about self-modeling (Gal-

lagher, 2005b; cf. Metzinger, 2006), and perhaps help to bridge gaps between neuroscientific and philosophical approaches to the self.

REFERENCES

- Anderson, M. L., and Perlis, D. R. (2005). The roots of self-awareness. *Phenomenol. Cogn. Sci.* 4, 297–333. doi: 10.1007/s11097-005-4068-0
- Apps, M. A., and Tsakiris, M. (2013). The free-energy self: a predictive coding account of self-recognition. *Neurosci. Biobehav. Rev.* doi: 10.1016/j.neubiorev.2013.01.029. [Epub ahead of print].
- Apps, M. A., Green, R., and Ramnani, N. (2013). Reinforcement learning signals in the anterior cingulate cortex code for others' false beliefs. *Neuroimage* 64, 1–9. doi: 10.1016/j.neuroimage.2012.09.010
- Armell, K. C., and Ramachandran, V. S. (2003). Projecting sensations to external objects: evidence from skin conductance response. *Proc. Biol. Sci.* 270, 1499–1506. doi: 10.1098/rspb.2003.2364
- Aspell, J. E., Lenggenhager, B., and Blanke, O. (2009). Keeping in touch with one's self: multisensory mechanisms of self-consciousness. *PLoS One* 4:e6488. doi: 10.1371/journal.pone.0006488
- Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., and Frith, C. D. (2010). Optimally interacting minds. *Science* 329, 1081–1085. doi: 10.1126/science.1185718
- Baker, C. L., Saxe, R. R., and Tenenbaum, J. B. (2011). “Bayesian theory of mind: modeling joint belief-desire attribution,” in *Proceedings of the Thirty-Second Annual Conference of the Cognitive Science Society*, Boston, 2469–2474.
- Bastos, A. M., Urey, W. M., Adams, R. A., Mangun, G. R., Fries, P., and Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron* 76, 695–711. doi: 10.1016/j.neuron.2012.10.038
- Bechara, A., Damasio, H., and Damasio, A. R. (2000). Emotion, decision making and the orbitofrontal cortex. *Cereb. Cortex* 10, 295–307. doi: 10.1093/cercor/10.3.295
- Behrens, T. E., Hunt, L. T., Woolrich, M. W., and Rushworth, M. F. (2008). Associative learning of social value. *Nature* 456, 245–249. doi: 10.1038/nature07538
- Bermúdez, J. L., Marcel, A. J., and Eilan, N. (Eds.). (1998). *The Body and the Self*. Cambridge: MIT Press.
- Bernhardt, B. C., and Singer, T. (2012). The neural basis of empathy. *Annu. Rev. Neurosci.* 35, 1–23. doi: 10.1146/annurev-neuro-062111-150536
- Blakemore, S. J., Wolpert, D. M., and Frith, C. D. (2002). Abnormalities in the awareness of action. *Trends Cogn. Sci.* 6, 237–242. doi: 10.1016/S1364-6613(02)01907-1
- Blanke, O. (2012). Multisensory brain mechanisms of bodily self-consciousness. *Nat. Rev. Neurosci.* 13, 556–571. doi: 10.1038/nrn3292
- Blanke, O., and Metzinger, T. (2009). Full-body illusions and minimal phenomenal selfhood. *Trends Cogn. Sci.* 13, 7–13. doi: 10.1016/j.tics.2008.10.003
- Botvinick, M. (2004). Probing the neural basis of body ownership. *Science* 305, 782–783. doi: 10.1126/science.1101836
- Botvinick, M., and Cohen, J. (1998). Rubber hands ‘feel’ touch that eyes see. *Nature* 391, 756. doi: 10.1038/35784
- Brown, E. C., and Brüne, M. (2012). The role of prediction in social neuroscience. *Front. Hum. Neurosci.* 6:147. doi: 10.3389/fnhum.2012.00147
- Brown, H., Adams, R. A., Parees, I., Edwards, M., and Friston, K. (2013). Active inference, sensory attenuation and illusions. *Cogn. Process.* doi: 10.1007/s10339-013-0571-3. [Epub ahead of print].
- Brown, H., and Friston, K. J. (2012). Free-energy and illusions: the corn-sweet effect. *Front. Psychol.* 3:43. doi: 10.3389/fpsyg.2012.00043
- Bubic, A., Von Cramon, D. Y., and Schubotz, R. I. (2010). Prediction, cognition and the brain. *Front. Hum. Neurosci.* 4:25. doi: 10.3389/fnhum.2010.00025
- Clark, A. (1999). An embodied cognitive science? *Trends Cogn. Sci.* 3, 345–351. doi: 10.1016/s1364-6613(99)01361-3
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204. doi: 10.1017/s0140525x12000477
- Conant, R. C., and Ashby, W. R. (1970). Every good regulator of a system must be a model of that system. *Int. J. Syst. Sci.* 1, 89–97. doi: 10.1080/00207727008920220
- Costantini, M., and Haggard, P. (2007). The rubber hand illusion: sensitivity and reference frame for body ownership. *Conscious. Cogn.* 16, 229–240. doi: 10.1016/j.concog.2007.01.001
- Critchley, H., and Seth, A. (2012). Will studies of macaque insula reveal the neural mechanisms of self-awareness? *Neuron* 74, 423–426. doi: 10.1016/j.neuron.2012.04.012
- Critchley, H. D., Wiens, S., Rotshstein, P., Ohman, A., and Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nat. Neurosci.* 7, 189–195. doi: 10.1038/nn1176
- Damasio, A. (1999). *The Feeling of What Happens: Body, Emotion and the Making of Consciousness*. London: Vintage Books.
- de Preester, H., and Tsakiris, M. (2009). Body-extension versus body-incorporation: is there a need for a body-model? *Phenomenol. Cogn. Sci.* 8, 307–319. doi: 10.1007/s11097-009-9121-y
- de Vignemont, F. (2007). How many representations of the body? *Behav. Brain Sci.* 30, 204–205. doi: 10.1017/s0140525x07001434
- de Vignemont, F., Tsakiris, M., and Haggard, P. (2005). “Body mereology,” in *Human Body Perception from the Inside Out*, eds G. Knoblich, I. M. Thornton, M. Grosjean and M. Shiffrar (New York: Oxford University Press), 147–170.
- Decety, J., and Sommerville, J. A. (2003). Shared representations between self and other: a social cognitive neuroscience view. *Trends Cogn. Sci.* 7, 527–533. doi: 10.1016/j.tics.2003.10.004
- Edelman, S. (2008). *Computing the Mind*. New York: Oxford University Press.
- Edwards, M. J., Adams, R. A., Brown, H., Pareés, I., and Friston, K. J. (2012). A Bayesian account of ‘hysteria’. *Brain* 135, 3495–3512. doi: 10.1093/brain/aww129
- Ehrsson, H. H., Spence, C., and Passingham, R. E. (2004). That's my hand! Activity in premotor cortex reflects feeling of ownership of a limb. *Science* 305, 875–877. doi: 10.1126/science.1097011
- Ehrsson, H. H. (2007). The experimental induction of out-of-body experiences. *Science* 317, 1048. doi: 10.1126/science.1142175
- Ehrsson, H. H., Holmes, N. P., and Passingham, R. E. (2005). Touching a rubber hand: feeling of body ownership is associated with activity in multisensory brain areas. *J. Neurosci.* 25, 10564–10573. doi: 10.1523/jneurosci.0800-05.2005
- Ehrsson, H. H., Wiech, K., Weiskopf, N., Dolan, R. J., and Passingham, R. E. (2007). Threatening a rubber hand that you feel is yours elicits a cortical anxiety response. *Proc. Natl. Acad. Sci. U S A* 104, 9828–9833. doi: 10.1073/pnas.0610011104
- Farmer, H., and Tsakiris, M. (2012). The bodily social self: a link between phenomenal and narrative selfhood. *Rev. Philos. Psychol.* 3, 125–144. doi: 10.1007/s13164-012-0092-5
- Fletcher, P. C., and Frith, C. D. (2008). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nat. Rev. Neurosci.* 10, 48–58. doi: 10.1038/nrn2536
- Fotopoulou, A. (2012). “Towards psychodynamic neuroscience,” in *From the Couch to the Lab: Trends in Psychodynamic Neuroscience*, eds A. Fotopoulou, M. Conway and D. Pfaff (New York: Oxford University Press), 25–48.
- Friston, K. (2005a). A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 815–836. doi: 10.1098/rstb.2005.1622
- Friston, K. J. (2005b). Hallucinations and perceptual inference. *Behav. Brain Sci.* 28, 764–766. doi: 10.1017/s0140525x05290131
- Friston, K. (2008). Hierarchical models in the brain. *PLoS Comput. Biol.* 4:e1000211. doi: 10.1371/journal.pcbi.1000211
- Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends Cogn. Sci.* 13, 293–301. doi: 10.1016/j.tics.2009.04.005
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787
- Friston, K. (2011). Embodied inference: or ‘I think therefore I am, if I am what I think’. *The Implications of Embodiment: Cognition and Communication*. 89–125.
- Friston, K. (2012a). Prediction, perception and agency. *Int. J. Psychophysiol.* 83, 248–252. doi: 10.1016/j.jpsycho.2011.11.014

- Friston, K. (2012b). Embodied inference and spatial cognition. *Cogn. Process.* 13(Suppl. 1), S171–S177. doi: 10.1007/s10339-012-0519-z
- Friston, K. (2013a). Active inference and free energy. *Behav. Brain Sci.* 36, 212–213. doi: 10.1017/S0140525X12002142
- Friston, K. (2013b). The fantastic organ. *Brain* 136, 1328–1332. doi: 10.1093/brain/awt038
- Friston, K. (2013c). Life as we know it. *J. R. Soc. Interface* 10, 1–12. doi: 10.1098/rsif.2013.0475
- Friston, K. J., and Stephan, K. E. (2007). Free-energy and the brain. *Synthese* 159, 417–458. doi: 10.1007/s11229-007-9237-y
- Friston, K., and Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1211–1221. doi: 10.1098/rstb.2008.0300
- Friston, K., Kilner, J., and Harrison, L. (2006). A free energy principle for the brain. *J. Physiol. Paris* 100, 70–87. doi: 10.1016/j.jphysparis.2006.10.001
- Friston, K., Mattout, J., and Kilner, J. (2011). Action understanding and active inference. *Biol. Cybern.* 104, 137–160. doi: 10.1007/s00422-011-0424-z
- Friston, K., Thornton, C., and Clark, A. (2012). Free-energy minimization and the dark-room problem. *Front. Psychol.* 3:130. doi: 10.3389/fpsyg.2012.00130
- Frith, C. (2007). *Making Up the Mind: How the Brain Creates Our Mental World*. Oxford, UK: Blackwell.
- Frith, C. (2012). Explaining delusions of control: the comparator model 20 years on. *Conscious. Cogn.* 21, 52–54. doi: 10.1016/j.concog.2011.06.010
- Frith, C. D., and Frith, U. (2012). Mechanisms of social cognition. *Annu. Rev. Psychol.* 63, 287–313. doi: 10.1146/annurev-psych-120710-100449
- Fuchs, T. (2012). The phenomenology and development of social perspectives. *Phenomenol. Cogn. Sci.* 1–29. doi: 10.1007/s11097-012-9267-x
- Gallagher, S. (1986). Lived body and environment. *Res. Phenomenol.* 16, 139–170. doi: 10.1163/156916486x00103
- Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends Cogn. Sci.* 4, 14–21. doi: 10.1016/s1364-6613(99)01417-5
- Gallagher, S. (2005a). *How the Body Shapes the Mind*. Oxford: Clarendon Press.
- Gallagher, S. (2005b). Metzinger's matrix: living the virtual life with a real body. *Psyche* 11, 1–9.
- Gallagher, S. (2009). "Philosophical antecedents to situated cognition," in *Cambridge Handbook of Situated Cognition*, eds P. Robbins and M. Aydede (Cambridge: Cambridge University Press), 35–51.
- Gallagher, S., and Cole, J. (1995). Body schema and body image in a deaf-fermented subject. *J. Mind Behav.* 16, 369–390.
- Gallagher, S., and Meltzoff, A. N. (1996). The earliest sense of self and others: Merleau-Ponty and recent developmental studies. *Philos. Psychol.* 9, 211–233. doi: 10.1080/09515089608573181
- Gallese, V., and Metzinger, T. (2003). Motor ontology: the representational reality of goals, actions and selves. *Philos. Psychol.* 16, 365–388. doi: 10.1080/0951508032000121760
- Gallese, V., and Sinigaglia, C. (2011). What is so special about embodied simulation? *Trends Cogn. Sci.* 15, 512–519. doi: 10.1016/j.tics.2011.09.003
- Gibson, J. J. (1977). "The theory of affordances," in *Perceiving, Acting and Knowing*, eds R. Shaw, and J. Bransford (Hillsdale: Erlbaum), 67–82.
- Goldman, A., and de Vignemont, F. (2009). Is social cognition embodied? *Trends Cogn. Sci.* 13, 154–159. doi: 10.1016/j.tics.2009.01.007
- Grafton, S. T. (2009). Embodied cognition and the simulation of action to understand others. *Ann. N Y Acad. Sci.* 1156, 97–117. doi: 10.1111/j.1749-6632.2009.04425.x
- Hamilton, A. F. D. C., Brindley, R. M., and Frith, U. (2007). Imitation and action understanding in autistic spectrum disorders: how valid is the hypothesis of a deficit in the mirror neuron system? *Neuropsychologia* 45, 1859–1868. doi: 10.1016/j.neuropsychologia.2006.11.022
- Hamilton, A. F., and Grafton, S. T. (2008). "The motor hierarchy: from kinematics to goals and intentions," in *Sensorimotor Foundations of Higher Cognition: Attention and Performance xxii*, eds Y. Rosetti, M. Kawato and P. Haggard (New York: Oxford University Press), 381–408.
- Hampton, A. N., Bossaerts, P., and O'Doherty, J. P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proc. Natl. Acad. Sci. U S A* 105, 6741–6746. doi: 10.1073/pnas.0711099105
- Head, H., and Holmes, G. (1911–1912). Sensory disturbances from cerebral lesions. *Brain* 34, 102–245. doi: 10.1093/brain/34.2-3.102
- Hering, E. (1942). *Spatial Sense and Movements of the Eye* (A. Ruddle, Trans.). Baltimore, MD: American Academy of Optometry (original work published 1879).
- Hohwy, J. (2007). The sense of self in the phenomenology of agency and perception. *Psyche (Stuttg)* 13, 1–20.
- Hohwy, J. (2010). "The hypothesis testing brain: some philosophical applications," in *ASCS09: Proceedings of the 9th Conference of the Australasian Society for Cognitive Science*, Sydney, Australia, 135–144.
- Hohwy, J. (2012). Attention and conscious perception in the hypothesis testing brain. *Front. Psychol.* 3:96. doi: 10.3389/fpsyg.2012.00096
- Hohwy, J. (2013). Delusions, illusions and inference under uncertainty. *Mind Lang.* 28, 57–71. doi: 10.1111/mila.12008
- Hohwy, J., and Paton, B. (2010). Explaining away the body: experiences of supernaturally caused touch and touch on non-hand objects within the rubber hand illusion. *PLoS One* 5:e9416. doi: 10.1371/journal.pone.0009416
- Holmes, N. P., and Spence, C. (2004). The body schema and multisensory representation(s) of peripersonal space. *Cogn. Process.* 5, 94–105. doi: 10.1007/s10339-004-0013-3
- Ijsselstein, A. W., de Kort, Y. A. W., and Haans, A. (2006). Is this my hand i see before me? The rubber hand illusion in reality, virtual reality, and mixed reality. *Presence: Teleop. Virt.* 15, 455–464. doi: 10.1162/pres.15.4.455
- James, W. (1890). *The Principles of Psychology*. New York: Dover.
- Kessler, K., and Thomson, L. A. (2010). The embodied nature of spatial perspective taking: embodied transformation versus sensorimotor interference. *Cognition* 114, 72–88. doi: 10.1016/j.cognition.2009.08.015
- Keyser, C., Kaas, J. H., and Gazzola, V. (2010). Somatosensation in social perception. *Nat. Rev. Neurosci.* 11, 417–428. doi: 10.1038/nrn2833
- Kiebel, S. J., Daunizeau, J., and Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS Comput. Biol.* 4:e1000209. doi: 10.1371/journal.pcbi.1000209
- Kilner, J. M., Friston, K. J., and Frith, C. D. (2007). Predictive coding: an account of the mirror neuron system. *Cogn. Process.* 8, 159–166. doi: 10.1007/s10339-007-0170-2
- Knoblich, G., Elsner, B., Aschersleben, G., and Metzinger, T. (2003). Grounding the self in action. *Conscious. Cogn.* 12, 487–494. doi: 10.1016/j.concog.2003.08.009
- Legrand, D. (2006). The bodily self: the sensori-motor roots of pre-reflective self-consciousness. *Phenomenol. Cogn. Sci.* 5, 89–118. doi: 10.1007/s11097-005-9015-6
- Lenggenhager, B., Tadi, T., Metzinger, T., and Blanke, O. (2007). Video ergo sum: manipulating bodily self-consciousness. *Science* 24, 1096–1099. doi: 10.1126/science.1143439
- Lloyd, D. M. (2007). Spatial limits on referred touch to an alien limb may reflect boundaries of visuo-tactile peripersonal space surrounding the hand. *Brain Cogn.* 64, 104–109. doi: 10.1016/j.bandc.2006.09.013
- Longo, M. R., Schüür, F., Kammers, M. P. M., Tsakiris, M., and Haggard, P. (2008). What is embodiment? A psychometric approach. *Cognition* 107, 978–998. doi: 10.1016/j.cognition.2007.12.004
- Longo, M. R., Schüür, F., Kammers, M. P. M., Tsakiris, M., and Haggard, P. (2009). Self awareness and the body image. *Acta Psychol. (Amst)* 132, 166–172. doi: 10.1016/j.actpsy.2009.02.003
- Makin, T. R., Holmes, N. P., and Ehrsson, H. H. (2008). On the other hand: dummy hands and peripersonal space. *Behav. Brain Res.* 191, 1–10. doi: 10.1016/j.bbr.2008.02.041
- Maselli, A., and Slater, M. (2013). The building blocks of the full body ownership illusion. *Front. Hum. Neurosci.* 7:83. doi: 10.3389/fnhum.2013.00083
- Meltzoff, A. N., and Moore, M. K. (1983). Newborn infants imitate adult facial gestures. *Child Dev.* 54, 702–709. doi: 10.2307/1130058
- Merker, B. (2007). Consciousness without a cerebral cortex: a challenge for neuroscience and medicine. *Behav. Brain Sci.* 30, 63–81. doi: 10.1017/s0140525x07000891
- Merleau-Ponty, M. (1962). *Phenomenology of Perception* (C. Smith, Trans.). London: Routledge and Kegan Paul (original work published 1945).
- Metzinger, T. (2003). Phenomenal transparency and cognitive self-reference. *Phenomenol. Cogn. Sci.* 2, 353–393. doi: 10.1023/b:phen.0000007366.42918.eb
- Metzinger, T. (2004a). *Being No One: The Self-Model Theory of Subjectivity*. Cambridge: MIT Press.
- Metzinger, T. (2004b). The subjectivity of subjective experience: a representationalist analysis of the

- first-person perspective. *Networks* 3–4, 33–64.
- Metzinger, T. (2005). Précis: being no one. *Psyche* 11, 1–35.
- Metzinger, T. (2006). Reply to Gallagher: different conceptions of embodiment. *Psyche* 12, 4.
- Moseley, G. L., Olthof, N., Venema, A., Don, S., Wijers, M., Gallace, A., et al. (2008). Psychologically induced cooling of a specific body part caused by the illusory ownership of an artificial counterpart. *Proc. Natl. Acad. Sci. U S A* 105, 13169–13173. doi: 10.1073/pnas.0803768105
- Mumford, D. (1992). On the computational architecture of the neocortex. *Biol. Cybern.* 66, 241–251. doi: 10.1007/BF00202389
- Oberman, L. M., and Ramachandran, V. S. (2007). The simulating social mind: the role of the mirror neuron system and simulation in the social and communicative deficits of autism spectrum disorders. *Psychol. Bull.* 133, 310. doi: 10.1037/0033-2909.133.2.310
- Pennebaker, J. W., and Skelton, J. A. (1981). Selective monitoring of bodily sensations. *J. Pers. Soc. Psychol.* 41, 213–223.
- Petkova, V. I., and Ehrsson, H. H. (2008). If I were you: perceptual illusion of body swapping. *PLoS One* 3:e3832. doi: 10.1371/journal.pone.0003832
- Petkova, V. I., Björnsdóttir, M., Gentile, G., Jonsson, T., Li, T. Q., and Ehrsson, H. H. (2011a). From part-to whole-body ownership in the multisensory brain. *Curr. Biol.* 21, 1118–1122. doi: 10.1016/j.cub.2011.05.022
- Petkova, V. I., Khoshnevis, M., and Ehrsson, H. H. (2011b). The perspective matters! Multisensory integration in ego-centric reference frames determines full-body ownership. *Front. Psychol.* 2:35. doi: 10.3389/fpsyg.2011.00035
- Preston, S. D., and de Waal, F. (2002). Empathy: its ultimate and proximate bases. *Behav. Brain Sci.* 25, 1–20. doi: 10.1017/s0140525x02000018
- Rao, R. P., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87. doi: 10.1038/4580
- Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192. doi: 10.1146/annurev.neuro.27.070203.144230
- Roelofs, C. O. (1959). Considerations on the visual egocentre. *Acta Psychol. (Amst)* 16, 226–234. doi: 10.1016/0001-6918(59)90096-4
- Schachter, S., and Singer, J. E. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychol. Rev.* 69, 379–399. doi: 10.1037/h0046234
- Schwabe, L., and Blanke, O. (2008). The vestibular component in out-of-body experiences: a computational approach. *Front. Hum. Neurosci.* 2:17. doi: 10.3389/neuro.09.017.2008
- Seth, A. (2009). Explanatory correlates of consciousness: theoretical and computational challenges. *Cognit. Comput.* 1, 50–63. doi: 10.1007/s12559-009-9007-x
- Seth, A. K., Suzuki, K., and Critchley, H. D. (2011). An interoceptive predictive coding model of conscious presence. *Front. Psychol.* 2:395. doi: 10.3389/fpsyg.2011.00395
- Singer, T., and Lamm, C. (2009). The social neuroscience of empathy. *Ann. N Y Acad. Sci.* 1156, 81–96. doi: 10.1111/j.1749-6632.2009.04418.x
- Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., and Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science* 303, 1157–1162. doi: 10.1126/science.1093535
- Slater, M., Perez-Mercos, D., Ehrsson, H. H., and Sanchez-Vives, M. V. (2009). Inducing illusory ownership of a virtual body. *Front. Neurosci.* 3:29. doi: 10.3389/neuro.01.029.2009
- Srinivasan, M. V., Laughlin, S. B., and Dubs, A. (1982). Predictive coding: a fresh view of inhibition in the retina. *Proc. R. Soc. Lond. B Biol. Sci.* 216, 427–459. doi: 10.1098/rspb.1982.0085
- Suzuki, K., Garfinkel, S., Critchley, H., and Seth, A. K. (2013). Multisensory integration across interoceptive and exteroceptive domains modulates self-experience in the rubber-hand illusion. *Neuropsychologia* doi: 10.1016/j.neuropsychologia.2013.08.014. (in press).
- Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* 28, 675–690. doi: 10.1017/s0140525x05000129
- Tsakiris, M. (2010). My body in the brain: a neurocognitive model of body-ownership. *Neuropsychologia* 48, 703–712. doi: 10.1016/j.neuropsychologia.2009.09.034
- Tsakiris, M., and Haggard, P. (2005). The rubber hand illusion revisited: visuotactile integration and self-attribution. *J. Exp. Psychol. Hum. Percept. Perform.* 31, 80–91. doi: 10.1037/0096-1523.31.1.80
- Tsakiris, M., Hesse, M. D., Boy, C., Haggard, P., and Fink, G. R. (2007). Neural signatures of body ownership: a sensory network for bodily self-consciousness. *Cereb. Cortex* 17, 2235–2244. doi: 10.1093/cercor/bhl131
- Varela, F., Thompson, E., and Rosch, E. (1994). *The Embodied Mind*. Cambridge, MA: MIT Press.
- Verschure, P. F., Voegtlin, T., and Douglas, R. J. (2003). Environmentally mediated synergy between perception and behaviour in mobile robots. *Nature* 425, 620–624. doi: 10.1038/nature02024
- Vogeley, K., and Fink, G. R. (2003). Neural correlates of the first-person-perspective. *Trends Cogn. Sci.* 7, 38–42. doi: 10.1016/s1364-6613(02)00003-7
- von Helmholtz, H. (1962). *Helmholtz's Treatise on Physiological Optics* (J. P. C. Southall, Trans. and Ed.). New York: Dover Publication (Original work published 1866).
- Zahavi, D. (1999). *Self-Awareness and Alterity: A Phenomenological Investigation*. Evanston: Northwestern Univ Press.
- Zaki, J., and Ochsner, K. N. (2012). The neuroscience of empathy: progress, pitfalls and promise. *Nat. Neurosci.* 15, 675–680. doi: 10.1038/nn.3085

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 16 July 2013; accepted: 20 August 2013; published online: 12 September 2013.

Citation: Limanowski J and Blankenburg F (2013) Minimal self-models and the free energy principle. *Front. Hum. Neurosci.* 7:547. doi: 10.3389/fnhum.2013.00547

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2013 Limanowski and Blankenburg. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

ADVANTAGES OF PUBLISHING IN FRONTIERS



FAST PUBLICATION

Average 90 days
from submission
to publication



COLLABORATIVE PEER-REVIEW

Designed to be rigorous –
yet also collaborative, fair and
constructive



RESEARCH NETWORK

Our network
increases readership
for your article



OPEN ACCESS

Articles are free to read,
for greatest visibility



TRANSPARENT

Editors and reviewers
acknowledged by name
on published articles



GLOBAL SPREAD

Six million monthly
page views worldwide



COPYRIGHT TO AUTHORS

No limit to
article distribution
and re-use



IMPACT METRICS

Advanced metrics
track your
article's impact



SUPPORT

By our Swiss-based
editorial team