

PERCEPTION-COGNITION INTERFACE AND CROSS-MODAL EXPERIENCES: INSIGHTS INTO UNIFIED CONSCIOUSNESS

EDITED BY : Aleksandra Mroczko-Wąsowicz
PUBLISHED IN : Frontiers in Psychology



frontiers

Frontiers Copyright Statement

© Copyright 2007-2017 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 978-2-88945-071-8

DOI 10.3389/978-2-88945-071-8

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

PERCEPTION-COGNITION INTERFACE AND CROSS-MODAL EXPERIENCES: INSIGHTS INTO UNIFIED CONSCIOUSNESS

Topic Editor:

Aleksandra Mroczko-Wąsowicz, National Yang Ming University, Taiwan



Cover image: “Land - Hora et Labora” by Matteo Boato, used with permission.

Available as Supplementary Material to: Albertazzi L, Canal L and Micciolo R (2015) Cross-modal associations between materic painting and classical Spanish music. *Front. Psychol.* 6:424.

doi: 10.3389/fpsyg.2015.00424

The present Research Topic explores closely related aspects of mental functioning, namely an interplay between perception and cognition, interactions among various sensory modalities, and finally, more or less unified conscious experiences arising in the context of these relations. Contributions emphasize a high flexibility observed in perception and may be seen as potential challenges to the traditional modular architecture of perceptual systems. Although the articles describe different phenomena, they follow one common theme — to investigate broadly understood unified experience — by studying either perception-cognition integration or the integration between sensory modalities. These integrative processes may well apply to subpersonal unconscious representations. However, the aim here is to approach phenomenal experience and thus a straightforward way of thinking about it is in terms of conscious perception.

Putting together scientific and philosophical concerns, this special issue encourages extending the study of perceptual experience beyond the single sense perception to advance our understanding of the complex interdependencies between different sensory modalities, other mental domains, and various kinds of unifying relations within conscious experience. It exhibits a remarkable need to study these phenomena in tangent, and so, the authors examine a variety of ways in which our perceptual experiences may be cross-modal or multisensory, integrated, embodied, synesthetic, cognitively penetrated, or otherwise affected by top-down influences.

The Research Topic comprises theoretical and empirical contributions of such fields as philosophy of mind, cognitive science, psychology, and neuroscience in the form of hypothesis and theory articles, original research articles, opinion papers, reviews, and commentaries.

Citation: Mroczko-Wąsowicz, A., ed. (2017). Perception-Cognition Interface and Cross-Modal Experiences: Insights into Unified Consciousness. Lausanne: Frontiers Media.
doi: 10.3389/978-2-88945-071-8

Table of Contents

- 05 Editorial: Perception–Cognition Interface and Cross-Modal Experiences: Insights into Unified Consciousness**
Aleksandra Mroczko-Wąsowicz
- 09 Rethinking the senses and their interactions: the case for sensory pluralism**
Matthew Fulkerson
- 23 Multisensory perception as an associative learning process**
Kevin Connolly
- 33 Body ownership and experiential ownership in the self-touching illusion**
Caleb Liang, Si-Yan Chang, Wen-Yeo Chen, Hsu-Chia Huang and Yen-Tung Lee
- 46 The Merit of Synesthesia for Consciousness Research**
Tessa M. van Leeuwen, Wolf Singer and Danko Nikolić
- 55 Synesthesia and release phenomena in sensory and motor grounding. Cases of disinhibited embodiment?**
Brian F. Gray and Julia Simmer
- 59 Revisiting the empirical case against perceptual modularity**
Farid Masrour, Gregory Nirshberg, Michael Schon, Jason Leardi and Emily Barrett
- 76 Cognitive penetrability and emotion recognition in human facial expressions**
Francesco Marchi and Albert Newen
- 88 Cognitive penetration and the gallery of indiscernibles**
Bence Nanay
- 91 Do intentions for action penetrate visual experience?**
Robert E. Briscoe
- 93 Consciousness doesn't overflow cognition**
Richard Brown
- 96 It does belong together: cross-modal correspondences influence cross-modal integration during perceptual learning**
Lionel Brunel, Paulo F. Carvalho and Robert L. Goldstone
- 106 Cross-modal metaphorical mapping of spoken emotion words onto vertical space**
Pedro R. Montoro, María José Contreras, María Rosa Elosúa and Fernando Marmolejo-Ramos
- 119 Cross-modal associations between materic painting and classical Spanish music**
Liliana Albertazzi, Luisa Canal and Rocco Micciolo
- 132 The coherent and fluent mind: how unified consciousness is constructed from cross-modal inputs via integrated processing experiences**
Piotr Winkielman, Michał Ziembowicz and Andrzej Nowak



Editorial: Perception–Cognition Interface and Cross-Modal Experiences: Insights into Unified Consciousness

Aleksandra Mroczko-Wąsowicz *

Institute of Philosophy of Mind and Cognition, National Yang-Ming University, Taipei, Taiwan

Keywords: perception, cognition, concepts, cognitive penetrability of perception, cross-modal experience, multisensory integration, multimodal binding, the unity of consciousness

Editorial on the Research Topic

Perception–Cognition Interface and Cross-Modal Experiences: Insights into Unified Consciousness

OPEN ACCESS

Edited by:

Morten Overgaard,
Aarhus University, Denmark

Reviewed by:

Luis Lemus,
National Autonomous University of
Mexico, Mexico
Liad Mudrik,
Tel Aviv University, Israel

*Correspondence:

Aleksandra Mroczko-Wąsowicz
mroczko-wasowicz@hotmail.com

Specialty section:

This article was submitted to
Consciousness Research,
a section of the journal
Frontiers in Psychology

Received: 14 June 2016

Accepted: 30 September 2016

Published: 22 November 2016

Citation:

Mroczko-Wąsowicz A (2016) Editorial:
Perception–Cognition Interface and
Cross-Modal Experiences: Insights
into Unified Consciousness.
Front. Psychol. 7:1593.
doi: 10.3389/fpsyg.2016.01593

The present Research Topic explores closely related aspects of mental functioning, namely an interplay between perception and cognition, interactions among various sensory modalities, and finally, more or less unified conscious experiences arising in the context of these relations. Contributions emphasize a high flexibility observed in perception and may be seen as potential challenges to the traditional modular architecture of perceptual systems. Although the articles describe different phenomena, they follow one common theme—to investigate broadly understood unified experience—by studying either perception–cognition integration or the integration between sensory modalities. These integrative processes may well apply to subpersonal unconscious representations. However, the aim here is to approach phenomenal experience and thus a straightforward way of thinking about it is in terms of conscious perception.

One of the seemingly principled divisions in the human mind is between sense perception and high-level cognition. Traditionally, perception and cognition have been viewed as distinct, encapsulated domains operating independently of each other (Fodor, 1983, 1984, 2008; Pylyshyn, 1999, 2003; Barrett, 2005; Heck, 2007; Firestone and Scholl, 2014). However, recent studies support a different view about the impact of perception on cognition (Barsalou, 2009, 2012; Goldstone and Hendrickson, 2010; Prinz, 2011; Weiskopf, 2015) as well as the various ways in which perceptual experiences can be influenced by cognitive states such as thoughts, judgments, beliefs, intuitions, expectations, desires, mental images, and emotions (Brockmole et al., 2002; Raftopoulos, 2011; Lupyan, 2012; Macpherson, 2012, 2016; Siegel, 2012; Stokes, 2012; Bannert and Bartels, 2013; Vetter and Newen, 2014; Raftopoulos and Zeimbekis, 2016). Thus, although the mentioned division between perceiving and reasoning may seem conceptually clear and unambiguous, these mental domains become closely intertwined when our beliefs, expectations, or desires affect what we see, hear, or taste, leading to complex phenomenal states of a hybrid nature. No matter whether we assume that there is no dichotomy between perception and cognition (Clark, 2013; Lupyan, 2015) or instead assume that there is a principled difference and a joint in nature between perception and cognition (Block, submitted; cf. Firestone and Scholl, 2015), cognitive-sensory interactions can and need to be accommodated within any of these accounts.

Perception has typically been studied in a single sense, mostly in the visual or auditory modality (Haynes et al., 2005; Gutschalk et al., 2008; Bekinschtein et al., 2009; Dehaene and Changeux, 2011; De Graaf et al., 2012). However, cross-modal experiences and heterogeneous

multisensory interactions, in which input in one sensory modality elicits or modulates contents in another modality, reveal that such perceptual experiences cannot be easily categorized as belonging to only one of the senses (De Gelder and Bertelson, 2003; Stein, 2012). Furthermore, recent studies suggest examining the role of multisensory signals in perceptual consciousness (Chen and Spence, 2010; De Meo et al., 2015; Deroy et al., 2016). While processing sensory information in cross-modal cases is generally multisensory, the result of that processing can be interpreted as either just a sum of coexisting modality-specific representations or an intrinsically multisensory whole. Determining whether multisensory processing results in a decomposable conjunction of independent unisensory contents or in a multimodal holistic state that cannot be parceled out into modality-specific components would provide the needed characterization of the basic units of perceptual consciousness (Bayne, 2014). Still, it is important to realize that instances of successful multisensory integration and cross-modal binding facilitated by spatio-temporal or semantic congruence are not necessarily accompanied by unified experiences of objects across the senses and that the complex relationship between multisensory integration and perceptual consciousness remains to be clarified (Deroy, 2014; c.f. Spence and Bayne, 2015).

Recent years have seen a surge of novel interdisciplinary work questioning the received view of separate sensory systems and traditional conceptions of different mental domains operating independently (Shimojo and Shams, 2001; Driver and Noesselt, 2008; Bayne, 2010; Macpherson, 2011a,b; Mroczko-Wąsowicz, 2013, 2016; Bennett and Hill, 2014; Deroy et al., 2014; de Vignemont, 2014a,b; Mroczko-Wąsowicz and Nikolić, 2014; O'Callaghan, 2014; Matthen, 2015; Stokes et al., 2015). For instance, the occurrence of cross-domain interchange going beyond the link between the sense perception and the domain of abstract, conceptually represented entities, i.e., extending to the domains of bodily, motor, and emotional states provides challenges to standard methods individuating our epistemic abilities (O'Regan and Noë, 2001; Barsalou, 2008; Tajadura-Jiménez et al., 2011, 2015; Mroczko-Wąsowicz and Werning, 2012; De Coster et al., 2013; Weiss et al., 2013; Shapiro, 2014; Goldstone et al., 2015). Results of these studies point to a significant change in our understanding of perception; they demonstrate an emerging agreement on an integrative picture of perception incorporating informational interactions. All this indicates a need for a new research methodology. A full understanding of how the mind works requires considering multifaceted links holding between various mental domains and their mutual impact. Our mental faculties should not only be studied separately. They require a more holistic approach in order to uncover their extensive capacity for interactions producing differently unified conscious experiences.

Putting together scientific and philosophical concerns, this special issue encourages extending the study of perceptual experience beyond the single sense perception to advance our understanding of the complex interdependencies between different sensory modalities, other mental domains, and various kinds of unifying relations within conscious experience. It

exhibits a remarkable need and benefit to study these phenomena in tangent, and so, the articles in this Research Topic examine a variety of ways in which our perceptual experiences may be cross-modal or multisensory, integrated, embodied, synesthetic, or affected by top-down influences.

Fulkerson argues that there are many forms of sensory interaction and unity, therefore classification of sensory systems and generated experiences is a matter of particular explanatory projects.

Connolly suggests that it is not an automatic feature binding mechanism that is responsible for our multimodal perceptual experiences, but rather an associative learning process that couples features from different sensory modalities so that we experience them as part of the same event.

Liang et al. investigate experiential ownership of bodily sensation and if it is guaranteed that a subject cannot be wrong about whether it is him who feels the sensation.

Van Leeuwen et al. propose several reasons for why the phenomenon of synesthesia and related alterations of brain networks and functional connectivity can be of merit for consciousness research.

Gray and Simner consider synesthesia and release phenomena in terms of disinhibited embodiment in sensory and motor systems respectively.

The following papers explore how perceptual processes can fail to be modular. They discuss a range of questions regarding cognitive effects on perception, including the issue of cognitive penetrability of perception.

In their contribution, Masrour et al. guide readers through philosophical issues of the debate on perceptual modularity, emphasizing results from cognitive neuroscience against the encapsulation thesis.

Marchi and Newen address the possibility of cognitive penetrability of perceptual experience in the domain of social cognition, namely visual experience of the facial expressions of emotion.

Nanay argues that the attribution of aesthetically relevant properties supervenes on one's perceptual experience, i.e., if there is a difference in such an attribution, there must also be a difference in perceptual experience.

Briscoe considers whether intentions for action penetrate visual experience of an object's size by analyzing various explanations of mechanisms possibly involved in such penetration.

Brown claims that theories of consciousness that see it as cognitive in nature or as an aspect of cognitive functioning such as the higher-order thought theory of consciousness provide a reasonable working hypothesis in the explanation of conscious experience.

There is also some grouping among the contributions discussing empirical results from their own studies on cross-modal associations, sensory integration, and unified conscious experience.

Brunel et al. demonstrate that cross-modal correspondences influence cross-modal integration during perceptual learning, leading to new learned units that have different stability over time.

Montoro et al. explore cross-modal metaphorical mapping of auditory emotion words onto vertical visual space and conclude that this association is not automatically activated but requires an explicit semantic evaluation of the emotion concepts to obtain an embodied effect.

Albertazzi et al. examine the existence of cross-modal associations between highly complex stimuli (i.e., materic painting and classical guitar music) due to patterns of qualitative similarity present in stimuli of different sensory modalities.

Finally, Winkielman et al. propose that unified consciousness is constructed from cross-modal inputs via integrated processing

experiences, an experiential mechanism that combines signals of processing quality.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

ACKNOWLEDGMENTS

This work was supported by the Ministry of Science and Technology (project: MOST104-2628-H-010-002-MY3).

REFERENCES

- Bayne (2014). "The multisensory nature of perceptual consciousness," in *Sensory Integration and the Unity of Consciousness*, eds D. J. Bennett and C. S. Hill (Cambridge, MA: MIT Press), 15–36.
- Bannert, M. M., and Bartels, A. (2013). Decoding the yellow of a gray banana. *Curr. Biol.* 23, 2268–2272. doi: 10.1016/j.cub.2013.09.016
- Barrett, H. C. (2005). Enzymatic computation and cognitive modularity. *Mind Lang.* 259–287. doi: 10.1111/j.0268-1064.2005.00285.x
- 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. Biol. Sci.* 364, 1281–1289. doi: 10.1098/rstb.2008.0319
- Barsalou, L. W. (2012). "The human conceptual system," in *The Cambridge Handbook of Psycholinguistics*, eds M. Spivey, K. McRae, and M. Joanisse (New York, NY: Cambridge University Press), 239–258.
- Bayne, T. (2010). *The Unity of Consciousness*. New York, NY: Oxford University Press.
- Bekinschtein, T. A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., and Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. *Proc. Natl Acad. Sci. U.S.A.* 106, 1672–1677. doi: 10.1073/pnas.0809667106
- Bennett, D. J., and Hill, C. S. (eds.). (2014). *Sensory Integration and the Unity of Consciousness*. Cambridge, MA: MIT Press.
- Brockmole, J. R., Wang, R. F., and Irwin, D. E. (2002). Temporal integration between visual images and visual percepts. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 315–334. doi: 10.1037//0096-1523.28.2.315
- Chen, Y.-C., and Spence, C. (2010). When hearing the bark helps to identify the dog: semantically-congruent sounds modulate the identification of masked pictures. *Cognition* 114, 389–404. doi: 10.1016/j.cognition.2009.10.012
- 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
- De Coster, L., Verschuere, B., Goubert, L., Tsakiris, M., and Brass, M. (2013). I suffer more from your pain when you act like me: being imitated enhances affective responses to seeing someone else in pain. *Cogn. Affect. Behav. Neurosci.* 13, 519–532. doi: 10.3758/s13415-013-0168-4
- De Gelder, B., and Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends Cogn. Sci.* 7, 460–467. doi: 10.1016/j.tics.2003.08.014
- De Graaf, T. A., Hsieh, P. J., and Sack, A. T. (2012). The "correlates" in neural correlates of consciousness. *Neurosci. Biobehav. Rev.* 36, 191–197. doi: 10.1016/j.neubiorev.2011.05.012
- Dehaene, S., and Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron* 70, 200–227. doi: 10.1016/j.neuron.2011.03.018
- De Meo, R., Murray, M. M., Clarke, S., and Matusz, P. J. (2015). Top-down control and early multisensory processes: chicken vs. egg. *Front. Integr. Neurosci.* 9:17. doi: 10.3389/fnint.2015.00017
- Deroy, O. (2014). "The unity assumption and the many unities of consciousness," in *Sensory Integration and the Unity of Consciousness*, eds D. Bennett and C. Hill (Cambridge, MA: MIT Press), 105–124.
- Deroy, O., Chen, Y.-C., and Spence, C. (2014). Multisensory constraints on awareness. *Phil. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130207. doi: 10.1098/rstb.2013.0207
- Deroy, O., Faivre, N., Lunghi, C., Spence, C., Aller, M., and Noppeney, U. (2016). The complex interplay between multisensory integration and perceptual awareness. *Multisens. Res.* 29, 585–606. doi: 10.1163/22134808-00002529
- de Vignemont, F. (2014a). "Multimodal unity and multimodal binding," in *Sensory Integration and the Unity of Consciousness*, eds D. Bennett and C. Hill (Cambridge, MA: MIT Press), 125–150.
- de Vignemont, F. (2014b). A multimodal conception of bodily awareness. *Mind* 123, 989–1020. doi: 10.1093/mind/fzu089
- Driver, J., and Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. *Neuron* 57, 11–23. doi: 10.1016/j.neuron.2007.12.013
- Firestone, C., and Scholl, B. J. (2014). "Top-down" effects where none should be found: the El Greco fallacy in perception research. *Psychol. Sci.* 25, 38–46. doi: 10.1177/0956797613485092
- Firestone, C., and Scholl, B. J. (2015). Cognition does not affect perception: evaluating the evidence for "top-down" effects. *Behav. Brain Sci.* doi: 10.1017/S0140525X15000965. [Epub ahead of print].
- Fodor, J. A. (1983). *The Modularity of Mind: An Essay in Faculty Psychology*. Cambridge, MA: MIT Press.
- Fodor, J. A. (1984). Observation reconsidered. *Philos. Sci.* 51, 23–43.
- Fodor, J. A. (2008). *LOT 2: The Language of Thought Revisited*. Oxford: Oxford University Press.
- Goldstone, R. L., de Leeuw, J. R., and Landy, D. H. (2015). Fitting perception in and to cognition. *Cognition* 135, 24–29. doi: 10.1016/j.cognition.2014.11.027
- Goldstone, R. L., and Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdiscipl. Rev. Cogn. Sci.* 1, 69–78. doi: 10.1002/wcs.26
- Gutschalk, A., Micheyl, C., and Oxenham, A. J. (2008). Neural correlates of auditory perceptual awareness under informational masking. *PLoS Biol.* 6:138. doi: 10.1371/journal.pbio.0060138
- Haynes, J. D., Driver, J., and Rees, G. (2005). Visibility reflects dynamic changes of effective connectivity between V1 and fusiform cortex. *Neuron* 46, 811–821. doi: 10.1016/j.neuron.2005.05.012
- Heck, R. G. (2007). "Are there different kinds of content?" in *Contemporary Debates in Philosophy of Mind*, eds B. P. McLaughlin and J. Cohen (New York, NY: Blackwell Publishing), 117–138.
- Lupyan, G. (2012). Linguistically modulated perception and cognition: the label-feedback hypothesis. *Front. Psychol.* 3:54. doi: 10.3389/fpsyg.2012.00054
- Lupyan, G. (2015). Cognitive penetrability of perception in the age of prediction: predictive systems are penetrable systems. *Rev. Philos. Psychol.* 6, 547–569. doi: 10.1007/s13164-015-0253-4
- Macpherson, F. (2011a). Cross-modal experiences. *Proc. Aristotelian Soc.* 111, 429–468. doi: 10.1111/j.1467-9264.2011.00317.x
- Macpherson, F. (2011b). "Introduction: The admissible contents of experience," in *The Admissible Contents of Experience*, eds F. Macpherson and K. Hawley (Chichester: Wiley-Blackwell), 1–15.
- Macpherson, F. (2012). Cognitive penetration of colour experience: Rethinking the issue in light of an indirect mechanism. *Philos. Phenomenol. Res.* 84, 24–62. doi: 10.1111/j.1933-1592.2010.00481.x

- Macpherson, F. (2016). The relationship between cognitive penetration and predictive coding. *Conscious. Cogn.* doi: 10.1016/j.concog.2016.04.001. [Epub ahead of print]
- Matthen, M. (ed.). (2015). *The Oxford Handbook of Philosophy of Perception*. Oxford: Oxford University Press.
- Mroczko-Wąsowicz, A. (2013). *The Unity of Consciousness and the Phenomenon of Synesthesia [Die Einheit des Bewusstseins und das Phänomen der Synästhesie]*. Published doctoral dissertation, Johannes Gutenberg University of Mainz.
- Mroczko-Wąsowicz, A. (2016). "The unity and complexity of flavour perception: multisensory phenomenology and processing in human-food interactions," in *Proceedings of ICMI 2016 Workshop on Multi-sensorial Approaches to Human-Food Interaction* (New York, NY: ACM Digital Library). doi: 10.1145/3007577.3007583
- Mroczko-Wąsowicz, A., and Nikolić, D. (2014). Semantic mechanisms may be responsible for developing synesthesia. *Front. Hum. Neurosci.* 8:509. doi: 10.3389/fnhum.2014.00509
- Mroczko-Wąsowicz, A., and Werning, M. (2012). Synesthesia, sensory-motor contingency and semantic emulation: how swimming style-color synesthesia challenges the traditional view of synesthesia. *Front. Psychol.* 3:279. doi: 10.3389/fpsyg.2012.00279
- O'Callaghan, C. (2014). "Intermodal binding awareness," in *Sensory Integration and the Unity of Consciousness*, eds D. Bennett and C. Hill (Cambridge, MA: MIT Press), 73–104.
- O'Regan, J. K., and Noë, A. (2001). A sensorimotor approach to vision and visual consciousness. *Behav. Brain Sci.* 24, 883–975. doi: 10.1017/S0140525X0100115
- Prinz, J. (2011). "The sensory basis of cognitive phenomenology," in *Cognitive Phenomenology*, eds T. Bayne and M. Montague (Oxford: Oxford University Press), 174–196.
- Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behav. Brain Sci.* 22, 341–423.
- Pylyshyn, Z. (2003). *Seeing and Visualizing: It's Not What You Think*. Cambridge, MA: MIT Press.
- Raftopoulos, A. (2011). Late vision: processes and epistemic status. *Front. Psychol.* 2:382. doi: 10.3389/fpsyg.2011.00382
- Raftopoulos, A., and Zeimbekis, J. (eds.). (2016). *Cognitive Penetrability*. Cambridge, MA: MIT Press.
- Shapiro, L. (ed.). (2014). *The Routledge Handbook of Embodied Cognition*. London; New York, NY: Routledge.
- Shimojo, S., and Shams, L. (2001). Sensory modalities are not separate modalities: plasticity and interactions. *Curr. Opin. Neurobiol.* 11, 505–509. doi: 10.1016/S0959-4388(00)00241-5
- Siegel, S. (2012). Cognitive penetrability and perceptual justification. *Noûs* 46, 201–222. doi: 10.1111/j.1468-0068.2010.00786.x
- Spence, C., and Bayne, T. (2015). "Is consciousness multisensory?" in *Perception and its Modalities*, eds D. Stokes, M. Matthen, and S. Biggs. (New York, NY: Oxford University Press), 95–132.
- Stein, B. E. (ed.). (2012). *The New Handbook of Multisensory Processing*. Cambridge, MA: MIT Press.
- Stokes, D. (2012). Perceiving and desiring: a new look at the cognitive penetrability of experience. *Philos. Stud.* 158, 479–492. doi: 10.1007/s11098-010-9688-8
- Stokes, D., Matthen, M., and Biggs, S. (eds.). (2015). *Perception and its Modalities*. New York, NY: Oxford University Press.
- Tajadura-Jiménez, A., Bianchi-Berthouze, N., Furfaro, E., and Bevilacqua, F. (2015). Sonification of surface tapping changes behavior, surface perception, and emotion. *IEEE MultiMedia* 22, 48–57. doi: 10.1109/MMUL.2015.14
- Tajadura-Jiménez, A., Pantelidou, G., Rebacz, P., Västfjäll, D., and Tsakiris, M. (2011). I-space: the effects of emotional valence and source of music on interpersonal distance. *PLoS ONE* 6:e26083. doi: 10.1371/journal.pone.0026083
- Vetter, P., and Newen, A. (2014). Varieties of cognitive penetration in visual perception. *Conscious. Cogn.* 27, 62–75. doi: 10.1016/j.concog.2014.04.007
- Weiskopf, D. (2015). "Observational concepts," in *The Conceptual Mind: New Directions in the Study of Concepts*, eds E. Margolis and S. Laurence (Cambridge: MIT Press), 223–247.
- Weiss, C., Tsakiris, M., Haggard, P., and Schütz-Bosbach, S. (2013). Agency in the sensorimotor system and its relation to explicit action awareness. *Neuropsychologia*. 52, 82–92. doi: 10.1016/j.neuropsychologia.2013.09.034

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

Copyright © 2016 Mroczko-Wąsowicz. 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.



Rethinking the senses and their interactions: the case for sensory pluralism

Matthew Fulkerson *

Department of Philosophy, University of California, San Diego, La Jolla, CA, USA

Edited by:

Aleksandra Mroczko-Wąsowicz,
National Yang Ming University,
Taiwan

Reviewed by:

Tom Froese, Universidad Nacional
Autónoma de México, Mexico
Kathleen Akins, Simon Fraser
University, Canada

***Correspondence:**

Matthew Fulkerson, Department of
Philosophy, University of California,
San Diego, 9500 Gilman Drive, La
Jolla, CA 92093, USA
e-mail: mfulkerson@ucsd.edu

I argue for sensory pluralism. This is the view that there are many forms of sensory interaction and unity, and no single category that classifies them all. In other words, sensory interactions do not form a single natural kind. This view suggests that how we classify sensory systems (and the experiences they generate) partly depends on our explanatory purposes. I begin with a detailed discussion of the issue as it arises for our understanding of thermal perception, followed by a general account and defense of sensory pluralism.

Keywords: perception, multisensory processing, sense organs, pluralism in modeling, thermoreception

1. INTRODUCTION

Start with two seemingly true statements: (i) We have many senses; (ii) They often interact.

These statements are now widely acknowledged and incorporated into recent work on perception, but they are also in deep tension with one another. Once we allow that the sensory modalities interact, and do so pervasively at multiple levels of sensory processing, with effects at all levels of our psychology (subpersonal, behavioral, and phenomenal), then it becomes difficult to make sense of what, exactly, these individual senses might be. Vision is less a single coherent modality than a complex collection of interacting subsystems. And that collection has features different in kind from those found in the auditory, vestibular, and nociceptive systems (of course, there are many similarities too). Indeed, it can become difficult to maintain the idea that we can have anything like a unified conception of sensory modalities and their interactions.

I start with a detailed discussion of human thermoreception, using it as a case study for the sort of tensions I describe above. I then discuss the general implications of this example, and propose a robust theoretical framework for addressing this tension. My claim is that we should abandon any single theoretical account of sensory interaction, and adopt a view according to which sensory systems and their interactions are classified in part by our explanatory purposes. The upshot of this proposal is that it allows us to fully acknowledge the deep interactions between sensory subsystems without thereby giving up entirely on the very idea of separate sensory modalities. The main target of my view is any form of sensory monism that assumes there will be a single, authoritative, and context free account of what it is to be a sensory modality and for an interaction between them to be “multisensory” or “multimodal.” On such a monist view, there should be a single determinate answer to the question of whether

vestibular awareness or pain or any other putative sense counts as a sensory modality. I believe such a view is implausible and deeply problematic, and in what follows I offer a substantive alternative account.

2. CASE STUDY: THERMAL PERCEPTION

We have a sensory system—commonly called the thermoreceptive system—that involves a series of distinct receptor populations in the skin (Schepers and Ringkamp, 2010). There are several different kinds of receptors involved, including thinly myelinated Aδ afferents that have receptive fields tuned to cooling and unmyelinated C afferents that code for both warming and cooling¹. These various receptor populations systematically combine with other cutaneous systems (like those that code for pressure, vibration, and shape) to inform us about thermal properties in the distal environment (Jones and Lederman, 2006; Lumpkin and Caterina, 2007). They thus seem to be a crucial component of haptic touch (Fulkerson, 2014b). They also play an important role in our bodily awareness and the regulation of body temperature (Hammel and Pierce, 2002; Jones and Lederman, 2006), and so seem also to belong to our general systems of bodily awareness (which includes proprioception, kinesthesia, and our body schema). And finally, thermoreceptors also play an important role in the nociceptive system, informing us of bodily damage caused by extreme hot and cold stimuli².

¹As we'll see, the different response patterns of these many afferent channels play different roles in different contexts. For instance, the thinly myelinated Aδ fibers play an especially important role in our perception of wetness (Ackerley et al., 2012).

²That pain you feel when eating spicy foods? It's caused by the activation of thermoreceptors. Black pepper contains *piperine* and chili peppers contain *capsaicin*, both of which both activate TRPV1 thermoreceptors. See Caterina et al. (1997).

How should we classify this thermoreceptor system? Is it even one thing, given its many different afferent populations with distinct receptive fields and activation profiles? Maybe thermoreception itself is multisensory? We can also ask whether it is a part of touch. Should it be examined and investigated along with the other constituents of haptic awareness? Or are these thermoreceptors really part of the nociceptive system? Exposure to extreme heat and cold are, after all, among our most intense causes of pain. Then again, perhaps it is part of our general system for bodily awareness, since such thermoreceptors play such an important role in the regulation of a comfortable bodily state. In each of these cases, we can ask whether that makes touch, pain, and bodily awareness essentially multisensory³.

Similarly, we might wonder whether thermoreception is its own independent sensory modality (multisensory or not). Is it *perceptual*, or do we only become informed of distal thermal properties indirectly, through inference from our bodily thermal state?⁴ Each of these positions has been defended (sometimes tacitly) in the literature (Martin, 1992; Schepers and Ringkamp, 2010; Gray, 2012). We are unlikely to make much progress on these claims, I believe, until we realize that there really is no such thing as the thermoreceptive system. The starting assumption that there is such a single system leads, I shall argue, to insurmountable theoretical and practical difficulties. Instead of a single thermalreceptive system, I believe that we have a complex series of receptors and processing units that perform *multiple* overlapping functions, and thus there are many, equally good ways of categorizing these various systems (see **Figure 1**). On this view, relative to one schema (its role in detecting and co-assigning features to distal objects), the thermoreceptive system is indeed continuous with (and therefore an essential part of) the sense of touch (itself a context-sensitive construct). If we focus purely on the physiological features of thermoreception, on the other hand, we have strong reason to classify (some elements of) this system as continuous with other elements of the nociceptive system. Like those other systems, many thermal channels involve slow, unmyelinated afferent nerve fibers that project contralaterally in the spinal column (unlike discriminatory touch afferents, which are typically

myelinated and project ipsilaterally, Welsh, 2001). According to a third schema, we can see that thermoreceptors also play an important role in the awareness and regulation of body temperature, and can be classified as part of a larger system of bodily awareness that includes proprioception, vestibular awareness, and other regulatory systems (Wenger, 1995).

Let's focus on the details of this last claim, that our thermoreceptive system is part of a larger system of bodily awareness (the details here are useful as an illustration; this is not intended as an exhaustive argument about how to understand the thermal system). One useful way of categorizing sensory systems concerns whether they are outward-facing, giving us information about the external world, or body-facing, giving us information and performing functions primarily dedicated to bodily awareness and homeostatic self-maintenance. This distinction is often thought to be one between two separate systems: the *exteroceptive system* and the *interoceptive system*⁵. Exteroception, the story goes, provides us with information about the external world. It is largely informational and descriptive, giving us an evidence-like connection with the world around us. It is helpful for practical purposes (helping us find food and shelter and avoid dangers) but also for epistemic matters (helping us learn, form beliefs, and plan). Interoception, on the other hand, concerns the present state of our bodies. It is simply not in the business of directly reporting on what's going on in the external environment. Instead, this system is wholly concerned with *regulating* the present state of the body.

I want to focus on the defense of the view that thermoreception is interoceptive found in Craig (2002).

Craig suggests on the basis of physiological and functional connections that the thermal system should be categorized as part of the interoceptive system; that it tells us more about the present state of our bodies than it does about the external world (and that it does the latter only as a kind of secondary function). In doing so, it functions to maintain balance in our bodily system, and it does so in a way very similar to the operation of other homeostatic systems like those for hunger, thirst, and pain (see also Nakamura and Morrison, 2007). Here Craig (2002) describes the interoceptive system (emphasis mine):

This system is a homeostatic afferent pathway that conveys signals from small-diameter primary afferents that represent the physiological status of all tissues of the body. It projects first to autonomic and homeostatic centers in the spinal cord and brainstem, thereby providing the long-missing afferent complement of the efferent autonomic nervous system. Together with afferent activity that is relayed by the nucleus of the solitary tract (NTS), it generates a direct thalamocortical representation of the state of the body in primates that is crucial for temperature, pain, itch and other

³Nociception represents another ideal case study in the difficulties facing any unified account of sensory interactions. For a convincing argument here see Corns (2014). For earlier discussion see Aydede and Guzeldere (2002).

⁴See for example Gray (2012) for a nuanced discussion of what our thermal experiences might represent even if focussed on the body.

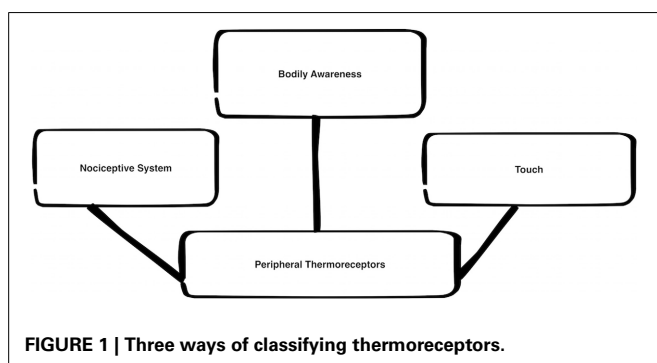


FIGURE 1 | Three ways of classifying thermoreceptors.

⁵It is interesting to compare this with the distinction between *emotional* and *discriminative* touch (McGlone et al., 2007). These categories overlap in several respects; indeed, McGlone et al. (2007) use some of Craig's findings and terminology to help mark the difference between emotional and discriminative. However, they note some interesting cases of overlap between physiologically distinct afferent populations (p. 176), and, as we'll see, they end up defending a dualist (i.e., pluralist) view of touch.

somatic feelings. This anatomical organization shows that these feelings are highly resolved, sensory aspects of ongoing homeostasis that represent the physiological condition of the body itself—a *distinct shift from the concept that pain and temperature are aspects of touch* (p. 655).

Notice that the key evidence for lumping these elements together into a single system are physiological. On my view, as we'll see, this is a perfectly appropriate context for categorizing these constituent systems. It just isn't the only such context.

While debates continue about whether hunger, thirst, and pain should be seen as perceptual (see e.g., Aydede, 2009), there is little debate about whether they represent something external to the body. On nearly all views, when we feel hungry, we are learning something about the present physiological state of our bodies rather than something about the external environment.

But why think that our thermoreceptive system (or rather many of its constituent parts) can play only one role, and must be either interoceptive or exteroceptive? A much better alternative here is to go pluralist: Craig is correct that there are explanatory schemes according to which it makes most sense to classify thermoreception along with other systems of bodily awareness like hunger and thirst. As Nakamura and Morrison (2007) write:

To evoke behavioral, autonomic, somatic and hormonal responses that counteract changes in environmental temperature before they affect body core temperature, thermoregulatory command neurons in the POA [the preoptic area] need to receive feedforward signaling of environmental temperature information from skin thermoreceptors through the spinal and trigeminal dorsal horns (p. 62).

Thermoreceptors, when they project to the POA and other areas that control homeostatic control, play a critical role in regulating our overall body temperature. We can recognize this without denying that we can *also* directly sense thermal features of the external environment when, for instance, our cutaneous thermoreceptors are co-activated with other constituents of externally directed haptic perception, whose afferents project directly to other areas of the somatosensory cortex.

Let's consider this second role in more detail⁶. When we actively explore an object with our hands, for instance, the synchronized motor engagement and cutaneous activation generate awareness of external objects and their thermal properties (Fulkerson, 2014b). This is, after all, how we successfully check whether the bath water is too hot, or whether the white wine is sufficiently chilled (cf. Jones and Lederman, 2006).

When we touch the bath water, we are attempting to determine the thermal state *of the water*. We reach our hand (or wrist, elbow, etc.) into the water, feel its temperature, and then decide whether or not the water is too hot. When we feel that the water is merely warm, we seem to succeed just fine in determining something about the state of the world, and not simply through some kind of explicit inference (Schepers and Ringkamp, 2010). Our experience is, it seems, about the state

of the water. I've already suggested a functional reason for this: the activated thermoreceptors are not acting or interpreted on their own; they are temporally and spatially aligned with our exploratory actions and other cutaneous afferents in a way that unifies and enriches their informational content. This point can be supplemented by the fact that thermal properties bind to other external properties, forming complex tangible blends that involve the association of distinct distal properties. Thermal properties, for instance, turn out to be one of the essential elements in our experience of wetness (Sullivan, 1923; Ackerley et al., 2012) and material composition (Jones and Lederman, 2006), allowing us to differentiate an equally smooth wooden surface from a metal one. As Ackerley et al. (2012, p. 73) note: "Skin afferents are rarely composed of just one sensory modality and some sensory receptors are polymodal. Furthermore, perception of the sensation usually occurs from a blend of inputs, for example, when we sense that something is wet, it is typically due to changes in both touch and temperature afferents. There is no evidence to suggest that we have wetness receptors in the skin." While these researchers often speak of touch and temperature awareness as separate things, their own evidence suggests otherwise. Relative to a purely physiological criteria, we can categorize touch and temperature as separate modalities, but when they function reliably to bring awareness of wetness and material composition they are better categorized as part of a single haptic system⁷. This is because material composition is one of the most important elements in tactual object-recognition (Klatzky and Lederman, 2008). None of these would be the case if thermal reception *only* informed us of the state of our bodies, or if external awareness involved a separate inferential step beyond the sensory level. Indeed, this issue is not unique to thermal awareness, since touch itself involves a great variety of distinct receptor types, projection sites, and downstream behavioral and psychological effects. As McGlone et al. (2007) end their discussion:

In conclusion, a dual role for touch serving both a discriminative and an affective role in human behavior has been described. The human hand has clearly evolved to perform a wide range of exploratory and manipulative tasks, and far surpasses this function in any other primate (p. 181).

So we can see that the thermoceptive system really does play (at least) dual roles in creatures like us. In addition, of course, the use of physiological and functional criteria to categorize the various constituents of the interoceptive system brings together a diverse range of systems that cross-cut in other interesting ways. While Craig suggests that all of these interoceptive systems have affective valence, for instance, so too do most externally-directed perceptual systems. Seeing something gruesome or disgusting often brings a strong emotional reaction. Vision and audition also play an important role in proprioception and wayfinding (Campos et al., 2012). While several of these systems also

⁶I set aside for now the details concerning the role of thermoreception in nociception.

⁷Indeed, I would argue that our tendency to talk of "thermoreception" and "touch" as unified entities also involves eliding important physiological and functional differences between the large number of distinct receptor populations involved in each system. We'll return to this general tendency in a later section.

have some homeostatic function, they also differ in many other respects. The idea that there is a single “interoceptive system” is itself a useful explanatory context. We should not at all be surprised that there will be equally useful alternative ways of categorizing these systems.

Contrast Craig’s view with that of Akins (1996). Like Craig, she denies that the senses are always in the business of veridically reporting on conditions external to an organism⁸.

She describes the traditional naturalistic account of senses, one that she will go on to deny, as follows: “The senses show the brain, otherwise blind, how things stand “out there,” both in the external world and in its own distal body” (p. 342). Later, she adds: “On the traditional picture, then, the senses, using a system of signals that capture the structure of a domain of external properties, tell the brain, without exaggeration or omission, “what is where” (p. 344).

This traditional (though still prevalent) view can motivate the idea that the senses have a single, hard-wired role to play (i.e., reporting external conditions directly to the brain). Conveniently for my purposes, she argues against this view with a detailed consideration of the peripheral thermoreceptive system. On the traditional view, “The receptors ... must react with a unique signal, one that correlates with a particular temperature state.” (p. 342). Of course, this is not how peripheral thermoreceptors function. They have highly context sensitive and variable response rates that depend on the present state of the skin and embedded receptors, the context of activation, and the homeostatic needs of the organism (consider as Akins does the contrasting experiences generated by placing a warmed and a chilled hand in a neutral glass of water).

These facts lead Akins to suggest that sensory systems are “narcissistic”: while they sometimes convey information about the external environment, they always do so in a way that reflects first and foremost the needs and priorities of the organism. These needs, in turn, are often variable and highly context-sensitive. The senses involve many interacting parts, playing many different and important roles, but always *for the organism*. This perspective supports Craig’s insights about the homeostatic and internally-directed nature of interoceptive thermal responses. Interoceptive contents, after all, will almost by definition be narcissistic. However, once the peripheral transducers are seen correctly as the initial components of much larger downstream neural networks subserving a variety of distinct psychological and behavioral activities, we can more easily see how the several channels involved in thermoreception can, in different contexts, and when connected with different downstream systems, be (literally) a part of several distinct sensory systems⁹.

The upshot then is not that Craig is wrong to apply the interoceptive category to some sensory systems. It’s that he can be correct that there is an interesting and important way of connecting these systems, without excluding alternative ways of categorizing them. On the moderate pluralist view I will go on

to defend, we can allow that thermal perception plays multiple different roles. Indeed, we can think of this system as a single system only by applying such a scheme of classification. There are a variety of distinct overlapping systems involved in thermoreception, and there are thus many different ways of classifying them. There is, on my view, both an internal and an externally-directed role. Thermoreception *really* is an important part of touch. It *really* is a part of our pain system. It *really* is part of bodily awareness. Which aspect we focus on depends on which aspects of the system we’re interested in, and our explanatory purposes.

One thing is clear: even if one of the key functions of thermal perception is to provide information about the present state of our bodies, it does not follow that this is the *only* thing that thermal perception does. Or at least, it does not follow that there aren’t multiple variants of thermal systems, all making use of the very same initial populations of peripheral thermoreceptors. One provides bodily information, another is connected with our haptic exploratory system, another plays a critical role in our pain experiences. Given this possibility, which I take to be an actuality, one should not make any inferences about perception generally on the basis of one function of thermal experience.

The upshot for us is that Craig highlights only *one* of the key functions of the thermal system, and his work allows us to see how (parts of) the same system can serve a variety of distinct roles. Some forms of thermal awareness only deliver awareness of the present state of our bodies; others inform us of the thermal properties of objects in our immediate environment. The real nature of thermoreception depends on what we are trying to explain, and on which associated features of the systems we are categorizing.

Thermoreception represents a kind of ideal case study here: it is a complex system, but well enough understood that we can use it to see exactly how plausible and powerful the moderate pluralist view can be. Now I will fill in the details of the sort of view I have in mind, starting with some essential background.

3. THE IMPORTANCE OF MULTISENSORY INTERACTION

Recall the statements that began this paper: (i) We have many senses; (ii) They often interact.

These two statements were for a long time discounted by those in the cognitive sciences. Many had what O’Callaghan (2007, 2008) has called a “visuocentric” conception of perceptual experience. Visual experience was discussed to the exclusion of other modalities, and it was tacitly assumed that the conclusions reached for visual experiences would translate smoothly over to the other senses.

Recent work in cognitive science has accepted a more nuanced, multisensory conception of perceptual experience¹⁰. Empirically informed philosophy of mind has similarly seen a transformation in our understanding of perception¹¹. Recent philosophy has seen

⁸Her focus is on sensory *contents* rather than processing, but the general point is the same.

⁹I am grateful here to Kathleen Akins for extremely helpful discussion of this material.

¹⁰The literature here is enormous; a good place to start is Stein (2012).

¹¹The transition in philosophy was largely spearheaded by O’Callaghan (2008, 2007). Evans (1982) and Martin (1992) are other important early sources. For the most recent treatment, see the articles in Bennett and Hill (2014).

a increase in research on other modalities¹², sensory interactions¹³, and on the individuation of the senses¹⁴.

Of course, much of this philosophical work has been informed by and is directly responding to work in the various cognitive sciences exploring the deep interconnections and interactions between the senses.

Researchers have focused extensively on many different elements of sensory interaction, from cross-modal illusions, in which activations in one modality alter or suppress activations in another, to other categories of interaction like sensory facilitation, dominance, and several distinct forms of sensory integration. An increasing focus recently has been on more complex instances of sensory interaction like those occurring in affective experience, cognitive penetration, and synesthesia¹⁵. In all cases, researchers have extensively documented deep and pervasive interactions between sensory modalities.

These discoveries have largely undercut the “visuocentric” assumptions found in earlier research, and challenge many simplistic conceptions of sensory experience. Of course, one can still find work devoted entirely to vision (and to other individual modalities), but now such work is typically much more self-conscious about the limitations of focusing on a single modality studied in isolation. This recent shift has brought with it many important advances in our understanding of sensory interactions and the nature of perceptual consciousness, and has been a good thing for those of us trying to better understand the nature of perception.

As with any large shift in the scientific landscape, the new multisensory focus has also raised a number of important theoretical questions and posed novel challenges. I want to suggest that the move from our prior conception of separate individual senses requires more than merely investigating non-visual modalities or considering some sensory interactions. The move to a multisensory framework requires a more substantial reorientation of the theoretical landscape and of our investigative practices. At the same time, we should resist the urge to completely abandon all talk of senses and sensory systems. Instead, I will argue for an intermediate view that rejects any single, unified account of sensory modalities and their interactions, instead embracing a multitude of such accounts.

Before discussing these details, it's necessary to make two caveats. First, my focus in this paper is on the cognitive science classification of sensory systems. When I talk about *vision* or *audition*, I'm primarily interested in how we individuate and classify for the purposes of scientific explanation a particular part of our

psychological biology. I am interested in the systems on the plausible assumption that it is those systems that are the constitutive and computational basis of the experiences generated¹⁶.

While this is a substantive assumption, the pluralist view does not depend on it (see the discussion of sensory substitution in §7.4 where this commitment is eased). It is an important advantage of my view that it allows and indeed embraces the idea that our perceptual experiences can be investigated and understood in multiple ways. So while the discussion that follows focuses almost exclusively on the sensory systems underlying our perceptual experiences rather than on their phenomenological, dynamic, or epistemic features, the view is ultimately sympathetic to many seemingly different approaches to understanding perceptual experience¹⁷.

Second, pluralist views have been discussed in a range of areas, especially in philosophy of biology (Kitcher, 1984; Mishler and Brandon, 1987; Ereshefsky, 1992; Steel, 2004; Cleland, 2013), but also in cognitive science (Dale et al., 2009), in general philosophy of science (Cartwright, 1999; Mitchell, 2002), aesthetics (Mag Uidhir and Magnus, 2011), and elsewhere. The view I defend in what follows was not initially inspired by this general move toward pluralism. Instead, it arose as a specific reaction to recent work on sensory interactions. It is not, therefore, the application of a form of pluralism defended in another domain to the sensory case. Instead, the view is motivated entirely by considerations internal to issues of explaining sensory interaction. For this reason, in what follows I will not engage in any systematic examination or comparisons between sensory pluralism and the many similar views defended in other domains, nor do I claim any special affiliation with such views.

4. THEORETICAL OPTIONS

In this section, I want to spell out in general terms the nature of the tension forced on us by the move to a multisensory conception of perception, and consider the theoretical options.

We start with assumption (i) that we have many senses. An implicit assumption here is that these senses are more or less self-contained entities (it doesn't matter whether we think of them

¹²See for instance, recent work on smell (Batty, 2010; Richardson, 2011); on taste (Smith, 2009, 2012); on sounds (Nudds, 2001; Matthen, 2010), and on touch (O'Shaughnessy, 1989; Scott, 2001; Ratcliffe, 2008; Richardson, 2013; Fulkerson, 2014b). In addition, there has been work on many forms of experience outside the traditional five senses, for instance on temporal experience (Grush, 2005; Phillips, 2008; Lee, 2014) and on bodily awareness (de Vignemont, 2007; Schwenkler, 2011).

¹³O'Callaghan (2012); Macpherson (2011); Bennett and Hill (2014)

¹⁴See for example Keeley (2002); Gray (2005); Nudds (2003); Macpherson (2010); Matthen (in press).

¹⁵See for example Mroczko-Wasowicz and Nikolic (2014); Vuilleumier and Driver (2007); Stokes (2012); Siegel (2011); Villemure et al. (2003).

¹⁶While my own preference is for representational and computational approaches to the mind, this assumption is not intended to rule out approaches to cognition and perception that emphasize the strong connections between this biology and the external world. For instance, it should not be taken to exclude views like those recently defended by Noë (2004); Hurley (1998); Thompson (2007). Indeed, as I say in the text, it is an advantage of my position that it leaves space for a variety of explanatory approaches to the study of perceptual experience, including embodied and enactive views. For example, my view leaves space for contexts in which a view of modality along the lines defended by McGann (2010) is appropriate. (A critical difference is that McGann's view is strongly eliminativist, holding that “there is no such thing as an experience that is purely visual, auditory, or otherwise modal” (p. 72). By my lights, whether there are such experiences is dependent on the explanatory context, and what we mean by “visual,” “auditory,” and the like.)

¹⁷Though I should emphasize that, in my own view, such approaches are useless unless constrained and informed by the empirical facts.

at this point as systems, modes of awareness, or forms of experience, etc.). One monist view that has been very influential is the claim that the senses are *modular input systems* (Fodor, 1983; Pylyshyn, 2006). On this view, the senses are domain specific, informationally encapsulated, hard-wired, and fast systems that function to process incoming sensory information. According to the modular account, we have a strong physiological, informational, functional, and computational distinction between sensory modalities. Vision uses different biological hardware than audition, to carry different information, for different computational and behavioral purposes¹⁸. The modular account is just one influential monist accounts in the literature. Its strength is supplemented by our strong intuitive sense that the senses represent very different forms of conscious awareness. What could be more clear than the difference between seeing something and hearing something? The view, and other weaker versions of monism, are systematically unable to adjust to the known facts about sensory interactions. This brings us to our second beginning statement.

That our senses interact (ii) seriously undermines any monist conception of sensory modality and interaction. We have learned that the senses interact in many interesting ways, often completely hidden from introspection. It has taken careful investigation to realize just how pervasive these influences can be. Much has been made, rightly, about the existence of cross-modal illusions (O'Callaghan, 2008). The McGurk effect shows that very often, what we hear is determined by what we see McGurk and MacDonald (1976); Skipper et al. (2007). The motion-bounce illusion shows that what we see is often partly determined by what we hear (Sekuler et al., 1997). The use of brain scans and single-recording techniques has shown that many distinct areas of sensory cortex are active and engaged in the generation of experiences in single modalities (Ghazanfar and Schroeder, 2006). Similarly, vestibular and proprioceptive information influences activations in other modalities (Frissen et al., 2011; Campos et al., 2012). Motor movements influence cutaneous activations (Chapman, 1994). Thermal receptors influence pressure awareness (Jones and Lederman, 2006). What we see influences what we smell (Herz and von Clef, 2001). And on and on.

Once we realize just how pervasive and varied these interactions can be, we really start to lose grip on our what these separate senses involved are supposed to be. If they are not domain-specific, if they are not physiologically and informationally isolated, if they serve many varied and interactive functions, if the experiences they generate are fused into complexes that aren't easily decomposed or isolated in experience, then in what sense are they really distinct sensory systems *at all*? They certainly aren't isolated or independent. The facts of sensory interaction make it a very difficult theoretical challenge to say exactly what the senses referred to in (i) might actually be. Depending on how we think about multisensory interactions, it can become difficult to avoid

the conclusion that we really don't have separate senses after all. Vision becomes a complex of various subsystems, each connected in various ways with many other sensory subsystems and aspects of cognition. Instead, we just have a vast mess of sensory interactions (maybe at the lowest level of sensory subsystems)¹⁹. I am not the first to notice these challenges. Consider the recent paper by Deroy et al. (2014), where they lay out many of the challenges facing the move to a multisensory conception of perception. As they note, there seem to be at present no clear experimental methods to directly investigate multisensory awareness or to distinguish between various models of sensory interaction. This problem is compounded, I believe, by appeal to several distinct forms of sensory interaction, including distinct levels of investigation (at the neurophysiological, behavioral, and introspective levels) and different forms of interaction (cross-modal influences, sensory blends, multimodal conjunctions). They are asking the right questions:

Can we simply take the current theories and protocols used to try and understand unisensory cases and then import them into the field of multisensory research? This is the approach that we wish to question here ...shifting to multisensory cases is not cost-free for the study of perceptual awareness. It introduces both methodological and theoretical pressures. (Deroy et al., 2014, p. 3).

These pressures are compounded by the diversity of theoretical questions and experimental methods involved in these investigations. As they note later, "The recycling of unisensory protocols is unlikely to provide good ways to study multisensory awareness, if there is indeed such a thing" (Deroy et al., 2014, p. 8). My proposal suggests that these difficulties are not simply temporary impediments in our understanding of sensory awareness; they are the inevitable result of trying to fit a heterogeneous class of interactions under a single category (either unisensory vs multisensory, full stop). Consider a simplified example to support this claim.

Suppose that we are thinking about sensory interactions as occurring fundamentally between informational systems, and we characterize these (roughly) in terms of informational processing. If we do this, we can think about the interactions of the senses as constituted by the sharing and interaction among separate informational channels (for details on how this might go, see my 2011). What happens immediately, however, is that vision and audition no longer constitute anything like a single coherent sensory modality. They are complex systems that themselves involve interactions among disparate sensory subsystems sharing information in lots of interesting ways. *This happens for any other criteria we try to use to define sensory modalities*²⁰.

Since the interactions that operate in vision cross all kinds of boundaries, it becomes difficult to make sense of what counts as *the* visual system. Do those auditory processing centers that function reliably and consistently to contribute to the nature of our visual experiences count as part of vision? What about the pervasive influence of vestibular and proprioceptive systems on vision? And what do we mean by *visual experiences*? Once you start taking

¹⁸Keeley (2002) can be read in some ways as regimenting many of these criteria as a means of individuating sensory modalities. In my view it is the strongest statement of sensory monism, but as it predates the current interest in multisensory perception, it does not even make an attempt to incorporate facts about sensory interactions.

¹⁹A view like this seems to be advocated by Shimojo and Shams (2001).

²⁰See Fulkerson (2011) for one elaboration of this claim.

seriously the fact that the senses interact, and you start looking at the details of these interactions, it can be incredibly difficult to make sense of what we're actually talking about. The very idea of a *visual* system, or of a *visual* experience, starts to break down. So the worries we encountered with thermoreception are not unique to that domain; they are pervasive issues that arise for all putative perceptual modalities²¹.

As I see it, there are three ways to settle respond to this tension. We can preserve and supplement the status quo through some form of sensory monism, we can reject the entire project of sensory classification and go eliminativist, or we can go pluralist²².

4.1. OPTION ONE

We can reject the claim that there is any tension or threat to the notion of a sensory modality posed by pervasive sensory interactions. One could, for instance, maintain the notion of the individual senses and try to explain multisensory interactions in ways that don't challenge the orthodox view of the senses. Connolly (2014) makes such an argument. In my commentary on Connolly's paper (Fulkerson, 2014a), I called such a view *sensory conservatism*; however, in this context I would describe it as a form of *sensory monism*. The idea is that we find some unified account that preserves the traditional notion of separate sensory modalities. Part of what this means is that we account for the wide range of sensory interactions by appeal to a criteria of sensory interaction that is independent from our criteria for being a sensory modality. We find a way to show that the traditional five (or more) senses remain of explanatory importance, and we account for multisensory interactions in a way that doesn't undermine these very kinds.

This is not an easy thing to pull off. For one, no one has yet suggested a criterion of sensory individuation that preserves the notion in the light of pervasive sensory interactions. The senses, whatever they are, cannot be domain specific, or functionally-unified, or marked in phenomenology, or physiologically specified, since what we call vision and audition and touch and olfaction and gustation have none of these features²³.

4.2. OPTION TWO

Instead of sensory monism, one could opt for eliminativism. One could hold that the traditional senses (and their various interactions) are a kind of false construct or simplified idealization, and propose that we reject all such talk from our theorizing. Recent

advances have demonstrated that our experience of the world is generated by a large number of interacting processing units. The natural way of thinking about sensory systems, on this view, is at a much finer grain than anything like modalities. Modalities are huge, messy collections of complex systems that involve mutually-interacting connections with numerous areas of the brain. They aren't natural kinds at all. On this view, to take the idea of sensory interactions seriously requires a much more radical shift in our thinking than we might have originally expected. In fact, it seems to require a rejection of (i). That is, it seems we ought to reject our intuitive notion of separate sensory modalities, and understand sensory interactions as pervasive "all the way down." In the end, there really are no senses. This view has been defended most explicitly by Shimojo and Shams (2001), and one can see echoes of it in the work of many others (e.g., Driver and Spence, 2000; O'Callaghan, 2008).

4.3. OPTION THREE

Instead of adopting monism or eliminativism, I argue instead that we should adopt sensory pluralism. This is the view that there are indeed separate modalities, and natural ways of carving up sensory systems and their interactions, just like the monist believes; but like the eliminativist, the pluralist holds that no *single* account of modality and interaction is forthcoming. Against these views, the pluralist holds that there are *many* criteria of sensory interaction and unity, and these criteria in turn partly depend on our explanatory purposes and the investigative context. In other words, we should be pluralists about the senses and their interactions.

While some versions of pluralism can involve a radical ontology, the moderate view I have in mind is neither radical nor ontologically profligate. It simply holds that sensory systems are complexes that can be fruitfully engaged in many ways. Instead of calling it *pluralism*, one could, following Evans (1982) on reference, simply catalog and describe the *variety* of sensory interactions. Or, like Matthen (2010), one could focus on the *diversity* of sensory classification. These differences in terms do not track a real difference in the view I have in mind. I simply use the label *sensory pluralism* to name the view that sensory interactions come in many different forms, and therefore do not form a (single) natural kind.

My brand of sensory pluralism is moderate and constrained by the fact that there are indeed better and worse ways of dividing sensory interactions (more strongly: there are legitimate and illegitimate forms of sensory classification). Yet among the good ways, there are multiple equally useful options relative to our purposes. These cross-classify our perceptual systems in ways that can seem deeply at odds with each other, though in reality they enrich and mutually support our understanding of sensory experience. For instance, we can investigate as a single entity the causal-detection system composed of several seemingly distinct sensory modalities (Michotte, 1963). There is a legitimate theoretical and empirical question about whether this system of classification really is legitimate (as far as know, the jury is still out on this question). The pluralism I defend is thus modest rather than revisionary: it acknowledges the inherent complexity and deep interconnections between sensory systems at different

²¹I want to emphasize that, while I speak here and throughout about the classification of sensory *modalities*, my target is broader than issues about the individuation of the senses. My primary concern, in fact, concerns how best to classify and understand lower level interactions between the senses.

²²This list is not exhaustive; there could be various forms of hybrid view in the area. I have difficulty imagining any hybrid view that was not consistent with the moderate pluralism I'm advocating here.

²³I have defended the idea that our best account of the individual modalities is that they are collections of sensory subsystems that function together to group or bind sets of sensory features together (Fulkerson, 2014a,b). This claim was made in the context of sensory pluralism: this is just *one* way to categorize the subsystems that make up vision. In many other respects, vision really is multisensory. On my view, it really is both, depending on what framework of investigation we are using.

processing levels, yet maintains that within this complexity there can be multiple robust explanatory systems of classification. We can, for instance, acknowledge in one explanatory context that all perception is inherently multisensory, while genuinely allowing in other contexts that some of our experiences are unisensory. My view is that what counts as a good classification of sensory interaction *partly* depends on the explanatory context. Hence “moderate sensory pluralism.”

In the next section, I want to highlight those general features of pluralist systems that ground legitimate schemes of classification, and support their explanatory utility.

5. THE CASE FOR PLURALISM

I put forward here some general claims in defense of the kind of moderate pluralism I have in mind. This will necessarily be a simplified discussion concerning various explanatory strategies we might take with respect to a domain. There are many discussions of pluralism in the literature, and the basic tenants are well understood. As Mitchell (2002, p. 55) remarks, “The ‘fact’ of pluralism in science is no surprise. On scanning contemporary journals, books, and conference topics in some sciences, one is struck by the multiplicity of models, theoretical approaches, and explanations.” This seems especially true of cognitive science (Dale, 2008; Dale et al., 2009). And, I shall argue, it is also the way we ought to be thinking about sensory interactions.

There are some general formal features that any complex system subject to moderate pluralism should exhibit. These are *decomposition*, *functional overlap*, and *bounded recombination*. We can find analogs of these features in something as simple as a Necker cube (see Figure 2).

The Necker Cube is a basic, simplified model of the kind of pluralist view I have in mind: it has constituent parts (decomposition). These parts in turn equally satisfy two inconsistent high-level descriptions, and they do so because any particular part

(a line or intersection) can play more than one role (functional overlap). There are also limits to the shapes that the cube can take on (bounded recombination); while there are multiple ways of seeing the cube, these ways are highly constrained²⁴.

Look at the point in the upper corner of the image picked out by the arrow. According to one high-level description, the point is a *front-facing* top corner of a cube. According to the other, it is a *rear-facing* point on the bottom corner of the cube. Which is correct? Well, the natural answer is *both*, depending on how the cube is seen. The parts themselves don’t settle the answer since they are consistent with both views. The lines and points on the page satisfy two distinct high-level descriptions. These are highly constrained descriptions: there are only two of them for these points, and they are very precise. In fact, fixing the high-level description completely fixes the role played by these elements. In context, there are correct and incorrect descriptions of these features. Many options for this point are completely ruled out: this particular point cannot be a rear-facing top corner. Nor can it stop being a point, and so on.

In this very basic example we can see the features of a moderate form of pluralism, one without problematic metaphysical commitments. Let us take these simple lessons and formalize them a bit for proper application to the sensory domain.

5.1. DECOMPOSITION

Let’s start with decomposition. It is not enough that a system has parts, but that it has *functionally salient* parts. This means that the system must have some *functional* decomposition. I am being intentionally broad about my use of “function” here. The parts of sensory systems I’m interested in play different causal, informational, mechanical, computational, and structural roles. As might be suspected from this list, I’m not here interested in, and nothing in my view hangs on, defending a particular theoretical account of *function*²⁵. The only constraint is that the relevant notion of function not be sensitive to our interpretive or conventional uses. There must be some actual intrinsic basis on which the low-level functions are assigned²⁶. I have in mind something like a “natural” or “proper” function, though very broadly construed (for a discussion of function in this sense, see Dretske, 1991 and Millikan, 1989).

A system is functionally decompositional—in the broad sense of function alluded to above—just in case its operation can be broken down into simpler parts that operate separately from the

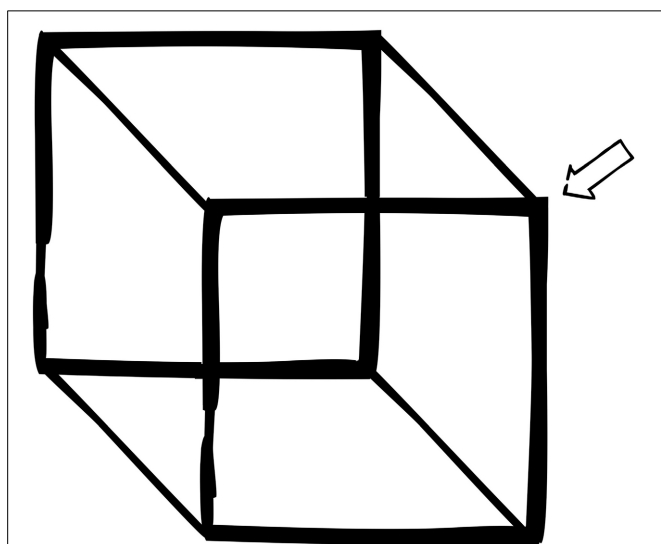


FIGURE 2 | Necker cube.

²⁴I’m using the Necker cube here as a basic model to point out those features of sensory systems that make them amenable to a pluralist view. Obviously, the analogy is not perfect. In particular, the cube lacks the required functional complexity, and its competing views are not sensitive to the *explanatory* context. Still, I think it is a useful toy example for clarifying the view I actually have in mind. I am grateful to an anonymous referee for pressing me to clarify this point.

²⁵There is, of course, a very large literature on the nature of function and mechanism in the special sciences (especially cognitive science and biology). For a very brief introduction, see Bechtel (2008); Cummins (1985); Machamer et al. (2000); Feest (2003); Millikan (1989)

²⁶Consider again the Necker cube. While its high level description depends on the visual context, the status of the lines and points on the page do not depend on the context. Each cube can be broken down into distinct functional complexes (corners and sides and edges, etc.).

other constituent parts²⁷. In other words, the complex system needs to be composed of parts that have a specifiable functional identity. A simple feature, like a point in space, or a complex object with lots of parts that each play no distinctive functional role, does not admit of the kind of moderate pluralism I have in mind. A hunk of iron for instance, is not subject to *this* sort of pluralism. It is surely a natural kind, one that can be used to do lots of different things²⁸. There are versions of pluralism that would apply to hunks of iron (Havstad, 2014), but for my purposes it doesn't count because its parts and their functions are too simple to admit of multiple appropriate schemes of scientific classification.

This should not be surprising: work in the metaphysics of mind has long recognized that psychological systems require a minimal level of functional complexity²⁹. Some entities simply do not have the structure or complexity necessary to allow for equally robust categorization at high levels³⁰.

Such decomposition is a necessary condition, but not a sufficient one. Just because something can be broken down into functional parts, it does not follow that we should be pluralists about its high-level nature. Another feature is needed, and that is functional overlap.

5.2. FUNCTIONAL OVERLAP

A pluralist friendly system must have a minimal level of complexity. In addition, however, the relationship between the constituent parts is also important. In particular, those functional parts should each contribute to distinct high-level systems.

It's unlikely that a system that had single-function parts (or more accurately, parts that contributed only to a single more complex system) could generate any interesting form of pluralism. This is simply because there would be only a single role played by that part, and so only a limited number of ways to reconceive its role in the larger system. Think about the cutting wheel on a can opener. It is an essential constituent part of the opener that plays a specialized functional role. So we have decomposition. But there's only the one role for the cutting wheel to play. It doesn't serve any other purpose, or contribute in different ways to different complexes at higher levels of the

can-opener. It's therefore difficult to suggest that our understanding and classification of the can opener varies in any way with the explanatory context³¹. Typically speaking, it doesn't. Only when the parts start to take on multiple roles will we start to see interesting ways of combining them into higher level systems.

Think again about the toy example of the Necker cube. Each point on the page functions both as a facing side and a rear side, depending on the view taken. That single point on the page (or the screen) plays *both* roles. If there were not some functional overlap, then there would not be multiple perspectives available on the cube as a whole. The parts serve dual functions, and we are able to see them play these different roles in distinct high-level structures. What role it plays thus depends on which high-level structure we're interested in (more on this soon).

5.3. LIMITED RECOMBINATION

The final feature is limited recombination. We can think of this as an upper bound on our pluralism. There are many ways of being a sensory system, on my view. And so our theoretical accounts of such systems are open to several distinct systems of classification. And yet, despite the existence of multiple forms of sensory interaction and methods of classification, these are highly constrained. There are clear limits on the roles played by the constituent parts and on the larger systems in which they participate. The view is not "anything goes." This is what makes this form of pluralism *moderate*. There are clear objective constraints that ground the admissible conceptions of the constituent systems³².

While one can look at the marks of the Necker cube on the page and see distinct but legitimate shapes represented, one cannot find spheres or other shapes in the mix. The objective locations of the points and lines rule out most shape interpretations. So there are clear constraints on how we understand this relatively simple system. Not only does this make the view ontologically moderate, it is also what makes the multiple views explanatory. The claim is not that sensory systems can be understood however we like, or that there are not facts of the matter concerning the natures of sensory interactions. Instead, the idea is that there are multiple, objectively robust roles played by the constituent elements of sensory systems, and so for any particular constituent subsystem, there will be more than one role it plays in distinct higher-level systems, but not an unlimited number of such roles. As we've seen, this perfectly describes the peripheral thermoreceptor system. This system contributes to pain awareness and is thus part of the nociceptive system. It also contributes to object recognition and externally-directed thermal awareness, and so it is also an essential part of the sense of touch. It also plays an important role (along with central thermoreceptors) in the regulation of body temperature, and so is part of our

²⁷In Fulkerson (2011) I discuss a version of this view, which I called the *functional dissociation criterion*. The notion I'm using here is broader than what I had in mind previously.

²⁸Perhaps this is a good place to note that what I have in mind differs significantly from the notion of *multiple realizability* in the cognitive sciences. The idea here isn't that the "same high level description" can be realized by different underlying constituents as in multiple realization, but that the very same set of underlying constituents can be parts of equally salient, but distinct high-level systems. My view is thus something closer to the converse of multiple realizability.

²⁹See for instance, see the discussion in Block (1997) concerning the "Disney Principle," the idea that in the real world, anything with a mind needs to have a minimal level of complexity (unlike the sentient teapots and spoons in the world of cartoons). These debates arise in several domains, involving debates about reduction, emergence, and the relationship between low-level realizers and high-level descriptions (Batterman, 2000).

³⁰It is of course my view, given the above, that there are several distinct notions of pluralism and that we need to be careful not to assume that they are equivalent.

³¹Of course, this is a bit simplified. Maybe for some explanatory purposes the material composition of the cutting wheel matters (why did it rust?), for others it might be its size or shape (why won't it cut this can?). These points suggest that pluralism is a robust phenomenon throughout our explanatory practices. The present point is that, above and beyond these basic forms of explanatory pluralism, sensory systems exhibit an additional layer of complexity.

³²Compare the *sorting* and *motivating* principles discussed by Ereshefsky (1992).

homeostatic regulatory system (along with thirst and hunger). There is no single classification of these peripheral thermoreceptor populations, because they play many varied roles in our lives.

We can now see that certain complex systems are amenable to a modest pluralist view. To suggest that a single complex system can be understood and classified in multiple ways does not commit us to a problematic ontology. We can make this clear by making explicit the contextual operator in our sensory classification. A claim such as “System X is multisensory” leaves out this operator, and thus cannot be properly evaluated. It is not an explanatory statement. Instead, we should be evaluating claims of the form “System X is multisensory according to explanatory schema Y.” This schema specifies the respect in which something counts as multisensory or not. Similarly for other claims.

Adding this sort of clause will allow researchers to avoid mere terminological disputes and help clarify the nature of the investigation in question. One worry about pluralist views is that they can foster confusion and hinder scientific progress. I find compelling the reply in Ereshefsky (1992, p. 680) to such worries about pluralistic views of species in biology:

[B]iologists should categorize those lineages by the criteria used to segment them: interbreeding units, monophyletic units, and ecological units. The term “species” is superfluous beyond the reference to a segmentation criterion; and when the term is used alone it leads to confusion. The term “species” has out-lived its usefulness and should be replaced by terms that more accurately describe the different types of lineages that biologists refer to as “species.”

Similarly, philosophers and others talking about sensory experience should avoid using terms like “multisensory,” “multi-modal,” or “cross-modal” without being clear about the way in which they are using those terms. They should not assume that there is a single, theoretically interesting way in which senses interact, or that we can have, say, a single unified account of what qualifies as a “cross-modal” form of interaction. Some interactions are legitimately unisensory, others involve activations of processing units distributed widely in other systems (and these often overlap!). There is thus no single way for these systems to interact; they are complexes that interact in many theoretically interesting ways.

Putting all of these elements together, we can see that sensory systems should be ideally situated to the kind of pluralistic view I’ve been outlining. They are, after all, evolved biological systems that serve many functions, and are subject to many constraints. Indeed, all of cognition seems amenable to this perspective. As Dale et al. (2009) write: “The mind, as somehow constituted by brain/body/environment interaction, is extraordinarily complex. In addition, we have many and assorted interests in that interaction” (p. 1). And these parts play these roles in a number of ways, through informational extraction and computation, through behavioral and bodily features and reactions, and so on.

6. SENSORY SYSTEMS

It should be clear from the gloss above that sensory systems are ideal candidates to satisfy all three constraints. If we’ve learned anything over the last few decades, it’s that our sensory systems are deeply complex structures that involve a large number of interacting elements. Very often these elements are put to different uses by various downstream systems. As such, sensory systems are decomposable into functionally salient parts. These parts (rods, cones, retinal ganglion cells, etc.) in turn perform different functions depending on which downstream systems they are contributing to³³. And so it should not be surprising that the explanatory context—the kinds of systems we’re investigating and what behaviors and capacities we seek to explain—can have a significant impact on how the various systems are categorized and understood.

Even entirely within vision this should be clear. A cone cell examined in isolation performs one function (converting electromagnetic energy into neural signals), but the function it serves can be influenced by, and in turn influence, other cones connected to it (for instance, when detecting edges). At higher levels of complexity, these same constituent elements can perform many other functions. For instance, these early visual elements are essential parts of a complex object-recognition system, but also play a role in guiding our motor actions. Recent debates about the “two visual streams hypothesis” arise partly because of these dual roles (Milner and Goodale, 1995). Which stream is *really* vision? Various options are available here, but taking the pluralist conception, one can see that we shouldn’t expect a single answer. What we call “vision” is in reality a complex set of distinct systems and subsystems. There are *many* things that count as vision (this is the pluralism). Which one is going to be explanatory and relevant for scientific purposes depends on making clear the explanatory context. Even so, it does not depend *entirely* on the context; there are clear objective constraints limiting the ways we can think about visual experiences. The view is heavily grounded in the actual capacities and functions of the constituent elements of the system.

While it strengthens the claims I’ll be making that they mesh with actual practice in the cognitive sciences, the case for sensory pluralism doesn’t rest *entirely* on this descriptive enterprise. It is neither necessary nor sufficient for the truth of sensory pluralism that researchers engage in these strategies of classification (they could simply be mistaken in their current practices). My purpose also is similarly not to weigh in on or take sides on these first-order debates. Instead, the discussion is meant to show how fruitful, plausible, and powerful the pluralist perspective can be in helping further our understanding of difficult issues in recent work on perceptual experience. Moderate sensory pluralism is, ideally, a form of what Mitchell (2002) calls “compatible pluralism.” On this view, the various explanations involved are not strict competitors, but mutually supporting accounts of complex phenomena:

³³The notion of “downstream” systems can be deeply misleading. Sensory systems, like much of cognition, is deeply heterarchical, and involves processing going in multiple directions at the same time. These complications only add additional support to the claims made here.

[C]omplex phenomenon harbor multiple interacting causal processes and multiple levels of organization which all may be involved in the generation of the feature to be explained. By disambiguating the question to be answered by an explanation—i.e., what is the evolutionary origin of a trait or behavior we observe now—one is still left with a plurality of potential causes acting at a number of levels of organization which may well constitute compatible answers to that single question (Mitchell, 2002, p. 57)

We have seen how this perspective enriches our understanding of the thermoreceptive system. Let us see how it might apply to other recent debates in the literature³⁴.

7. IMPLICATIONS

I will now briefly discuss some potential applications of the account described here in several domains of active research on sensory awareness.

7.1. OLFACTION

The olfactory system is another obvious case where sensory pluralism finds ample support. Intuitively, we believe that we have a single “sense of smell” and that we can understand the components of this system as a single, coherent thing. The reality is a bit more complicated. We in fact have two senses of smell, a orthonasal and a retronasal system. The orthonasal system involves molecules that are picked up in the surrounding environment through the nasal cavity, often by exploratory acts of sniffing (Wilson and Stevenson, 2006). These inputs provide reliable information about the nature of environmental chemical stimulants (See also Batty, 2009). We can even use smell for wayfinding and to help influence our emotional reactions (Herz, 2007; Rosenblum, 2011).

Retronasal olfaction by contrast involves chemical irritants that rise from inside the mouth and pass through the olfactory epithelium from the other direction. Though the initial activation sites are more or less the same in both instances, the resulting perceptual experiences and functional interactions are very different from orthonasal ones. Here the smell becomes fused and combined with other taste information and generates a complex experience of flavor (Auvray and Spence, 2008). So is smell a unified sensory modality? Is it externally directed? Or is it part of a multisensory system of flavor detection? According to the sensory pluralist, the answer is all of the above. The initial chemoreceptors involved in both systems might be the same, but they play very different roles when combined with distinct inputs (external vs internal sources of chemical irritants) and co-processing elements (sniffing and head movements in externally directed tasks and coordinated taste and texture activations in the mouth, respectively). Here again we see that what seems like a single modality is really a complex collection of interacting elements that can be appropriately classified in a variety of ways.

³⁴The pluralist view offers a robust explanation for the prevalence of such debates. More importantly, it suggests a way to move forward on such debates, as I hope will become clear in each example.

7.2. AUDITORY PROCESSORS

The auditory system also admits of several distinct schemes of classification: the initial processing units involved in auditory experience play a role in several interacting systems: general sound perception, our awareness of speech, and in the perception of music. There are reasons for thinking of these as very different systems, and thus there are multiple ways of classifying and accounting for our auditory awareness; these ways involve different functional roles, associated interactions with other systems, and behavioral capacities. In addition, of course, we can understand audition as part of larger networks connected to causal detection (as in the motion bounce illusion, Sekuler et al., 1997). All of these forms of classification are robust and explanatory, and often involve the same initial processing units and transducer populations. We shouldn't expect a single, unified account of audition. Like thermoreception and smell, it involves a range of capacities that admit of distinct forms of classification.

7.3. SYNESTHESIA

Synesthesia is another interesting case for understanding sensory pluralism. This condition involves (roughly) the reliable activation of one modality by stimuli presented to another. In this way, it seems to be a kind of cross-modal interaction, but one importantly different from typical cases of multisensory integration or facilitation. In addition, it poses a number of basic definitional and phenomenological questions. Researchers have long known, for instance, that synesthesia comes in a variety of forms, and it is difficult to find a single account that covers all (and only) genuine cases (Macpherson, 2007; Mroczko-Wasowicz and Werning, 2012)³⁵. Given the difficulties in presenting a robust, unified account of synesthesia, we should not be surprised if it turns out that there are multiple forms of the condition, each distinctive in various ways. The pluralist perspective suggests that we should not (simply) hold out for a single mechanism underlying the overall condition, but explore the possibility that the condition arises in a variety of distinctive ways. One could even allow that so-called normal subjects might exhibit features continuous with the possession of synesthesia (see Auvray and Deroy, *in press*; Cohen, *in press*). The question of what counts and what doesn't count as synesthesia in general might not be a well-formed question. Maybe synesthesia isn't a natural kind at all?³⁶

The sensory pluralist can allow that, in some respects, many cases of synesthesia are extensions of ordinary perceptual capacities, part of the same functional units that underlie our general experience of the world. On the other hand, from a slightly different explanatory context, we can see discontinuities as well. In addition, some forms of the condition might be more strongly connected with one context, and might exploit resources typical in ordinary perceptual interaction, whereas others might involve

³⁵See also Simner (2012a) target paper, commentaries by Eagleman (2012) and Cohen Kadosh and Terhune (2012), with a reply by Simner (2012b).

³⁶Gray (2001) makes the case that synesthesia poses problems for Fodorian modularity, and should be taken seriously in any account of psychological kinds. As one might expect, I agree with this assessment, and suggest in addition that facts about synesthesia also support a version of sensory pluralism.

interactions more difficult to reconcile with typical sensory interactions. There is no reason to choose sides here (at least not yet); once we make clear the explanatory context of our investigations, and the precise nature of the interactions under investigation, we can make clear the sense in which these phenomena are like and unlike other forms of sensory awareness. What this involves, as in the other cases I've discussed, is making clear the explanatory context and embracing the idea that there may be multiple useful ways of investigating and theorizing about these interactions³⁷.

7.4. SENSORY SUBSTITUTION

There is one final area of intensive investigation that would benefit from the pluralist perspective. Sensory substitution and enhancement devices pose many challenges for traditional monist accounts of sensory individuation. Such devices provide input usually provided by one modality through a device that interacts with a different modality. For instance, a camera might be used to provide inputs to touch for a subject without normal sight. If a subject is presented with visual information through a camera system that translated those signals into a vibrating plate on the tongue, does the resulting experience count as visual or tactual? There have been many discussions about such devices³⁸.

The pluralist view suggests that these devices ought to admit of distinct forms of classification. They pose such a difficulty because they often have characteristics from both modalities. If we focus on behavioral capacities we might classify the experiences generated by the device one way; if we focus instead on phenomenal character we might classify it differently. Enhancement systems might produce novel forms of awareness that don't fit into any of our current schemes of classification. They also suggest that the focus of this discussion—the multiple roles that our low level biological machinery can play—might be too narrow. Sensory enhancement and substitution might reveal that our sensory capacities outstrip the present functions of our hardware³⁹.

8. SUMMING UP

The main alternative view to pluralism would be some form of monism: the idea that a single scheme of classification should define each of the sensory modalities, and their interactions. But it should be clear from the preceding that it seems highly unlikely that we will find a unified criteria for defining each of the senses. Vision differs from the other senses in a multitude of ways, and plays many distinct roles at different levels of sensory processing. What single account of modality or interaction can capture that diversity, and then work equally well for audition, proprioception, touch, taste, vestibular awareness, sensory dominance, facilitation, suppression, and cross-modal blends (like flavor)?

Others might worry that I've left the details here are a little spare. That is intentional. I do not wish to commit myself to any particular account of scientific explanation here. If one takes a mechanistic or functional explanation as ideal for work in cognitive science, then what I say here suggests that we can (and should) focus on a diversity of functional explanations when it comes to the senses and their interactions. If one prefers a different explanatory framework (a computational or informational story, say), then my claims here should motivate us to look for a diversity of computational processes involved in the generation of sensory experience.

Nothing that I've said requires us to take a stand on intertheoretic relations, reductionism, emergence, or explanation. At no point do I claim that there are sensory systems that can or cannot be reduced to lower level functional or computational components. The claim is that, when it comes to sensory systems, we should expect distinct explanatory accounts to be available (cf. Dale et al., 2009). The only substantive commitment I make is that each system of classification be genuinely explanatory, and grounded in the objective basic features of the system. In this sense, it is a genuine ontological pluralism (cf. Ereshefsky, 1992), but a moderate one. My claim is not that we cannot know what senses "really are." It is that, as a matter of fact, senses really are lots of things, and what counts as explanatory in our theorizing about sensory interactions depends on how we're carving the systems up and what we are trying to explain. So while my point is distinct from claims about multiple realizability and about levels of explanation in the cognitive sciences (see Marr, 1982; Dennett, 1989), the view is both compatible with and offered in the spirit of these views.

As we've seen, there has been a lot of work recently on understanding the nature of multisensory awareness. Arguments abound concerning whether we need to completely reject our prior conceptions of sensory modalities and their interactions, or whether we can salvage some aspects of sensory unity and cohesion. The sensory pluralist view doesn't, in itself, settle these debates. But it does suggest that many of these debates are merely verbal disputes, where the contexts of investigation and explanation have not been clarified. There need be no debate, for instance, between those who think of thermoreception as continuous with pain and other interoceptive systems, and those who investigate the role of thermoreception in object recognition and sensory exploration. There should be no disputes between views on which flavor awareness forms a separate modality or not. In some explanatory contexts it most certainly does; in others it need not. The pluralist view doesn't give up on the idea of correct scientific theorizing, it just makes clear something that already is the case: sensory systems and their interactions are complex, multifaceted, and occur at many levels of processing. Our theorizing about these interactions needs to recognize and take on board these complexities⁴⁰.

³⁷I should emphasize, one final time, that nothing I've said here is meant to rule out genuine disagreements; there can and will be false accounts of the phenomena that can be definitely ruled out even if sensory pluralism is true.

³⁸For a representative discussion, see (among others): Deroy and Auvray (2014); Noë (2004); Auvray and Myin (2009); Deroy (2012); Farina (2013); Rita and Kercel (2003); Froese et al. (2012).

³⁹See Clark (2003) for discussion along these lines.

⁴⁰I would like to thank the editor and referees for an extremely helpful set of comments on earlier versions of this paper. I would also like to thank the audience and participants at the Network for Sensory Research Workshop on Multisensory Perception held at the University of Toronto where an early version of this material was presented.

REFERENCES

- Ackerley, R., Olsson, H., Wessberg, J., and McGlone, F. (2012). Wetness perception across body sites. *Neurosci. Lett.* 522, 73–77. doi: 10.1016/j.neulet.2012.06.020
- Akins, K. (1996). Of sensory systems and the “aboutness” of mental states. *J. Philos.* 93, 337–372. doi: 10.2307/2941125
- Auvray, M., and Deroy, O. (in press). “How do synesthetes experience the world,” in *Oxford Handbook of Philosophy of Perception*, ed M. Matthen (New York, NY: Oxford University Press).
- Auvray, M., and Myin, E. (2009). Perception with compensatory devices: from sensory substitution to sensorimotor extension. *Cogn. Sci.* 33, 1036–1058. doi: 10.1111/j.1551-6709.2009.01040.x
- Auvray, M., and Spence, C. (2008). The multisensory perception of flavor. *Conscious. Cogn.* 17, 1016–1031. doi: 10.1016/j.concog.2007.06.005
- Aydede, M. (2009). Is feeling pain the perception of something? *J. Philos.* 106, 531–567.
- Aydede, M., and Guzeldere, G. (2002). Some foundational problems in the scientific study of pain. *Philos. Sci.* 69, S265–S283. doi: 10.1086/341851
- Batterman, R. W. (2000). Multiple realizability and universality. *Br. J. Philos. Sci.* 51, 115–145. doi: 10.1093/bjps/51.1.115
- Batty, C. (2009). What’s that smell? *South. J. Philos.* 47, 321–348. doi: 10.1111/j.2041-6962.2009.tb00164.x
- Batty, C. (2010). Scents and sensibilia. *Am. Philos. Q.* 47, 103.
- Bechtel, W. (2008). *Mental Mechanisms. Philosophical Perspectives on Cognitive Neuroscience*. New York, NY: Psychology Press.
- Bennett, D., and Hill, C. (2014). *Sensory Integration and the Unity of Consciousness*. Cambridge, MA: MIT Press.
- Block, N. (1997). Anti-reductionism slaps back. *Philos. Perspect.* 31, 107–132.
- Campos, J. L., Butler, J. S., and Bühlhoff, H. H. (2012). Multisensory integration in the estimation of walked distances. *Exp. Brain Res.* 218, 551–565. doi: 10.1007/s00221-012-3048-1
- Cartwright, N. (1999). *The Dappled World. A Study of the Boundaries of Science*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139167093
- Caterina, M. J., Schumacher, M. A., Tominaga, M., Rosen, T. A., Levine, J. D., and Julius, D. (1997). The capsaicin receptor: a heat-activated ion channel in the pain pathway. *Nature* 389, 816–824. doi: 10.1038/39807
- Chapman, C. E. (1994). Active versus passive touch: factors influencing the transmission of somatosensory signals to primary somatosensory cortex. *Can. J. Physiol. Pharmacol.* 72, 558–570. doi: 10.1139/y94-080
- Clark, A. (2003). *Natural-Born Cyborgs*. New York; Oxford: Oxford University Press.
- Cleland, C. E. (2013). Pluralism or unity in biology: could microbes hold the secret to life? *Biol. Philos.* 28, 189–204. doi: 10.1007/s10539-013-9361-7
- Cohen, J. (in press). “Synesthetic perception as continuous with ordinary perception, or: we’re all synesthetes now,” in *Sensory Blendings: New Essays on Synaesthesia*, ed O. Deroy (New York, NY: Oxford University Press), 1–28.
- Cohen Kadosh, R., and Terhune, D. B. (2012). Redefining synaesthesia? *Br. J. Psychol.* 103, 20–23. doi: 10.1111/j.2044-8295.2010.02003.x
- Connolly, K. (2014). “Making sense of multiple senses,” in *Consciousness Inside and Out: Phenomenology, Neuroscience, and the Nature of Experience*, ed R. Brown (Dordrecht: Springer), 351–364.
- Corns, J. (2014). The inadequacy of unitary characterizations of pain. *Philos. Stud.* 169, 355–378. doi: 10.1007/s11098-013-0186-7
- Craig, A. D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. *Nat. Rev. Neurosci.* 3, 655–666. doi: 10.1038/nrn894
- Cummins, R. (1985). *The Nature of Psychological Explanation*. Cambridge, MA: MIT Press.
- Dale, R. (2008). The possibility of a pluralist cognitive science. *J. Exp. Theor. Artif. Intell.* 20, 155–179. doi: 10.1080/09528130802319078
- Dale, R., Dietrich, E., and Chemero, A. (2009). Explanatory pluralism in cognitive science. *Cogn. Sci.* 33, 739–742. doi: 10.1111/j.1551-6709.2009.01042.x
- de Vignemont, F. (2007). Habeas corpus: the sense of ownership of one’s own body. *Mind Lang.* 22, 427–449. doi: 10.1111/j.1468-0017.2007.00315.x
- Dennett, D. C. (1989). *The Intentional Stance*. Cambridge, MA: MIT.
- Deroy, M. A. O. (2012). Reading the world through the skin and ears: a new perspective on sensory substitution. *Front. Psychol.* 3:457. doi: 10.3389/fpsyg.2012.00457
- Deroy, O., and Auvray, M. (2014). “A crossmodal perspective on sensory substitution,” in *Perception and Its Modalities*, eds S. Biggs, M. Matthen, and D. Stokes (Oxford: Oxford University Press), 327–349.
- Deroy, O., Chen, Y. C., and Spence, C. (2014). Multisensory constraints on awareness. *Philos. Trans. R. Soc. B Biol. Sci.* 369:20130207. doi: 10.1098/rstb.2013.0207
- Dretske, F. I. (1991). *Explaining Behavior. Reasons in a World of Causes*. Cambridge, MA: MIT Press.
- Driver, J., and Spence, C. (2000). Multisensory perception: beyond modularity and convergence. *Curr. Biol.* 10, 731–735. doi: 10.1016/S0960-9822(00)00740-5
- Eagleman, D. M. (2012). Synaesthesia in its protean guises. *Br. J. Psychol.* 103, 16–19. doi: 10.1111/j.2044-8295.2011.02020.x
- Ereshefsky, M. (1992). Eliminative pluralism. *Philos. Sci.* 59, 671–690. doi: 10.1086/289701
- Evans, G. (1982). *The Varieties of Reference*. Oxford: Oxford University Press.
- Farina, M. (2013). Neither touch nor vision: sensory substitution as artificial synaesthesia? *Biol. Philos.* 28, 639–655. doi: 10.1007/s10539-013-9377-z
- Feest, U. (2003). Functional analysis and the autonomy of psychology. *Philos. Sci.* 70, 937–948. doi: 10.1086/377379
- Fodor, J. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.
- Frissen, I., Campos, J. L., Souman, J. L., and Ernst, M. O. (2011). Integration of vestibular and proprioceptive signals for spatial updating. *Exp. Brain Res.* 212, 163–176. doi: 10.1007/s00221-011-2717-9
- Froese, T., McGann, M., Bigge, W., Spiers, A., and Seth, A. K. (2012). The enactive torch: a new tool for the science of perception. *IEEE Trans. Haptics* 5, 365–375. doi: 10.1109/TOH.2011.57
- Fulkerson, M. (2011). The unity of haptic touch. *Philos. Psychol.* 24, 493–516. doi: 10.1080/09515089.2011.556610
- Fulkerson, M. (2014a). “Explaining multisensory experience,” in *Consciousness Inside and Out: Phenomenology, Neuroscience, and the Nature of Experience*, ed R. Brown (Dordrecht: Springer), 365–373.
- Fulkerson, M. (2014b). *The First Sense. A Philosophical Study of Human Touch*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/9780262019965.001.0001
- Ghazanfar, A. A., and Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends Cogn. Sci.* 10, 278–285. doi: 10.1016/j.tics.2006.04.008
- Gray, R. (2001). Cognitive modules, synaesthesia and the constitution of psychological natural kinds. *Philos. Psychol.* 14, 65–82. doi: 10.1080/09515080120033562
- Gray, R. (2005). On the concept of a sense. *Synthese* 147, 461–475. doi: 10.1007/s11229-005-1334-1
- Gray, R. (2012). What do our experiences of heat and cold represent? *Philos. Stud.* 166, 131–151. doi: 10.1007/s11098-012-0083-5
- Grush, R. (2005). “Brain time and phenomenological time,” in *Philosophy and the Neurosciences*, eds A. Brooks and K. Akins (Cambridge: Cambridge University Press), 160–207.
- Hammel, H. T., and Pierce, J. B. (2002). Regulation of internal body temperature. *Ann. Rev. Physiol.* 30, 641–710. doi: 10.1146/annurev.ph.30.030168.003233
- Havstad, J. (2014). *Biochemical Kinds and Selective Naturalism*. PhD thesis, University of California, San Diego, CA.
- Herz, R. (2007). *The Scent of Desire*. New York, NY: William Morrow.
- Herz, R. S., and von Clef, J. (2001). The influence of verbal labeling on the perception of odors: evidence for olfactory illusions? *Perception* 30, 381–391. doi: 10.1068/p3179
- Hurley, S. (1998). *Consciousness in Action*. Cambridge, MA: Harvard.
- Jones, L. A., and Lederman, S. J. (2006). *Human Hand Function*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780195173154.001.0001
- Keeley, B. (2002). Making sense of the senses: individuating modalities in humans and other animals. *J. Philos.* 99, 5–28. doi: 10.2307/3655759
- Kitcher, P. (1984). Species. *Philos. Sci.* 51, 308–333.
- Klatzky, R., and Lederman, S. (2008). “Object recognition by touch,” in *Blindness and Brain Plasticity in Navigation and Object Perception*, eds J. J. Rieser, D. H. Ashmead, F. F. Ebner, and L. L. Corn (New York, NY: Lawrence Erlbaum, Psychology Press), 185–207.
- Lee, G. (2014). Temporal experience and the temporal structure of experience. *Philosopher’s Imprint* 14, 1–21.
- Lumpkin, E. A., and Caterina, M. J. (2007). Mechanisms of sensory transduction in the skin. *Nature* 445, 858–865. doi: 10.1038/nature05662
- Machamer, P., Darden, L., and Craver, C. F. (2000). Thinking about mechanisms. *Philos. Sci.* 67, 1–25. doi: 10.1086/392759
- Macpherson, F. (2007). “Synaesthesia, functionalism and phenomenology,” in *Cartographies of the Mind*, eds M. Marraffa, F. Ferretti, and M. de Caro (Dordrecht: Kluwer; Springer), 65–80.

- Macpherson, F. (2010). Taxonomising the senses. *Philos. Stud.* 153, 123–142. doi: 10.1007/s11098-010-9643-8
- Macpherson, F. (2011). XV-cross-modal experiences. *Proc. Arist. Soc.* 111, 429–468. doi: 10.1111/j.1467-9264.2011.00317.x
- Mag Uidhir, C., and Magnus, P. D. (2011). Art concept pluralism. *Metaphilosophy* 42, 83–97. doi: 10.1111/j.1467-9973.2010.01678.x
- Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco, CA: W. H. Freeman.
- Martin, M. (1992). “Sight and touch,” in *The Contents of Experience*, ed T. Crane (Cambridge: Cambridge University Press), 196–216.
- Matthen, M. (2010). On the diversity of auditory objects. *Rev. Philos. Psychol.* 1, 63–89. doi: 10.1007/s13164-009-0018-z
- Matthen, M. (in press). “The individuation of the senses,” in *Oxford Handbook of the Philosophy of Perception*, ed M. Matthen (New York, NY: Oxford University Press).
- McGann, M. (2010). Perceptual modalities: modes of presentation or modes of interaction? *J. Conscious. Stud.* 17, 1–2.
- McGlone, F., Vallbo, A. B., Olsson, H., Loken, L., and Wessberg, J. (2007). Discriminative touch and emotional touch. *Can. J. Exp. Psychol.* 61, 173–183. doi: 10.1037/cjep.2007019
- McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748.
- Michotte, A. (1963). *The Perception of Causality*. New York, NY: Basic Books.
- Millikan, R. (1989). In defense of proper functions. *Philos. Sci.* 56, 288–302. doi: 10.1086/289488
- Milner, A. D., and Goodale, M. A. (1995). *The visual Brain in Action*. New York, NY: Oxford University Press.
- Mishler, B. D., and Brandon, R. N. (1987). Individuality, pluralism, and the phylogenetic species concept. *Biol. Philos.* 2, 397–414. doi: 10.1007/BF00127698
- Mitchell, S. D. (2002). Integrative pluralism. *Biol. Philos.* 17, 55–70. doi: 10.1023/A:1012990030867
- Mroczko-Wasowicz, A., and Nikolic, D. (2014). Semantic mechanisms may be responsible for developing synesthesia. *Front. Hum. Neurosci.* 8:509. doi: 10.3389/fnhum.2014.00509
- Mroczko-Wasowicz, A., and Werning, M. (2012). Synesthesia, sensory-motor contingency, and semantic emulation: how swimming style-color synesthesia challenges the traditional view of synesthesia. *Front. Psychol.* 3:279. doi: 10.3389/fpsyg.2012.00279
- Nakamura, K., and Morrison, S. F. (2007). A thermosensory pathway that controls body temperature. *Nat. Neurosci.* 11, 62–71. doi: 10.1038/nn2027
- Noë, A. (2004). *Action in Perception*. Cambridge, MA: MIT.
- Nudds, M. (2001). Experiencing the production of sounds. *Eur. J. Philos.* 9, 210–229. doi: 10.1111/1468-0378.00136
- Nudds, M. (2003). The significance of the senses. *Proc. Arist. Soc.* 104, 31–51.
- O’Callaghan, C. (2007). *Sounds: a Philosophical Theory*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780199215928.001.0001
- O’Callaghan, C. (2008). Seeing what you hear: cross-modal illusions and perception. *Philos. Issues* 18, 316–338. doi: 10.1111/j.1533-6077.2008.00150.x
- O’Callaghan, C. (2012). “Perception and multimodality,” in *Oxford Handbook of Philosophy of Cognitive Science*, eds E. Margolis, R. Samuels, and S. Stich (New York, NY: Oxford University Press), 73–91.
- O’Shaughnessy, B. (1989). The sense of touch. *Aust. J. Philos.* 67, 37–58. doi: 10.1080/00048408912343671
- Phillips, I. (2008). Perceiving temporal properties. *Eur. J. Philos.* 18, 176–202. doi: 10.1111/j.1468-0378.2008.00299.x
- Pylshyn, Z. (2006). *Seeing and Visualizing: It’s Not What You Think*. Cambridge, MA: MIT Press.
- Ratcliffe, M. (2008). Touch and situatedness. *Int. J. Philos. Stud.* 16, 299–322. doi: 10.1080/09672550802110827
- Richardson, L. (2011). Sniffing and smelling. *Philos. Stud.* 162, 401–419. doi: 10.1007/s11098-011-9774-6
- Richardson, L. (2013). Bodily sensation and tactile perception. *Philos. Phenomenol. Res.* 86, 134–154. doi: 10.1111/j.1933-1592.2011.00504.x
- Rita, P. B.-Y., and Kerckel, S. W. (2003). Sensory substitution and the human-machine interface. *Trends Cogn. Sci.* 7, 541–546. doi: 10.1016/j.tics.2003.10.013
- Rosenblum, L. D. (2011). *See What I’m Saying: The Extraordinary Powers of Our Five Senses*. New York, NY: W. W. Norton & Company.
- Schepers, R. J., and Ringkamp, M. (2010). Thermoreceptors and thermosensitive afferents. *Neurosci. Biobehav. Rev.* 34, 177–184. doi: 10.1016/j.neubiorev.2009.10.003
- Schwenkler, J. (2011). The objects of bodily awareness. *Philos. Stud.* 162, 465–472. doi: 10.1007/s11098-011-9777-3
- Scott, M. (2001). Tactual perception. *Aust. J. Philos.* 79, 149–160. doi: 10.1080/713931200
- Sekuler, R., Sekuler, A. B., and Lau, R. (1997). Sound alters visual motion perception. *Nature* 385, 308–308. doi: 10.1038/385308a0
- Shimojo, S., and Shams, L. (2001). Sensory modalities are not separate modalities: plasticity and interactions. *Curr. Opin. Neurobiol.* 11, 505–509. doi: 10.1016/S0959-4388(00)00241-5
- Siegel, S. (2011). Cognitive penetrability and perceptual justification*. *Nous* 46, 201–222. doi: 10.1111/j.1468-0068.2010.00786.x
- Simner, J. (2012a). Defining synaesthesia. *Br. J. Psychol.* 103, 1–15. doi: 10.1348/000712610X528305
- Simner, J. (2012b). Defining synaesthesia: a response to two excellent commentaries. *Br. J. Psychol.* 103, 24–27. doi: 10.1111/j.2044-8295.2011.02059.x
- Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., and Small, S. L. (2007). Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cereb. Cortex* 17, 2387–2399. doi: 10.1093/cercor/bhl147
- Smith, B. (2009). “Taste, philosophical perspectives,” in *Encyclopedia of the Mind*, ed H. Pashler (Thousand Oaks, CA: Sage Publications, Inc.), 731–734.
- Smith, B. (2012). Perspective: complexities of flavour. *Nature* 486, S6. doi: 10.1038/486S6a
- Steel, D. (2004). Can a reductionist be a pluralist? *Biol. Philos.* 19, 55–73. doi: 10.1023/B:BIPH.0000013263.35121.5a
- Stein, B. E. (2012). *The New Handbook of Multisensory Processes*. Cambridge, MA: MIT Press.
- Stokes, D. (2012). Perceiving and desiring: a new look at the cognitive penetrability of experience. *Philos. Stud. Int. J. Philos. Anal. Tradit.* 158, 477–492. doi: 10.1007/s11098-010-9688-8
- Sullivan, A. (1923). The perceptions of liquidity, semi-liquidity and solidity. *Am. J. Psychol.* 34, 531–541. doi: 10.2307/1414055
- Thompson, E. (2007). *Mind in life*. Cambridge, MA: Harvard.
- Villemure, C., Slotnick, B. M., and Bushnell, M. C. (2003). Effects of odors on pain perception: deciphering the roles of emotion and attention. *Pain* 106, 101–108. doi: 10.1016/S0304-3959(03)00297-5
- Vuilleumier, P., and Driver, J. (2007). Modulation of visual processing by attention and emotion: windows on causal interactions between human brain regions. *Philos. Trans. R. Soc. B Biol. Sci.* 362, 837–855. doi: 10.1098/rstb.2007.2092
- Welsh, M. J. (2001). Biochemical basis of touch perception: mechanosensory function of degenerin/Epithelial Na⁺ channels. *J. Biol. Chem.* 277, 2369–2372. doi: 10.1074/jbc.R100060200
- Wenger, C. B. (1995). “The regulation of body temperature,” in *Medical Physiology*, ed R. Rodney (Boston, MA: Little, Brown & Co), 527–550.
- Wilson, D. A., and Stevenson, R. J. (2006). *Learning to Smell: Olfactory Perception from Neurobiology to Behavior*. Baltimore, MD: Johns Hopkins University Press.

Conflict of Interest Statement: The author declares 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 September 2014; accepted: 22 November 2014; published online: 10 December 2014.

Citation: Fulkerson M (2014) Rethinking the senses and their interactions: the case for sensory pluralism. *Front. Psychol.* 5:1426. doi: 10.3389/fpsyg.2014.01426

This article was submitted to *Consciousness Research*, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Fulkerson. 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.



Multisensory perception as an associative learning process

Kevin Connolly*

Philosophy and Institute for Research in Cognitive Science, University of Pennsylvania, Philadelphia, PA, USA

Edited by:

Aleksandra Mroczko-Wasowicz,
National Yang Ming University, Taiwan

Reviewed by:

Nicholas Altieri, Idaho State
University, USA

Casey O'Callaghan, Washington
University in St. Louis, USA

*Correspondence:

Kevin Connolly, Philosophy and
Institute for Research in Cognitive
Science, University of Pennsylvania,
Philadelphia, PA, USA
e-mail: connok@sas.upenn.edu

Suppose that you are at a live jazz show. The drummer begins a solo. You see the cymbal jolt and you hear the clang. But in addition seeing the cymbal jolt and hearing the clang, you are also aware that the jolt and the clang are part of the same event. Casey O'Callaghan (forthcoming) calls this awareness "intermodal feature binding awareness." Psychologists have long assumed that multimodal perceptions such as this one are the result of an automatic feature binding mechanism (see Pourtois et al., 2000; Vatakis and Spence, 2007; Navarra et al., 2012). I present new evidence against this. I argue that there is no automatic feature binding mechanism that couples features like the jolt and the clang together. Instead, when you experience the jolt and the clang as part of the same event, this is the result of an associative learning process. The cymbal's jolt and the clang are best understood as a single learned perceptual unit, rather than as automatically bound. I outline the specific learning process in perception called "unitization," whereby we come to "chunk" the world into multimodal units. Unitization has never before been applied to multimodal cases. Yet I argue that this learning process can do the same work that intermodal binding would do, and that this issue has important philosophical implications. Specifically, whether we take multimodal cases to involve a binding mechanism or an associative process will have impact on philosophical issues from Molyneux's question to the question of how active or passive we consider perception to be.

Keywords: perceptual learning, crossmodal integration, feature binding, multimodal interaction, crossmodal interaction, binding, multisensory integration, associative learning

INTRODUCTION

Suppose that you are at a live jazz show. The drummer begins a solo. You see the cymbal jolt and you hear the clang. But in addition seeing the cymbal jolt and hearing the clang, you are also aware that the jolt and the clang are part of the same event. Casey O'Callaghan (forthcoming) calls this awareness "intermodal feature binding awareness." It is *intermodal*, meaning that it involves more than one sense modality. It is *feature binding* in that the features are perceived as jointly bound to the same object or event. And it is *awareness*, because you are conscious of the features being bound to the object or event in this way.

While I agree that we can have awareness that the jolt and the clang are part of the same event, I will argue that there is no automatic feature binding mechanism that binds features like the jolt and the clang together. Instead, when you experience the jolt and the clang as part of the same event, this is the result of an associative learning process. The cymbal's jolt and the clang are best understood as a single learned perceptual unit, rather than as automatically bound. More generally, my claim is that multimodal cases involve learned associations, and I will outline a specific learning process in perception whereby we come to "chunk" the world into multimodal units. A central contribution of the paper is this: unitization is an entirely undiscussed way that an associationist might implement an associative account of multimodal perception. It is one thing to say that features *x* and *y* are associated. It is another thing to give a detailed account (drawing on an established perceptual learning process) of how exactly that association happens. In what follows, I attempt to do exactly that.

It can be difficult to tease apart the difference between an account of multimodal perception based on intermodal feature binding and an account based on associative learning. For now, the key question to ask is how features, such as a jolt and a clang, come to be coupled. Specifically, did the coupling happen in past experience, or did it happen just prior to your current experience? In other words, if you experience a jolt and a clang as part of the same event, is this due to those features getting coupled in your past experience, or did the coupling of the jolt and the clang occur just prior to your experience of them?

If feature binding awareness does not involve feature binding—which is what I will argue—then this flies in the face of the way that scientists working on multimodal perception have been thinking about these cases. Consider four such representative passages highlighted by O'Callaghan (forthcoming, ms pp. 8–9):

When presented with two stimuli, one auditory and the other visual, an observer can perceive them either as referring to the same unitary audiovisual event or as referring to two separate unimodal events *There appear to be specific mechanisms in the human perceptual system involved in the binding of spatially and temporally aligned sensory stimuli.* (Vatakis and Spence, 2007, 744, 754, italics were added for emphasis).

As an example of such privileged binding, we will examine the relation between visible impacts and percussive sounds, which allows for a *particularly powerful form of binding that produces audio-visual objects.* (Kubovy and Schutz, 2010, 42, italics were added for emphasis).

In a natural habitat information is acquired continuously and simultaneously through the different sensory systems. As some of these inputs

have the same distal source (such as the sight of a fire, but also the smell of smoke and the sensation of heat) it is reasonable to suppose that the organism should be able to bundle or bind information across sensory modalities and not only just within sensory modalities. For one such area where intermodal binding (IB) seems important, that of concurrently seeing and hearing affect, *behavioural studies have shown that indeed intermodal binding takes place during perception.* (Pourtois et al., 2000, 1329, italics were added for emphasis).

[T]here is undeniable evidence that the visual and auditory aspects of speech, when available, contribute to an integrated perception of spoken language *The binding of AV speech streams seems to be, in fact, so strong* that we are less sensitive to AV asynchrony when perceiving speech than when perceiving other stimuli (Navarra et al., 2012, 447, italics were added for emphasis)¹.

The traditional view is that multimodal perception at the conscious-level is the result of intermodal feature binding at the unconscious-level in all the ways mentioned above, whether it is with spatially and temporally aligned stimuli, audio-visual objects, with facial expression and tone of voice, or audio-visual speech streams. I will argue that this view is mistaken.

Whether multimodal perception involves an automatic binding process or an associative process has been discussed before (in the case of speech perception, for instance, see Altieri and Townsend, 2011; Altieri et al., 2011). Starting with the former view, the theory that multimodal perception involves an automatic binding process is consistent with several other theories in the psychological literature on perception, including Gibson's (1950, 1972, 1979) theory of direct perception and Fowler's discussion of speech as an amodal phenomenon (Fowler, 2004). On Gibson's view, for instance, we directly perceive objects with their features already integrated. We do not have to associate the jolt and the clang, for instance, because we directly perceive the cymbal, and the jolt and clang features are already integrated into the cymbal. Similarly, Fowler (2004) discusses the view that listeners directly perceive speech gestures. A gestural percept is amodal, as she describes it, with information from different sense modalities already integrated into it. On this view, you perceive a speech gesture with the auditory and visual features already integrated into it.

What Gibson, Fowler, and the binding view have in common is that features are automatically bound outside of and prior to conscious perception. Since processing happens early, a good model is a coactive model, which Townsend and Nozawa (1995) define as, "A parallel architecture which assumes that input from the separate parallel channels is consolidated into a resultant common processor" (p. 323). The binding mechanism, in this case, would serve as the common processor that consolidates information channels from different sense modalities. On a coactive model, the jolt and the clang information would be consolidated into the binding mechanism, and the output of that mechanism results in those bound features being available to consciousness. This enables awareness that the features are part of the same event (see Figure 1).

The view that multimodal perception results from an associative learning process, on the other hand, is consistent with several

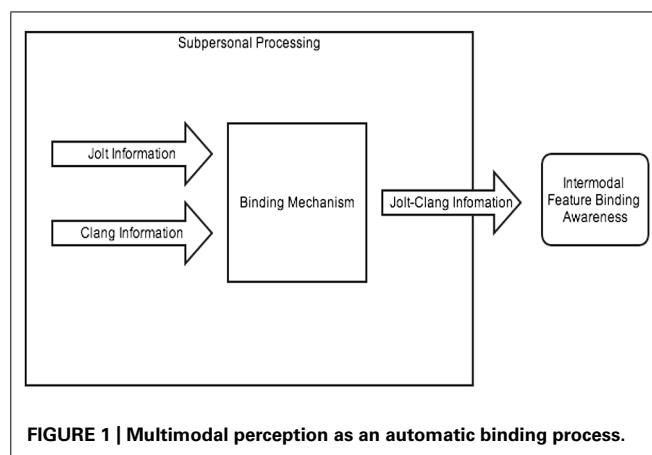


FIGURE 1 | Multimodal perception as an automatic binding process.

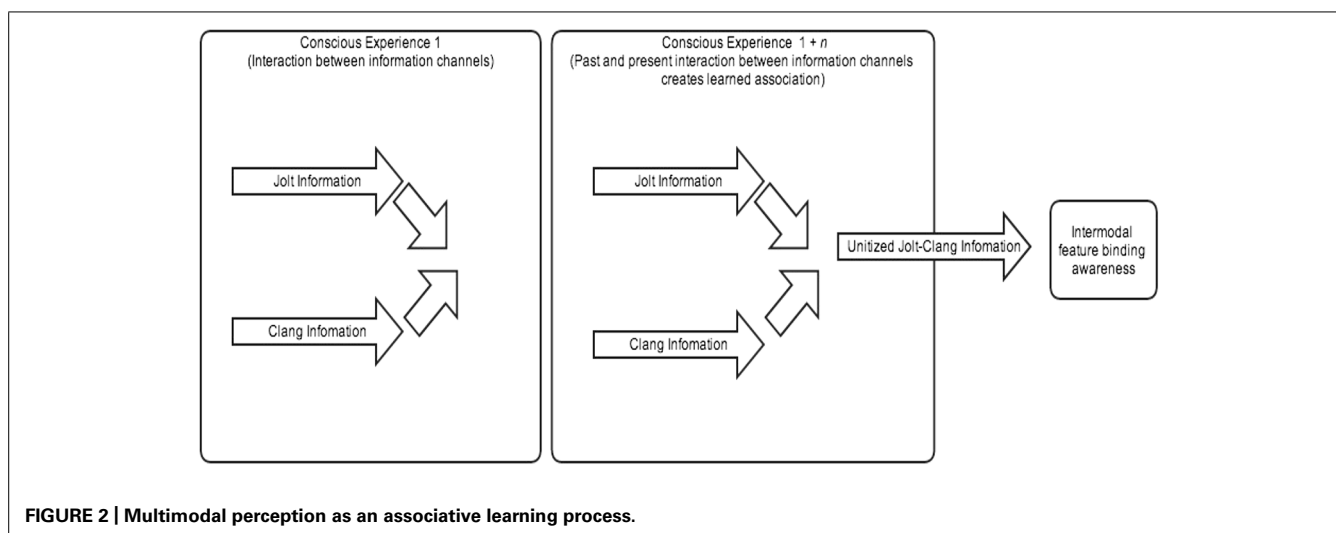
other theories in the psychological literature on perception. Smith and Yu (2007, 2008), for example, have studied how both infants and adults match words to scenes. As Quine (1960) pointed out, given a somewhat complex scene, for any given word, there are an infinite number of possible referents. Yet, Smith and Yu (2007, 2008) and Yu and Smith (2006, 2007, 2011, 2012) show how the binding of a word and a referent occurs through an associative learning process whereby infants and adults learn, across varying contexts, the statistical likelihood that a word refers to a particular kind of object). Along the same lines, Wallace (2004) describes how multisensory neurons require a protracted maturation process. Specifically, multisensory neurons get strengthened through experience.

What holds the theories of Smith and Yu, Wallace, and myself in common is that on our views features are not coupled by an automatic mechanism, but rather get coupled through an associative learning process. While associative accounts have been offered for speech perception, as Smith and Yu do, they have rarely been applied to multimodal cases outside of speech perception, and this paper shows how they can be applied in that way. A good model for associative learning is an interactive parallel processing model (see Townsend and Wenger, 2004). According to this model, the information channels from different sense modalities are processed in parallel, but can interact. The jolt and the clang information, for instance, involve parallel processing, and the interaction between the two information streams enables them to become associated. The result of this association is that the jolt and the clang are later experienced, not as distinct, but as part of one and the same event (see Figure 2)².

My plan for the rest of the paper is as follows. In section "Why is the Debate Significant?" I will say why it is significant whether we take multimodal cases to involve a binding mechanism or an associative process at the unconscious-level. In particular, it will have impact on philosophical issues from Molyneux's question to the question of how active or passive we should consider perception to be. In section "Intermodal Feature Binding Awareness," I will briefly explain O'Callaghan's notion of intermodal feature binding awareness—an account that details what is happening at

¹O'Callaghan adds to this list: Bushara et al. (2003), Bertelson and de Gelder (2004), Spence and Driver (2004), Spence (2007), and Stein (2012).

²Thanks to a reviewer for extensive suggestions about related literature, and the connections between that literature and the claims I make in the paper.



the conscious-level for cases that many psychologists have taken to involve intermodal feature binding at the unconscious-level. In section “Unitization,” I will offer a previously undiscussed alternative to intermodal feature binding—what’s called “unitization” in the literature on perceptual learning. In section “Applying Unitization to Multimodal Cases,” I will apply unitization to multimodal cases, and show how the phenomenon is consistent with O’Callaghan’s main argument for intermodal feature binding awareness. In section “Objections and Responses,” I will respond to some objections.

WHY IS THE DEBATE SIGNIFICANT?

Why does it matter whether intermodal feature binding awareness is the result of intermodal binding or of learned associations? One reason has to do with the implications the issue has for a long-standing philosophical problem. Molyneux’s question asks whether a man born blind, who can distinguish a cube and sphere by touch, could distinguish those shapes upon having his sight restored. If intermodal feature binding awareness is the result of learned associations, then we have a straightforward “no” answer to Molyneux’s question. You see the cube for the first time. No learned associations have taken place between sight and touch. So, no, you don’t recognize which is the cube and which is the sphere (see **Table 1**).

How we answer Molyneux’s question will, in turn, have ramifications for debates between nativists or rationalists such as Leibniz (1765/1982), on the one hand, and empiricists such as Locke (1690/1975) on the other hand. On Molyneux’s question, nativists hold that the association between the felt and seen cube is innate, while empiricists hold that it is learned. If the association between the felt and seen cube is learned, therefore yielding a “no” answer to Molyneux’s question, then this gives us an empiricist answer to Molyneux’s question, rather than a nativist one.

Recent experimental evidence lends support to the claim that the answer to Molyneux’s question is a “no.” A study conducted by Held et al. (2011) tested whether subjects who had just undergone cataract removal surgery for sight restoration, would be able to identify previously felt legos by sight. In the

study, subjects first were given one lego to touch. Next, they were visually presented with two distinctly shaped legos, and were asked which of the two legos they had previously been touching. Subjects performed at near-chance levels in answering this question. Held and colleagues interpret this result to mean that the answer to Molyneux’s question is likely to be “no,” since subjects born blind, who could distinguish between shapes by touch, could not distinguish those shapes upon having their sight restored (for a debate about the experimental design in Held et al., 2011; see Schwenkler, 2012, 2013; Connolly, 2013).

A second reason it matters whether intermodal feature binding awareness is the result of intermodal binding or of learned associations is that, as O’Callaghan has pointed out, one of the most important discoveries in the cognitive science of perception in the past two decades is that the senses involve extensive interaction and coordination. We want to understand how this works, and many cases of multisensory awareness are cases of binding awareness. But are cases of binding awareness the result of intermodal binding, or are they the result of learned associations? Depending on which one of these is our answer, we will have a different account of one of the most important discoveries in the cognitive science of perception in the past two decades.

A third reason the debate is important is that a view that makes multimodal perception a flexible, learned process (see, for instance, Connolly, 2014) fits more naturally with the emerging view of perception as a more active process than it has typically been taken to be. That is to say, it fits with a view of perception where the perceiver works to construct the world through learning and exploration rather than just passively receiving inputs that get transformed into a representation of the world. On an active view the perceiver does not simply look at the world as a passive observer, but has to look around, explore, and tweak the processes that are involved in perception to make them more useful to them for knowing what is out in the world, and for interacting with the world in an effective way.

Table 1 | Multimodal perception: associative learning vs. intermodal feature binding.

	Associative learning	Intermodal feature binding
When does the coupling happen?	In past experience	Jut prior to present experience
Can the theory accommodate intermodal feature binding awareness?	Yes (see “Applying Unitization to Multimodal Cases”)	Yes
How does the theory answer Molyneux’s question?	Definitive “No”	Likely “Yes”
What kind of processing model fits the Theory?	Interactive parallel processing model	Coactive model

INTERMODAL FEATURE BINDING AWARENESS

When you are listening to the drum solo, see the cymbal jolt, hear the clang, and are also aware that the jolt and the clang are part of the same event, this is a case of intermodal feature binding awareness. Why is intermodal feature binding awareness of theoretical significance? One reason is that it propels an argument, made by O’Callaghan (2014), that not all perceptual experience is modality specific, that is to say, that there are cases of multimodal perception which cannot be broken down into just seeing, hearing, touching, tasting, and smelling, happening at the same time. As he puts it, perceptual awareness is not just “minimally multimodal.” It is not just exhausted by perceptual awareness in each of the sense modalities happening at the same time.

Why does O’Callaghan deny minimal multimodality? One reason is due to intermodal feature binding awareness. Intermodal feature binding awareness occurs when you consciously perceive multiple features from more than one sense modality jointly to belong to the same object or event. O’Callaghan’s main argument for intermodal feature binding awareness runs as follows. Consider the difference between the following cases one and two. In case one, when the drummer begins a solo, you see the cymbal jolt and hear the clang, and you are aware that the jolt and the clang are part of the same event. In case two, you see the jolt and hear the clang, but you are not aware that the jolt and the clang are part of the same event. Perhaps you have never seen a cymbal before and are unaware of the sound that it makes. According to O’Callaghan, there may be a phenomenal difference between case one and case two. This difference is explicable in terms of intermodal feature binding awareness: case one involves such an awareness, while case two does not. O’Callaghan generalizes the point: “a perceptual experience as of something’s being F and G may differ in phenomenal character from an otherwise equivalent perceptual experience as of something F and something G, where F and G are features perceptually experienced through different modalities” (O’Callaghan, 2014, ms p. 8). This is just to say that in the cymbal example and others like it, case one differs from case two in terms of its phenomenology. O’Callaghan explains this difference in that the former, but not the latter case involves intermodal feature binding awareness.

Everything said so far is about feature binding awareness. This is something that happens at the conscious-level. But psychologists often talk about feature binding, and there they are referring to a unconscious process. As a representative view, Vatakis and Spence claim:

When presented with two stimuli, one auditory and the other visual, an observer can perceive them either as referring to the same unitary audiovisual event or as referring to two separate unimodal events There appear to be specific mechanisms in the human perceptual system involved in the binding of spatially and temporally aligned sensory stimuli. (Vatakis and Spence, 2007, 744, 754; quoted by O’Callaghan, 2014, ms p. 8)

But what is the connection between feature binding awareness and the feature binding process? The assumption in the empirical literature is that cases like the cymbal case depend upon feature binding at the unconscious-level—an assumption that I will argue is mistaken. Roughly and briefly, on my view, cases like the cymbal case are best explained through a process called “unitization,” whereby features (such as the jolt and the clang) that were once detected separately, are later detected as a single unit. For example, while someone who has never seen a cymbal before might plausibly experience the clash and the jolt not as the part of the same event, others unitize those features into the same event, due to learning.

O’Callaghan’s own argument is about feature binding awareness, which he describes as likely related to—but not the same as—feature binding itself. O’Callaghan explains the connection: “Feature binding awareness presumably depends upon feature binding processes. I say “presumably” because a feature binding process ... may require that features are detected or analyzed separately by subpersonal perceptual mechanisms” (forthcoming, ms p. 3). At the same time, O’Callaghan distances himself from feature binding processes. He allows that “it is possible that what I have characterized as feature binding awareness could occur without such a feature binding process” (forthcoming, ms p. 3). So, on O’Callaghan’s view, the existence of feature binding awareness does not imply a feature binding process.

O’Callaghan’s account of feature binding awareness is consistent with my view, since I deny a feature binding process, and his view does not imply such a process. But one place where O’Callaghan and I differ is with the name “feature binding awareness.” If there is an associative process involved rather than a feature binding mechanism and the result of the associative process manifests itself at the conscious-level (and I will argue that this is the case), it is hard to see why we should call the conscious upshot “feature binding awareness” rather than “associative awareness.” If there is an associative process, then since we will have ruled out a feature binding mechanism in favor of a different process, it would be inaccurate to call the conscious upshot “feature binding awareness.”³

At the same time, O’Callaghan and I are united in our departure from Spence and Bayne (2014), who say the following:

But are features belonging to different modalities bound together in the form of MPOs [multimodal perceptual objects]? ... [W]e think it

³I thank Casey O’Callaghan and Diana Raffman for clarifying the relationship between O’Callaghan’s position and my own.

is debatable whether the “unity of the event” really is internal to one’s experience in these cases, or whether it involves a certain amount of post-perceptual processing (or inference). In other words, it seems to us to be an open question whether, in these situations, one’s experience is of a MPO or whether instead it is structured in terms of multiple instances of unimodal perceptual objects. (Spence and Bayne, 2014, ms 27, 29; quoted by O’Callaghan, forthcoming, p. 5)

On Spence and Bayne’s account, it is debatable whether intermodal feature binding awareness occurs at all. So, in the cymbal case, where O’Callaghan and I think that you can see the cymbal jolt and hear the clang, and be aware that the jolt and the clang are part of the same event, Spence and Bayne think that is debatable. One alternative, they might say, is that you see the jolt of the cymbal, hear the clang, and infer that they are both associated with the same object. And on their view, it is an open question whether such an alternative is correct.

O’Callaghan’s account is restricted to the conscious-level. But we can ask what the unconscious processes are which produce it. Psychologists have assumed that intermodal feature binding produces multimodal perception, but I will now explore a previously un-discussed alternative to intermodal feature binding—what is called “unitization” in the literature on perceptual learning.

UNITIZATION

Robert Goldstone, one of the leading psychologists working on perceptual learning today, lists unitization as one of four mechanisms of *perceptual learning*. What is perceptual learning? Eleanor Gibson defines it as “any relatively permanent and consistent change in the perception of a stimulus array, following practice or experience with this array” (Gibson, 1963, p. 29). Perceptual learning involves perceptual changes. Perceptual changes occur so that we can better perform the cognitive tasks that we need to do. The idea is that to ideally perform cognitive tasks, it is better for perceptual systems to be flexible, rather than hardwired. As Goldstone puts it, one might be tempted to hold that the perceptual system is hardwired, the intuition being that “stable foundations make strong foundations” (Goldstone, 2010, p. v). But actually a better model of perception is a suspension bridge: “Just as a suspension bridge provides better support for cars by conforming to the weight loads, perception supports problem solving and reasoning by conforming to these tasks” (Goldstone, 2010, p. v). Perceptual systems are flexible rather than hardwired so that they can better support cognitive tasks. Specifically, the kind of flexibility on which I will focus is how perceptual systems are able to construct perceptual units of the various different sizes, which improve our ability to respond to our environment.

When people hear about perceptual learning, they often think of cases of improved discrimination abilities. William James, for instance, writes of a man who has learned to distinguish by taste between the upper and lower half of a particular type of wine (James, 1890, p. 509). What the man’s perceptual system had previously treated as a single thing is later treated as two distinct things. Psychologists who work on perceptual learning call this *differentiation*. But the converse happens as well. Sometimes, what has been treated previously by the perceptual system as two things, is later treated by it as one thing. Psychologists call this *unitization*. Perceptual units are created not just by breaking down larger units

(like the bottle of wine) into smaller one’s (like the top half and the bottom half), but also by merging smaller units into larger ones.

As Goldstone puts it, “Unitization involves the construction of single functional units that can be triggered when a complex configuration arises. Via unitization, a task that originally required detection of several parts can be accomplished by detecting a single unit [U]nitization integrates parts into single wholes” (Goldstone, 1998, p. 602)⁴. For example, consider someone who is developing an expertise in wine tasting and is learning to detect Beaujolais. Detecting it at first might involve detecting several features, such as the sweetness, tartness, and texture. But detecting the Beaujolais is later accomplished by just detecting it as a single unit. Since the Beaujolais gets unitized by your perceptual system, this allows you to quickly and accurately recognize it, when you taste it.

According to Goldstone and Byrge, unitization in perception is akin to “chunking” in memory (Goldstone and Byrge, 2014, ms p. 15). Normally, we are only able to commit 7 ± 2 items into short-term memory. Yet, we are easily able to do much better with the following string of 27 letters, by chunking them:

M O N T U E W E D F B I C I A K G B C B S N B C A B C

We can chunk the first nine letters as abbreviations for days of the week, the next nine as abbreviations for intelligence agencies, and the final nine as abbreviations for American television networks. Chunking is the building of new units that help to enable memory. Similarly, in perception, unitization allows us to encode complex information, which without unitization we might be unable to encode. Suppose, for instance, that you are drinking an extremely complex Beaujolais that you have never tasted before. Your perceptual system might unitize that type of wine, allowing you to recognize it, despite the fact that it is extremely complex.

A whole host of objects have been shown to be first processed as distinct parts, and later processed as a unit. Goldstone and Byrge offer the following diverse list: “birds, words, grids of lines, random wire structures, fingerprints, artificial blobs, and 3-D creatures made from simple geometric components” (Goldstone and Byrge, 2014, ms p. 17). Unitization occurs not just for things like cats and cars, but also for objects constructed in the lab. For instance, Gauthier and Tarr (1997) constructed a set of objects called “Greebles,” which shared a set of spatial features in common. When the subjects were exposed to the Greebles for long enough, they would begin to process them as units (Gauthier and Tarr, 1997). This showed up in the fact that people trained with the Greebles performed better than novices on speed and accuracy tests.

Many of the cases mentioned so far involve parts being treated as wholes after unitization, as when parts of a Greeble get treated as a whole unit. However, there are also cases in which attributes or properties become treated as units. For instance, a study by Shiffrin and Lightfoot (1997) showed that subjects are able to unitize the angular properties (i.e., *horizontality*, *verticality*, or *diagonality*) of a set of line segments. The study involved sets of three line segments, each of the segments angled either horizontally, vertically, or diagonally. Subjects were given a target set. Say,

⁴I will be drawing very closely from Goldstone (1998, pp. 602–604) and Goldstone and Byrge (2014) in explaining unitization.

for instance, that the target set is a set of two horizontal and one vertical line segments. Given that target, the subjects were asked to pick matching targets out (that is, all and only sets involving two horizontal and one vertical line segment), and ignore distractors (such as a set of two vertical and one horizontal line segments, or three horizontal line segments, among others). Subjects became very quick at this task through training, indicating that they had unitized the angular attributes of each of the three line segments.

As objects become unitized, the whole becomes easier to process perceptually than the part. At first, when one is learning what Greebles are, it is essential to identify them by their features. After they become unitized, however, it is easier to process them as whole units. Similarly, faces are unitized—they are easier to process as wholes than as parts. One interesting feature of face unitization is that inverting a face disrupts the unitization process. This means that faces are harder to recognize when presented upside-down than when presented right-side up (Diamond and Carey, 1986). Furthermore, if you distort features of a face, the distortions are quite apparent when the face is right-side up, but much less apparent when the face is upside-down. This effect, called the *Thatcher effect*, seems to show something important about the phenomenology of a unitized object. Specifically, what it is like to experience the upside-down distorted face is not simply what it is like to experience the right-side up distorted face plus inversion. Rather, there is something that it is like to experience, say, a distorted nose and lips in a unitized face, and that is different from what it is like to experience a distorted nose and lips in a non-unitized face.

APPLYING UNITIZATION TO MULTIMODAL CASES

My claim is that we unitize things, sometimes unimodally, as in the case of faces, birds, grids of lines, random wire structures, artificial blobs, and fingerprints. But sometimes unitization occurs multimodally as well. As Goldstone writes, “Neural mechanisms for developing configural units with experience are located in the superior colliculus and inferior temporal regions. Cells in the superior colliculus of several species receive inputs from many sensory modalities (e.g. visual, auditory, and somatosensory), and differences in their activities reflect learned associations across these modalities” (Goldstone, 1998). So, unitization occurs in part in the superior colliculus, a place that in cats and macaque monkeys receives multisensory inputs (see Stein and Wallace, 1996, p. 290).

Reconsider the difference between case one and case two of the cymbal example. In case one, you see the jolt of the cymbal, hear the clang, and are aware that the jolt and the clang are part of the same event. In case two, you see the jolt and hear the clang, but are not aware that they are part of the same event. My claim is that in case one, the jolt and the clang are unitized in the same event, while in case two they are not. Interestingly enough, one reason why case two might occur in the first place is if you have never seen a cymbal before, and so you have not built the association between what a cymbal looks like when it has been struck and what it sounds like.

This gives us a substantive reply to O’Callaghan’s main argument for intermodal feature binding awareness. O’Callaghan

argues for intermodal feature binding awareness by distinguishing between intermodal cases (1) and (2):

- (1) Perceiving a thing’s being both *F* and *G* (where *F* and *G* are features that are perceived through different sense modalities).
- (2) Perceiving a thing’s being *F* and a thing’s being *G*.

His idea is that (1) involves intermodal feature binding awareness, while (2) does not. But what I am saying is that the difference between (1) and (2) does not entail that intermodal feature binding has occurred (as psychologists have argued). We can distinguish between (1) and (2) phenomenally without appealing to intermodal feature binding. If (1) involves unitization, while (2) does not, then the phenomenal difference between them is that in (1), *F* and *G* are unitized in the thing, while in (2), they are not unitized in the thing.

Put more formally, in the case where you see the cymbal jolt and you hear the clang, let $E1[f(x)]$ and $E2[g(y)]$ denote that seeing the jolt *x* is a function *f* of vision and that the jolt is experienced as part of event 1, while hearing the clang *y* is a function *g* of audition and the clang is experienced as part of event 2. This is case one. Let $E1[f(x), g(y)]$ denote that seeing the jolt *x* is a function *f* of vision and the jolt is experienced as part of event one, while hearing the clang *y* is a function *g* of audition and the clang is also experienced as part of event one. This is case two, which is distinct from case one in that case two involves a single event while case one involves two events⁵.

Unitization is applicable to multimodal cases in other ways. Just as there are misfires in unimodal unitization, there are misfires in multimodal unitization cases as well. In the unimodal case, you might see a face in a grilled cheese sandwich. Your perceptual system is unitizing something that is not in fact a face. Now consider the multimodal case of ventriloquism. Typically, when you see moving lips and hear a congruent sound, the sound comes from the lips. You have built up an association between moving lips and the sounds that come from them. In the ventriloquist effect, you see the dummy’s lips move, and you hear a congruent sound. Your perceptual system unitizes the dummy’s lips and the sound. Yet, this unitization is a misfire. The sound is not in fact coming from the dummy’s lips.

In many cases, unitization enables more efficient processing. Instead of having to see the jolt of the cymbal, hear the clang, and judge that they are both associated with the same object, the unitization process efficiently does this for you. It would take a longer time to have to see the jolt, hear the clang, and judge that they are part of the same object. Unitization is a way of embedding that task into our quick perceptual system. We get the same information—that the jolt and the clang are part of the same event—without having to make time-consuming inferences to get there. This frees up cognition to make other, more sophisticated, inferences. To draw an analogy, an elite tennis player might not have to think about her footwork because that task has been embedded into motor memory, freeing her mind to make more sophisticated judgments about what to do in the match. As in such

⁵Thanks to a reviewer for the symbolism.

cases of motor learning, unitization can free up cognition to do more sophisticated tasks.

The units involved in unitization may have complex internal structures. Think about the unitization of faces, for instance. The associations involved are not just between two or so elements, but can be quite complicated associations between various different features of a face. Multimodal associations might be complicated in a similar way. There may not just be simple associations between two elements, but rather complicated associations between various different multimodal features of a single object or event.

When Pourtois et al. (2000), Vatakis and Spence (2007), Kubovy and Schutz (2010), and Navarra et al. (2012) others assume that intermodal feature binding occurs, their background reasoning is perhaps something like the following. We know that intramodal binding occurs, that is, that features detected by a single sense modality get bound together, as when the shape and color of a cup get bound to it. We know that multimodal perception occurs. So we can take binding and extrapolate from the intramodal case to apply it to multimodal perception. The overall argument that I am making is structurally similar. We know that unitization occurs, and we know that multimodal perception occurs. So I am taking unitization and extrapolating from the unimodal case to apply it to multimodal perception. But how do we know which point is the right starting point? How do we know whether we should start with binding or start with unitization? I now want to turn to a few cases that I think are potentially difficult for the intermodal binding view to handle, but easy for the unitization view.

Start by considering a case of illusory lip-synching—a case where someone appears to be lip-synching, but is actually singing. Sometimes this might occur due to a mismatch in association between the audio and the visual. In 2009, for instance, a Scottish singer named Susan Boyle gained worldwide fame from her appearance on the TV show “Britain’s Got Talent.”⁶ Her performance was captivating to many people because to them she did not look as if she could have such an impressive singing voice. They did not associate that sound with that look. And part of the good that came out of her case was that people broke their previous false association. Now imagine that you are in the audience as Susan Boyle steps on stage and sings. Plausibly, this would be a confusing experience. At first, you might not localize the sound at Susan Boyle’s moving lips. In your experience, it might be a case of illusory lip-synching. You might experience the sound as coming from elsewhere, even though it is actually coming from Susan Boyle.

Cases where vocal sounds are incongruous with the visual might be most vivid with pets, and amusing videos are often made documenting the results, showing animals that sound like human beings or like fire engine sirens. Consider one such example. Suppose you are listening to your radio with your dog nearby. A song comes on the radio that you haven’t heard before. You happen to glance over at your dog, who appears to be moving its mouth in synch with the vocals. Then you realize that what you thought were the vocals are actually coming from your dog.⁷

By appealing to learned associations, the singing dog case (and others like it) makes sense—the radio’s location and the dog’s sound get unitized. This happens because through experience, your perceptual system associates the sound that the dog makes with the radio. That sound is the kind of sound that would typically come from a radio. When the radio’s location and the dog’s sound get unitized, this is a misfire. The sound came from the dog and not from the radio. However, the misfire is understandable, given the fact that that type of sound typically comes from a radio and not a dog. We can apply the lesson of this case more generally. Past associations (between, say, types of sounds and types of things) determine the specific multimodal units that we experience.

It is unclear what psychologists who advocate intermodal feature binding would say about these sorts of cases. The dog’s mouth movement and the sound have happened at the exact same time, and from the same spatial location, but fail to be bound. But if binding were an automatic mechanism, wouldn’t intermodal binding just bind the dog’s voice to the dog’s mouth?

One option for the defender of binding is to hold that binding need not be automatic, but can be modulated by cognitive factors like whether or not the noise is the sound that a dog can make. For example, O’Callaghan (forthcoming, ms p. 15) quotes Vatakis and Spence (2007, p. 744), who claim that binding need not depend just on “low-level (i.e., stimulus-driven) factors, such as the spatial and temporal co-occurrence of the stimuli,” but can depend on “higher level (i.e., cognitive) factors, such as whether or not the participant assumes that the stimuli should ‘go together.’” If Vatakis and Spence (2007) and O’Callaghan (forthcoming) are right, then binding need not be automatic, since it can be modulated by cognitive factors.

If binding need not be automatic, but can be modulated by cognitive factors, then this presents a difficult challenge. My claim was that a view on which binding is automatic gets cases like the dog case wrong, since it would predict that the dog’s voice gets bound to the dog’s mouth, which is not what happens. Yet, if theorists defending an intermodal binding process can just weaken the automaticity requirement, then it seems that they can accommodate cases like the dog case into their model. One possible response is to appeal to parsimony. Given that it is difficult empirically to pull apart the associative account from the intermodal binding account, an appeal to the theoretical virtues of each view is warranted. If an associative view can handle all putative cases of intermodal binding, but an intermodal binding view cannot handle all cases without appealing to a learning mechanism (to deal with cases of involving the plausibility of combination), then it seems like parsimony supports the associative view. Of course, there may be other theoretical virtues to take into account when examining both views, as well as other empirical considerations, but it seems at the very least that parsimony tells in favor of an associative account.

OBJECTIONS AND RESPONSES

My claim is that appealing to learned associations (such as the association between the dog’s sound and the radio’s location) makes sense of cases like the dog case. But one might object that there are other equally good or better ways of making sense of such

⁶For a video of her initial performance, which has been viewed over 150 million times, see: <http://www.youtube.com/watch?v=RxPZh4AnWyk>

⁷As an example of this, see: <https://www.youtube.com/watch?v=KECwXu6qz8I>

cases. One alternative is that the singing dog case is just a straightforward crossmodal illusion, like the ventriloquist effect⁸. The idea is that just as in the ventriloquist effect, the auditory location of the sound gets bound to the moving lips, so too in the singing dog case, the sound gets bound to the location of the radio. In both cases, the experience is illusory. Just as the sound is not coming from the ventriloquist dummy in the ventriloquist case, so too is it not coming from the radio in the singing dog case.

My response is as follows. In the ventriloquist effect, both binding and association are viable explanations, at least on its face. For the associative explanation, it could be that we build an association between the sound of a voice and the movement of lips. On the other hand, an explanation just in terms of binding is equally plausible. It could be that we bind sounds with congruent movements together. In the singing dog case, however, only an explanation in terms of association will suffice. The associative explanation is that we build an association between voice sounds and radios, and so when the dog makes a voice sound, that sound gets unitized with the radio. An explanation just in terms of binding gives the wrong prediction for the singing dog case. If we bind sounds with congruent movements together, then the dog's sound should be bound to the congruent movement of the dog's mouth. Consider a second objection that there is another equally good or better way of making sense of cases like the singing dog case. According to this objection, feature binding can be guided by categorical perception. The idea is that in the singing dog case, and cases like it, the categories that you have (of dog voices and radio sounds, for instance) influence what gets bound to what. So, there is a story to be told about the selection of features with regard to which features get bound together. And it is natural to suppose that categorical learning might have a role to play in which features get selected and thus bound together. Traditionally, the literature on binding has been very much concerned with sensory primitives like colors and shapes, and there's a question about whether higher-level perceptual features get bound in that same way. According to this objection, we do not need to choose between feature binding and learned associations because they can play a role together⁹.

I find this objection to be plausible, yet currently unsubstantiated. To the best of my knowledge at least, there is no empirical evidence demonstrating the claim that categorical perception can guide feature binding. I take it to still be a plausible hypothesis, however, because there is some evidence that learning connections between sensory primitives can influence the binding process (Colzato et al., 2006). But as far as I know, this same influence has not been demonstrated for higher-level perceptual features. The objection is right in that it remains a live option that feature binding can be guided by categorical perception. Still, if the goal of the objection is to establish that there is another equally good or better way of making sense of cases like the singing dog case, in absence of empirical evidence to ground this alternative, the alternative is not a better explanation. There is empirical evidence,

due to studies on unitization, to ground the explanation of the singing dog case in terms of learned association. So, barring empirical evidence to ground the explanation of that case in terms of categorical perception guiding feature binding, this explanation is not equal or better than the explanation in terms of learned association.

A third objection is to the idea that unitization can explain multimodal cases. According to this objection, unitization implies that there was something there before to unitize. But in certain cases of multimodal perception, this seems implausible. Take the case of flavor perception. Flavor is a combination of taste, touch, and retronasal (inward-directed) smell (see Smith, 2013). Yet, flavors are always just given to you as single unified perceptions. You are never given just the parts. You don't start by having a retronasal smell experience, taste and touch, and then unitize those things¹⁰.

I think this objection points to an exception to the argument that I am making. Flavor perception is a special case of multimodal perception where a unitization account does not apply. This might seem *ad hoc*, but at the same time, it is well-recognized that flavor perception is a special case of multimodal perception in general. Flavor is special, because as O'Callaghan points out, it is a "type of feature whose instances are perceptible only multimodally" (O'Callaghan, 2014, ms p. 26). That is to say, where in the cymbal case, one can experience the jolt and the clang either together or separately (if one were to close one's eyes or shut one's ears, for instance), in the case of flavor properties, they are perceptible only through taste, touch, and retronasal smell. Given that, it should not be surprising that flavor has a special treatment.

A fourth objection continues on the third, but focuses on speech perception rather than flavor perception. According to this objection, there are documented cases of infant speech perception where an infant has a coupling without ever being exposed to either of the coupling's components. For instance, before eleven months, Spanish infants can match /ba/ and /va/ sounds with corresponding images of someone unambiguously saying /ba/ and /va/ (Pons et al., 2009). Spanish itself does not make a distinction between /ba/ and /va/. Even if an infant is not surrounded by English speakers, for example, the infant before eleven months can still match audio and visual English phonemes. But how can this be through association when the infant herself was not surrounded by English speakers? Why are infants able to match the sounds with the images, and how can an associative account explain it?¹¹.

This objection presents a difficult but not insurmountable challenge for the unitization view of multimodal perception. In the study in question (Pons et al., 2009), all infants initially underwent two 21 s trials in which they were presented with silent video clips of a bilingual speaker of Spanish and English, repeatedly producing a /ba/ syllable on one side of the screen and a /va/ syllable on the other side. So, while it is right to say that the Spanish infants had not been surrounded by English speakers, they had

⁸Thanks to Casey O'Callaghan for raising this possibility.

⁹Thanks to Tim Bayne for this objection.

¹⁰I owe this objection to Barry C. Smith.

¹¹I owe this objection to Barry C. Smith and Janet Werker.

been exposed to English speakers. And it remains a possibility that this exposure was sufficient for matching audio and visual English phonemes through association. This possibility is more plausible if we allow that some pairs are more easily unitized than others, in this case /ba/ and /va/ sounds with corresponding images of someone unambiguously saying /ba/ and /va/.

CONCLUSION

My account sides with O'Callaghan in one respect, and against the dominant view in psychology in another respect. With O'Callaghan, I accept that perceptual awareness is not just "minimally multimodal." It is not just exhausted by perceptual awareness in each of the sense modalities happening at the same time. The cymbal case shows this. There is something that it is like to be aware that the jolt of the cymbal and the clang are part of the same event. And this is different from what it is like to just see the jolt and hear the clang. In holding this view, I depart from Spence and Bayne (2014), who find it debatable that it is part of one's experience that the jolt and the clang are part of the same event, rather than part of post-perceptual processing or some kind of inference the subject makes.

According to the dominant view in psychology (including Pourtois et al., 2000; Vatakis and Spence, 2007; Kubovy and Schutz, 2010), multimodal experiences result from an intermodal feature binding process. Against this dominant view, however, I am a skeptic of intermodal feature binding. This is because I think that an associative process rather than a binding mechanism best explains multimodal perceptions. To show this, I outlined a specific associative process in the literature on perceptual learning that can explain multimodal perceptions: unitization. I argued, for instance, that unitization best explains what it is like to be aware that the jolt of the cymbal and the clang are part of the same event. The jolt and the clang are unitized in that event. So, I am skeptical of an explanation of this case, and cases like it, in terms of intermodal feature binding. Such multimodal perceptions are unitized, not bound¹².

REFERENCES

- Altieri, N., Pisoni, D. B., and Townsend, J. T. (2011). Some behavioral and neurobiological constraints on theories of audiovisual speech integration: a review, and suggestions for new directions. *Seeing Perceiving* 24, 513–539. doi: 10.1163/187847611X595864
- Altieri, N., and Townsend, J. T. (2011). An assessment of behavioral dynamic information processing measures in audiovisual speech perception. *Front. Psychol.* 2:238. doi: 10.3389/fpsyg.2011.00238
- Bertelson, P., and de Gelder, B. (2004). "The psychology of multimodal perception," in *Crossmodal Space and Crossmodal Attention*, eds C. Spence and J. Driver (New York: Oxford University Press), 141–177.
- Bushara, K. O., Hanakawa, T., Immisch, I., Toma, K., Kansaku, K., and Hallett, M. (2003). Neural correlates of cross-modal binding. *Nat. Neurosci.* 6, 190–195. doi: 10.1038/nn993
- Colzato, L. S., Raffone, A., and Hommel, B. (2006). What do we learn from binding features? Evidence for multilevel feature integration. *J. Exp. Psychol. Hum. Percept. Perform.* 32, 705–716. doi: 10.1037/0096-1523.32.3.705
- Connolly, K. (2013). How to test Molyneux's question empirically. *Iperception* 4, 508–510. doi: 10.1068/i0623jc
- Connolly, K. (2014). "Making sense of multiple senses," in *Consciousness Inside and Out: Phenomenology, Neuroscience, and the Nature of Experience*, ed. R. Brown (Dordrecht: Springer), 351–364.
- Diamond, R., and Carey, S. (1986). Why faces are and are not special: an effect of expertise. *J. Exp. Psychol. Gen.* 115, 107–117. doi: 10.1037/0096-3445.115.2.107
- Fowler, C. A. (2004). "Speech as a supramodal or amodal phenomenon," in *The Handbook of Multisensory Processes*, eds G. A. Calvert, C. Spence, and B. E. Stein (Cambridge, MA: MIT Press), 189–201.
- Gauthier, I., and Tarr, M. J. (1997). Becoming a "greeble" expert: exploring mechanisms for face recognition. *Vision Res.* 37, 1673–1682. doi: 10.1016/S0042-6989(96)00286-6
- Gibson, E. J. (1963). Perceptual learning. *Annu. Rev. Psychol.* 14, 29–56. doi: 10.1146/annurev.ps.14.020163.000333
- Gibson, J. J. (1950). *The Perception of the Visual World*. Boston: Houghton Mifflin.
- Gibson, J. J. (1972). "A theory of direct visual perception," in *The Psychology of Knowing*, eds J. R. Royce and W. W. Rozeboom (New York: Gordon & Breach), 215–240.
- Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.
- Goldstone, R. L. (1998). Perceptual learning. *Annu. Rev. Psychol.* 49, 585–612. doi: 10.1146/annurev.psych.49.1.585
- Goldstone, R. L. (2010). "Forward to gauthier," in *Perceptual Expertise: Bridging Brain and Behavior*, eds M. Tarr and D. Bub (Oxford New York: Oxford University Press).
- Goldstone, R. L., and Byrge, L. A. (2014). "Perceptual learning," in *The Oxford Handbook of the Philosophy of Perception*, ed. M. Matthen (Oxford: Oxford University Press).
- Held, R., Ostrovsky, Y., de Gelder, B., Gandhi, T., Ganesh, S., Mathur, U., et al. (2011). The newly sighted fail to match seen shape with felt. *Nat. Neurosci.* 14, 551–553. doi: 10.1038/nn.2795
- James, W. (1890). *Principles of Psychology*. New York: Holt. doi: 10.1037/11059-000
- Kubovy, M., and Schutz, M. (2010). Audio-visual objects. *Rev. Philos. Psychol.* 1, 41–61. doi: 10.1007/s13164-009-0004-5
- Leibniz, G. W. (1765/1982). "Perception," in *New Essays on Human Understanding*, trans. eds P. Remnant and J. Bennett. Cambridge: Cambridge University Press.
- Locke, J. (1690/1975). "Perception," in *An Essay Concerning Human Understanding Book II*, ed. P. H. Niddich (New York: Oxford University Press).
- Navarra, J., Yeung, H. H., Werker, J. F., and Soto-Faraco, S. (2012). "Multisensory interactions in speech perception," in *The New Handbook of Multisensory Processing*, ed. B. E. Stein, (Cambridge, MA: MIT Press), 435–452.
- O'Callaghan, C. (2014). "Not all perceptual experience is modality specific," in *Perception and Its Modalities*, eds D. Stokes, S. Biggs, and M. Matthen, (Oxford: Oxford University Press).
- O'Callaghan, C. (forthcoming). "Intermodal binding awareness," in *Sensory Integration and the Unity of Consciousness*, eds D. Bennett and C. Hill (Cambridge, MA: MIT Press).
- Pons, F., Lewkowicz, D. J., Soto-Faraco, S., and Sebastián-Gallés, N. (2009). Narrowing of intersensory speech perception in infancy. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10598–10602. doi: 10.1073/pnas.0904134106
- Pourtois, G., de Gelder, B., Vroomen, J., Rossion, B., and Crommelinck, M. (2000). The time-course of intermodal binding between seeing, and hearing affective information. *Neuroreport* 11, 1329–1333. doi: 10.1097/00001756-200004270-00036
- Quine, W. V. O. (1960). *Word and Object*. Cambridge, MA: MIT Press.
- Schwenkler, J. (2012). On the matching of seen and felt shapes by newly sighted subjects. *Iperception* 3, 186–188. doi: 10.1068/i0525ic
- Schwenkler, J. (2013). Do things look the way they feel? *Analysis* 73, 86–96. doi: 10.1093/analysis/ans137
- Shiffrin, R. M., and Lightfoot, N. (1997). Perceptual learning of alphanumeric-like characters. *Psychol. Learn. Motiv.* 36, 45–81. doi: 10.1016/S0079-7421(08)60281-9
- Smith, B. C. (2013). "Taste, philosophical perspectives," in *Encyclopedia of the Mind. Thousand Oaks*, ed. H. E. Pashler (California: SAGE Publications, Inc.).
- Smith, L., and Yu, C. (2007). "Infants rapidly learn words from noisy data via cross-situational statistics," in *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (Mahwah, NJ: Lawrence Erlbaum), 27.
- Smith, L., and Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition* 106, 1558–1568. doi: 10.1016/j.cognition.2007.06.010

¹²Thanks to Casey O'Callaghan, Tim Bayne, Barry C. Smith, Diana Raffman, Ophelia Deroy, Mohan Matthen, Peter Baumann, Janet Werker, Matthew Fulkerson, Adrienne Prettyman, Zoe Jenkin, Aaron Henry, an anonymous reviewer, and to the audience at the 2014 workshop on multisensory integration at the University of Toronto.

- Spence, C. (2007). Audiovisual multisensory integration. *Acoust. Sci. Technol.* 28, 61–70. doi: 10.1250/ast.28.61
- Spence, C., and Bayne, T. (2014). “Is consciousness multisensory?” in *Perception and its Modalities*, eds S. Biggs, D. Stokes, and M. Matthen (Oxford: Oxford University Press).
- Spence, C., and Driver, J. (ed.). (2004). *Crossmodal Space and Crossmodal Attention*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780198524861.001.0001
- Stein, B. E. (2012). *The New Handbook of Multisensory Processing*. Cambridge, MA: MIT Press.
- Stein, B. E., and Wallace, M. T. (1996). Comparisons of cross-modality integration in midbrain and cortex. *Prog. Brain Res.* 112, 289–299. doi: 10.1016/S0079-6123(08)63336-1
- Townsend, J. T., and Nozawa, G. (1995). Spatio-temporal properties of elementary perception: an investigation of parallel, serial, and coactive theories. *J. Math. Psychol.* 39, 321–359. doi: 10.1006/jmps.1995.1033
- Townsend, J. T., and Wenger, M. J. (2004). A theory of interactive parallel processing: new capacity measures and predictions for a response time inequality series. *Psychol. Rev.* 111, 1003–1035. doi: 10.1037/0033-295X.111.4.1003
- Vatakis, A., and Spence, C. (2007). Crossmodal binding: evaluating the “unity assumption” using audiovisual speech stimuli. *Percept. Psychophys.* 69, 744–756. doi: 10.3758/BF03193776
- Wallace, M. T. (2004). The development of multisensory processes. *Cogn. Proc.* 5, 69–83. doi: 10.1007/s10339-004-0017-z
- Yu, C., and Smith, L. B. (2006). “Statistical cross-situational learning to build word-to-world mappings,” in *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (Mahwah, NJ: Lawrence Erlbaum).
- Yu, C., and Smith, L. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychol. Sci.* 18, 414–420. doi: 10.1111/j.1467-9280.2007.01915.x
- Yu, C., and Smith, L. B. (2011). What you learn is what you see: using eye movements to study infant cross-situational word learning. *Dev. Sci.* 14, 165–180. doi: 10.1111/j.1467-7687.2010.00958.x
- Yu, C., and Smith, L. B. (2012). Modeling cross-situational word-referent learning: prior questions. *Psychol. Rev.* 119, 21–39. doi: 10.1037/a0026182

Conflict of Interest Statement: The author 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: 19 July 2014; accepted: 10 September 2014; published online: 26 September 2014.

Citation: Connolly K (2014) Multisensory perception as an associative learning process. *Front. Psychol.* 5:1095. doi: 10.3389/fpsyg.2014.01095

This article was submitted to *Consciousness Research*, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Connolly. 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.



Body ownership and experiential ownership in the self-touching illusion

Caleb Liang^{1,2*}, Si-Yan Chang¹, Wen-Yeo Chen², Hsu-Chia Huang² and Yen-Tung Lee¹

¹ Department of Philosophy, National Taiwan University, Taipei, Taiwan

² Graduate Institute of Brain and Mind Sciences, National Taiwan University, Taipei, Taiwan

Edited by:

Aleksandra Mroczko-Wasowicz,
National Yang Ming University,
Taiwan

Reviewed by:

Tom Froese, Universidad Nacional
Autónoma de México, Mexico
Bigna Lenggenhager, University
Hospital Zürich, Switzerland

*Correspondence:

Caleb Liang, Department of
Philosophy, Graduate Institute of
Brain and Mind Sciences, National
Taiwan University, No. 1, Section 4,
Roosevelt Road, Taipei 106, Taiwan
e-mail: yiliang@ntu.edu.tw

We investigate two issues about the subjective experience of one's body: first, is the experience of owning a full-body fundamentally different from the experience of owning a body-part? Second, when I experience a bodily sensation, does it guarantee that I cannot be wrong about whether it is me who feels it? To address these issues, we conducted a series of experiments that combined the rubber hand illusion (RHI) and the "body swap illusion." The subject wore a head mounted display (HMD) connected with a stereo camera set on the experimenter's head. Sitting face to face, they used their right hand holding a paintbrush to brush each other's left hand. Through the HMD, the subject adopted the experimenter's first-person perspective (1PP) as if it was his/her own 1PP: the subject watched either the experimenter's hand from the adopted 1PP, and/or the subject's own hand from the adopted third-person perspective (3PP) in the opposite direction (180°), or the subject's full body from the adopted 3PP (180°, with or without face). The synchronous full-body conditions generate a "self-touching illusion": many participants felt that "I was brushing my own hand!" We found that (1) the sense of body-part ownership and the sense of full-body ownership are not fundamentally different from each other; and (2) our data present a strong case against the mainstream philosophical view called the immunity principle (IEM). We argue that it is possible for misrepresentation to occur in the subject's sense of "experiential ownership" (the sense that I am the one who is having this bodily experience). We discuss these findings and conclude that not only the sense of body ownership but also the sense of experiential ownership call for further interdisciplinary studies.

Keywords: body ownership, experiential ownership, self-as-object, self-as-subject, self-touching illusion, pre-reflective immunity, bodily self-consciousness

INTRODUCTION

Many daily experiences involve the sense of *body ownership*, which concerns what it is like to feel this hand or body *as mine*. Walking into a coffee shop, I quickly get a cup of cappuccino and take a sip to enjoy the nice taste and aroma. I experience the hand holding the cup as *my* hand, and I experience this particular body that just walked in as *my* body. Although the sense of body ownership has been studied by many groups in recent years, two key issues are still to be addressed. First, what is the relationship between the sense of body-part ownership and the sense of full-body ownership? Is the latter fundamentally different from the former? Or is the difference only a matter of degree? The second issue concerns: *who* is undergoing the experiences that occur in this particular body or body-part? By walking in and taking the sip, I not only experience this hand and body as mine, but I also have an implicit sense that I am the unique subject of those experiences that involve my body or body parts. For instance, I have an implicit sense that it is *me* who is experiencing the specific aroma and taste of cappuccino, it is me who is having the tactile sensations of holding the coffee mug, and it is me who is experiencing this particular body that just walked into the coffee shop. We will

call this the sense of *experiential ownership*. The issue that we intend to investigate is: can one's sense of experiential ownership go wrong?

The sense of body ownership and the sense of experiential ownership correspond to the philosophical distinction between the sense of *self-as-object* and the sense of *self-as-subject* (Wittgenstein, 1958; Shoemaker, 1968; Gallagher, 2012). Wittgenstein once made a famous distinction between using the first-person pronoun "I" *as-object* and using same term *as-subject*. He says: "It is possible that, say in an accident, I should feel a pain in my arm, see a broken arm at my side, and think it is mine, when really it is my neighbor's ... On the other hand, there is no question of recognizing a person when I say I have toothache. To ask 'are you sure it is *you* who have pains?' would be nonsensical" (1958, p. 67). The idea is that when one is conscious of oneself-as-object, error is always possible. However, when one is conscious of oneself-as-subject, a specific type of mistake is impossible. Shoemaker (1968) has articulated this idea by explaining that we are "immune to error through misidentification relative to the first-person pronouns" (IEM). Focusing on the case of phenomenal state, IEM states that when I am aware of a phenomenal state

through first-personal access, such as introspection, somatosensation, proprioception, etc., *I cannot be wrong* about whether it is I who feels it.

The characterization of self-as-object above fits well with our current knowledge about body ownership. Researchers have shown how misrepresentations may occur in one's sense of body ownership. This fits well with the view that "I"-as object and consciousness of self-as-object can be mistaken. In the RHI, watching a rubber hand being stroked synchronously with one's own unseen hand causes many subjects to feel as if the rubber hand is their own hand (Botvinick and Cohen, 1998; Tsakiris and Haggard, 2005). Researchers have studied not only body-part but also full-body illusions. In the full-body experiments by Lenggenhager et al. (2007), many subjects experienced the illusion that the virtual body was their own, and that they saw themselves from outside the body (cf. Ehrsson, 2007 for a different set-up). In the body swap illusion, many subjects felt that another person's whole body became their own and reported that "I was shaking hands with myself" (Petkova and Ehrsson, 2008, p. 5). These experiments have been widely used as paradigms for studying the sense of body ownership (Bertamini et al., 2011; Rohde et al., 2011; Pfeiffer et al., 2013; Salomon et al., 2013; van Doorn et al., 2014).

In light of these studies, the first issue can be stated as follows: Is the sense of body ownership involved in full-body illusions fundamentally different from that involved in the body-part illusion? Or is the difference only a matter of degree? On the one hand, Blanke and Metzinger (2009) propose that the most basic sense of self, i.e., what they call *minimal phenomenal selfhood* (MPS), is "experienced as a single feature, namely a coherent representation of the whole, spatially situated body—and not as multiple representations of separate body parts" (p. 9). The key features of MPS, including self-identification, self-location and first-person perspective are not modulated during the RHI. So Blanke and Metzinger believe that MPS can be illuminated only by investigating the sense of full-body ownership. This suggests that the sense of full-body ownership is fundamentally distinct from the sense of body-part ownership. On the other hand, Tsakiris (2010) focuses primarily on the RHI and holds that "the necessary conditions for the experience of ownership over a body-part seem to be the same as the ones involved in the experience of ownership for full bodies" (p. 710). Also, Petkova et al. (2011b) in an fMRI study suggest that "the unitary experience of owning an entire body is produced by neuronal populations that integrate multisensory information across body segments" (p. 1118). These views lean toward the position that there is no essential difference between the sense of body-part ownership and the sense of full-body ownership. We think that, in order to solve this issue, the body-part and full-body illusions should not be treated separately. So in this study we combine both types of illusions. We test the hypothesis that the sense of body-part ownership and the sense of full-body ownership are not fundamentally distinct from each other. This psychophysical hypothesis, if correct, may provide a useful guide for investigating the relevant neural mechanisms.

In contrast to our current knowledge of body ownership, experiential ownership is almost neglected by researchers. To clarify the second issue, it will be very useful to distinguish between the

fact and the *sense* of experiential ownership. Consider the simple example again. On the one hand, it is a fact that right now it is me, not you, who is the subject of this particular experience. Call this the *fact* of experiential ownership. This fact is objective because, for every conscious experience we can ask "who is the subject of that experience?" and there is a fact about it. On the other hand, this fact is connected with a first-personal perspective: in taking the sip, I have an implicit sense that it is me who is having this experience. Call this the *sense* of experiential ownership. It is subjective in that one can experience oneself as the subject simply by experiencing phenomenal states; the former is a part of the latter. We suggest that the sense of self-as-subject introduced above can be captured by the sense of experiential ownership.

Now the second issue is: can misrepresentation occur in one's sense of experiential ownership? Can one's sense of experiential ownership misrepresent the relevant fact of experiential ownership? Influenced by Wittgenstein and Shoemaker, most philosophers believe that this is a purely conceptual or semantic issue, and the answer is negative (Coliva, 2006; cf. also papers in Prosser and Récanati, 2012). We are skeptical about this mainstream position, and propose that the sense of experiential ownership is open to empirical investigations. Our hypothesis is that, the fact of experiential ownership can be misrepresented by the subject's sense of experiential ownership. *Pace* Wittgenstein, sometimes it makes perfect sense to ask "are you sure it is you who is experiencing so-and-so?" and that Shoemaker's immunity principle (IEM), or at least some versions of it, fails to hold.

To address these two issues and test our hypotheses, we conducted a series of experiments that combined the rubber hand illusion (RHI, Botvinick and Cohen, 1998; Tsakiris and Haggard, 2005) and the "body swap illusion" (Petkova and Ehrsson, 2008). By manipulating the participant's visual perspective and allowing the participant to interact with the experimenter, many subjects experienced what we call the "self-touching illusion": the subject felt that "I was brushing my own hand!" The subject was touching someone and being touched at the same time, as well as watching his/her own body in front of him/herself. This subject-experimenter interaction makes the illusion quite different from both the standard RHI and full-body illusions (FBI, Ehrsson, 2007; Lenggenhager et al., 2007). As we will see below, the self-touching experiments enabled us to compare the sense of full-body ownership with the sense of body-part ownership. Moreover, they created a situation wherein, subjectively, it was not totally clear whether it was me or someone else who felt the touch.

METHODS

MATERIALS AND PARTICIPANTS

In this study, we used a head mounted display (HMD, Sony HMZ-T1) and a stereo camera (Sony HDR-TD20V). The skin conductance responses (SCR) were recorded with a Data Acquisition Unit-MP35 (Biopac Systems, Inc. USA). For questionnaires, we used a Likert scale from "strongly disagree" (−3) to "strongly agree" (+3). The questionnaire statements are randomly distributed and can be divided into the following categories: body-part ownership, full-body ownership, touch referral, agency, self-touching illusion, experiential ownership, and double body

Table 1 | The questionnaires consisting of 13 statements divided into seven categories.

Category		
Body-part ownership	Q1.	It felt as if the hand seen on the screen was my hand
	Q2.	It seemed as if the touch I felt was on the hand brushed by the paintbrush on the screen
Touch fererral	Q11.	It seemed as if the touch I felt was on the body in front of me
	Q3.	It felt as if I could control the hand holding the paintbrush on the screen
Agency	Q8.	It felt as if I could control the body in front of me
	Q6.	It felt as if the body in front of me was mine.
Full-body ownership	Q7.	It felt as if I was sitting in front of me
	Q4.	It felt as if I was brushing my own hand
Self-touching illusion	Q5.	The person whom I brushed was me, not someone else
	Q9.	It was me who felt being brushed, not someone else
Experiential ownership	Q10.	The person who felt being brushed was not me
	Q12.	It felt as if I had two bodies
Double body effect	Q13.	It felt as if I was looking at myself from the opposite side

Statements in questionnaires. The questionnaires were in Chinese when presented to participants. Here are the English translations, and the wordings of some statements were slightly adjusted to fit the different HMD images across experiments. The questionnaires for the full-body conditions did not contain body-part statements Q1–Q3.

effect (Table 1). We conducted three experiments, each with four conditions. See Table 2 below for the details of the participants. All participants gave written consent prior to the experiments. This study was approved by the Research Ethics Committee of National Taiwan University (NTU-REC: 201310HS026).

PROCEDURE

The subject wore a HMD connected with a stereo camera positioned on the experimenter's head. Sitting face to face, they used their right hand to hold a paintbrush to brush each other's left hand for 2 min (Figure 1A). We call this set-up the "Basic Setting." The brushing was either synchronous or asynchronous. In the asynchronous conditions, the subject was asked to maintain a constant speed of about 2 s per stroke, but the experimenter varied the frequency randomly from 1 to 3 s per stroke. The experimenter also randomly brushed different locations on the back of the subject's left hand, including the fingers and the wrist. Through the HMD, the subject adopted the experimenter's first person perspective (1PP) as if it was his/her own 1PP. We will call this *adopted 1PP*. The subject watched either the experimenter's hand from the adopted 1PP, and/or the subject's own hand from the experimenter's third person perspective (hereafter, *adopted 3PP*, 180°), or the subject's full body from the adopted 3PP (180°, with or without face).

To measure SCR, two single-use foam electrodes (Covidien, Inc., Mansfield, USA) were attached to the participant's left-hand middle finger and fourth finger, on the volar surfaces of the medial phalanges. Data were registered at a sample rate of 200 Hz, and analyzed with Biopac software AcqKnowledge v. 3.7.7. In Experiments 1 and 2, we presented a threat (kitchen knife) at the 90th second. The knife was shown in the scene and approached the stereo camera (i.e., the subject's adopted 1PP). We identified the amplitude of SCR as the difference between the maximal and minimal values of the responses within 5 s after the knife threat. Thus, what we measured was phasic SCR (Dawson et al., 2007). Those subjects who did not show any SCR amplitude were classified as non-responders, and were excluded from the analysis. Totally, we excluded 4 pieces of SCR data. After each experiment, the participant filled out a questionnaire.

Regarding statistical methods, based on many previous studies, we had strong prior expectations that the values measured in the synchronous conditions would be higher than in the asynchronous conditions, i.e., we assumed that $\mu_1(\text{synchronous}) > \mu_2(\text{asynchronous})$. So in Experiments 1 and 2, we used one-tailed *t*-tests to analyze both the questionnaires and SCR data. Then, to compare the sense of body-part ownership and the sense of full-body ownership, we conducted ANOVA and correlation analyses across five conditions selected from Experiments 1~3. Finally, we did a correlation analysis on the data about the sense of experiential ownership.

Experiment 1

In Experiment 1, the participant watched through the HMD the front side of his/her own virtual body, including not only the torso, legs, and face, but also his/her own right hand holding a paintbrush (Figure 1B). This experiment had two goals. The first was to verify whether this setting would create a variant of the full-body illusion, which may provide a new paradigm for studying full-body ownership. Second, we believe that in order to investigate the sense of experiential ownership, the experimental set-up should be arranged such that the subject may interact with another person. In Petkova and Ehrsson (2008), the subject and the experimenter squeezed each other's hands synchronously and, due to manipulation of visual perspective, some participants felt that they were shaking hands with themselves. However, Petkova and Ehrsson's research target was exclusively on the sense of body ownership. Like most studies, the sense of experiential ownership was not measured. Therefore, in Experiment 1 we used the Basic Setting to examine whether the subjective experience of "self-touching" is a solid effect, and we investigated not only the sense of body ownership but also the sense of experiential ownership. In condition 1, we performed synchronous brushing followed by a questionnaire. For the sake of later discussion, we will use "FB1" (Full-body condition 1) to indicate this condition. In condition 2 the brushing was asynchronous. Using the same set-up, in conditions 3 and 4 we measured SCR to provide objective support for conditions 1 and 2 respectively (Table 2).

Experiment 2

In order to exclude the possibility that the phenomena measured by Experiment 1 are merely isolated contingent effects, we

Table 2 | Overview of experiments.

Experiment		Description	Measures taken	Participants (n)	Statistics
Experiment 1	Cond. 1: Sync.	Full body, face, active hand	Questionnaire (FB1)	38 (♂ 21) $M = 21.55 \pm 2.90$	t -test (Sync. Vs. Async)
	Cond. 2: Async.	Full body, face, active hand	Questionnaire	35 (♂ 18) $M = 21.23 \pm 1.66$	
	Cond. 3: Sync.	Full body, face, active hand	SCR	15 (♂ 7) $M = 22.13 \pm 3.14$	t -test (Sync. Vs. Async)
	Cond. 4: Async.	Full body, face, active hand	SCR	15 (♂ 6) $M = 22.47 \pm 3.02$	
Experiment 2	Cond. 1: Sync.	Full body, no face, passive hand	Questionnaire (FB2)	28 (♂ 20) $M = 23.75 \pm 3.63$	t -test (Sync. Vs. Async)
	Cond. 2: Async.	Full body, no face, passive hand	Questionnaire	14 (♂ 11) $M = 23.43 \pm 3.80$	
	Cond. 3: Sync.	Full body, no face, passive hand	SCR	13 (♂ 9) $M = 24.08 \pm 3.52$	t -test (Sync. Vs. Async)
	Cond. 4: Async.	Full body, no face, passive hand	SCR	13 (♂ 10) $M = 23.69 \pm 3.82$	
Experiment 3	Cond. 1: Sync.	Full body, no face, no hand	Questionnaire (FB3)	24 (♂ 15) $M = 21.50 \pm 2.73$	ANOVA
	Cond. 2: Sync.	Full body, no face, active hand	Questionnaire (FB4)	27 (♂ 13) $M = 21.52 \pm 3.33$	ANOVA
	Cond. 3: Sync.	1PP passive hand	Questionnaire (BP1)	25 (♂ 10) $M = 21.48 \pm 2.26$	ANOVA
	Cond. 4: Sync.	1PP passive hand & 3PP active hand	Questionnaire (BP2)	27 (♂ 16) $M = 21.44 \pm 2.69$	ANOVA

applied the Basic Setting and constructed a different full-body condition. In Experiment 2, the participant watched through the HMD the front side of his/her own virtual body, including the torso and legs, but not the face. The participant also saw his/her own left hand being touched by a paintbrush held by the experimenter's hand (**Figure 1C**). This experiment consisted of four conditions as well, and the procedures and measurements were exactly the same as Experiment 1 (**Table 2**). The only difference between Experiments 1 and 2 were the HMD images described above. In Experiment 2, we will call condition 1 “FB2” (Full-body condition 2).

Experiment 3

We used the Basic Setting to conduct two other full-body conditions (FB3 and FB4) and two body-part conditions (BP1 and BP2). In this experiment, only the synchronous conditions were performed and measured by questionnaires. FB3: Through the HMD, the subject saw the front side of his/her own virtual body from below the neck. The subject saw his/her own torso, but not the face and hands (**Figure 1D**). FB4: Through the HMD, the subject saw the front side of his/her own virtual body, including not only the torso and legs, but also his/her own right hand holding a paintbrush (**Figure 1E**). BP1: Through the HMD, the subject saw the experimenter's hand from the adopted 1PP being touched by a paintbrush (**Figure 1F**). BP2: The subject saw two hands via the HMD: the subject's own hand viewed from the adopted 3PP in the opposite direction (180°) holding a paintbrush and brushing the experimenter's

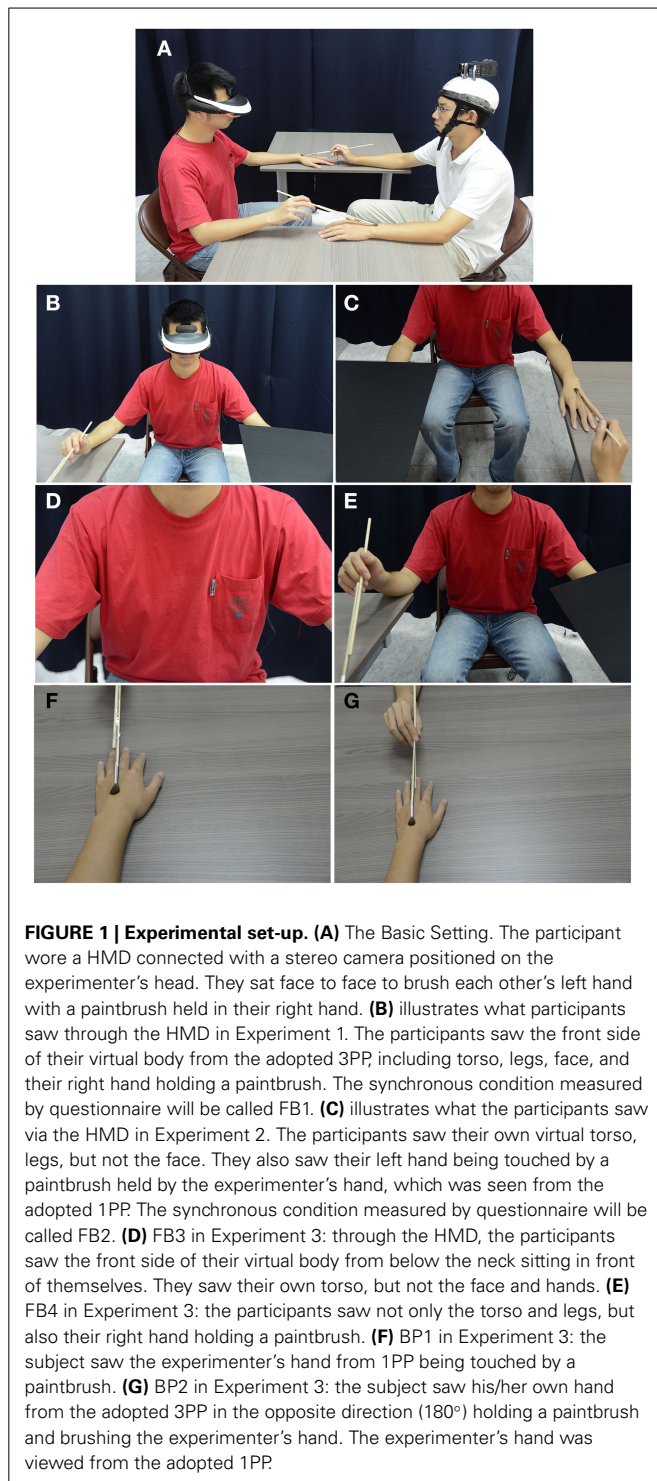
hand. The experimenter's hand was viewed from the adopted 1PP (**Figure 1G**).

RESULTS

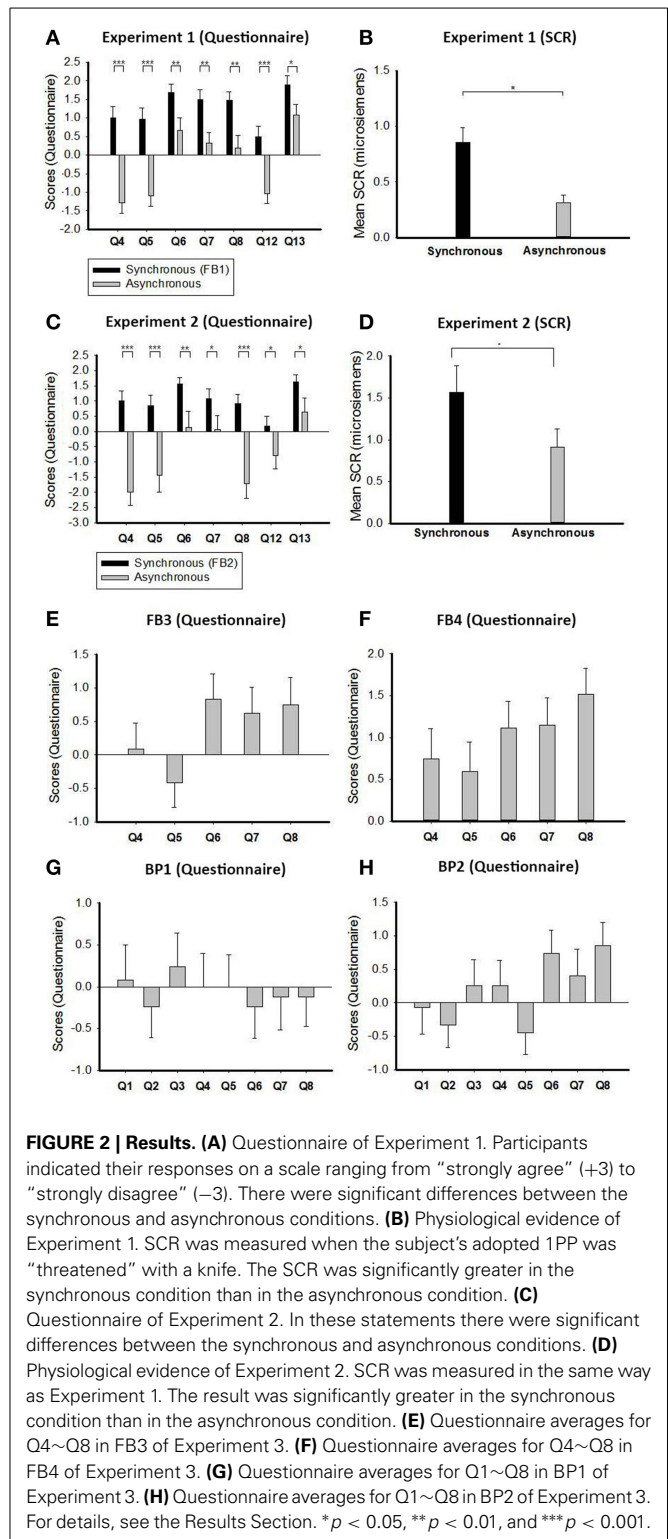
EXPERIMENT 1

We report two key observations from Experiment 1. First, the questionnaire contained statements regarding full-body ownership (Q6), self-location (Q7), full-body agency (Q8), and the double body effect (Q12 and Q13). The average scores on these statements were significantly higher in the synchronous condition (FB1) than in the asynchronous condition (Q6: $p = 0.0073$, Cohen's $d = 0.594$; Q7: $p = 0.0021$, Cohen's $d = 0.706$; Q8: $p = 0.0012$, Cohen's $d = 0.748$; Q12: $p = 0.0001$, Cohen's $d = 0.933$; Q13: $p = 0.0140$, Cohen's $d = 0.533$, independent one-tailed t -test, **Figure 2A**). The SCR measured in conditions 3 and 4 showed the same differences as well ($p = 0.0080$, Cohen's $d = 0.970$, independent one-tailed t -test, **Figure 2B**), which provided objective support for the questionnaire data. This suggests that FB1 successfully induced a new version of the full-body illusion, where the participants felt as if the body in front of them was theirs (Q6) and that they could control it (Q8), and they felt as if they were sitting in front of their own body (Q7). These results are consistent with previous studies (Ehrsson, 2007; Lenggenhager et al., 2007). Finally, they even felt as if they had two bodies (Q12 and Q13, cf. Supplement Materials for more discussion on this effect).

Second, compared with the asynchronous condition, the synchronous full-body condition generated a “self-touching



illusion": the subject felt that "I was brushing my own hand!" This was measured by two questionnaire statements: "It felt as if I was brushing my own hand" (Q4), and "The one whom I brushed was me, not someone else" (Q5). Since Q4 and Q5 involve both hand-touching and self-identification (Blanke and Metzinger, 2009), they are associated not only with body-part but also full-body representations. Statistics showed significant



differences between the synchronous and asynchronous conditions (Q4: $p < 0.0010$, Cohen's $d = 1.301$; Q5: $p < 0.0010$, Cohen's $d = 1.168$, independent one-tailed t -test, Figure 2A), and the SCR results provided objective evidence for this new type of full-body illusion (Figure 2B). This supports that the

self-touching illusion is a distinctive version of the full-body illusion.

EXPERIMENT 2

Using the same questionnaire in Experiment 2, we found that the average scores for full-body ownership (Q6), self-location (Q7), full-body agency (Q8), and the double body effect (Q12 and Q13) were significantly higher in the synchronous condition (FB2) than in the asynchronous condition (Q6: $p = 0.0023$, Cohen's $d = 1.009$; Q7: $p = 0.0457$, Cohen's $d = 0.581$; Q8: $p < 0.0010$, Cohen's $d = 1.675$; Q12: $p = 0.0446$, Cohen's $d = 0.585$; Q13: $p = 0.0171$, Cohen's $d = 0.735$, independent one-tailed t -test, **Figure 2C**). Also, the SCR values were significantly higher in the synchronous condition than in the asynchronous condition ($p = 0.0473$, Cohen's $d = 0.711$, independent one-tailed t -test, **Figure 2D**). This indicates that, like FB1 above, FB2 can induce a version of full-body illusion as well. These results nicely collaborate with the data collected from Experiment 1, suggesting that there are in fact many ways to induce full-body illusions.

Second, just like Experiment 1, the synchronous condition in Experiment 2, i.e., FB2, caused the subject to experience the self-touching illusion: the participants felt as if they were brushing their own hand. We observed significant differences between the synchronous and asynchronous conditions on Q4 and Q5 (Q4: $p < 0.0010$, Cohen's $d = 1.821$; Q5: $p = 0.0003$, Cohen's $d = 1.236$, independent one-tailed t -test, **Figure 2C**) and on the SCR values (**Figure 2D**). The data for the sense of experiential ownership will be presented later. Together with the results from Experiment 1, we confirm that the self-touching illusion is a solid effect.

EXPERIMENT 3

Figures 2E,F show the questionnaire data of the other two full-body conditions, FB3 and FB4. **Figures 2G,H** present the questionnaire data of the two body-part conditions, BP1 and BP2. We will see that, by combining the data from these and other synchronous conditions, an important lesson can be drawn regarding the relationship between the sense of body-part ownership and full-body ownership.

THE SENSE OF BODY OWNERSHIP

One distinct feature of our study is that totally we carried out six synchronous body-part and full-body conditions, which are more than previous studies (Ehrsson, 2007; Lenggenhager et al., 2007; Petkova and Ehrsson, 2008). Another feature is that we asked self-touching questions (Q4 and Q5) and full-body questions (Q6, Q7, and Q8) both in the body-part conditions (BP1 and BP2) and in the full-body conditions (FB1~FB4). These features allow us to compare the participants' responses in many different conditions. We hypothesized that the illusory sense of full-body ownership would gradually increase from the body-part conditions to the full-body conditions. To test this hypothesis, we used ANOVA to analyze the questionnaire data on Q5~Q8 across the following series of conditions: BP1, BP2, FB3, FB4, and FB1. The order of this series was determined by the scopes that the participants saw via the HMD, which systematically increase from the body-part to the full-body conditions: BP1 (passive hand only),

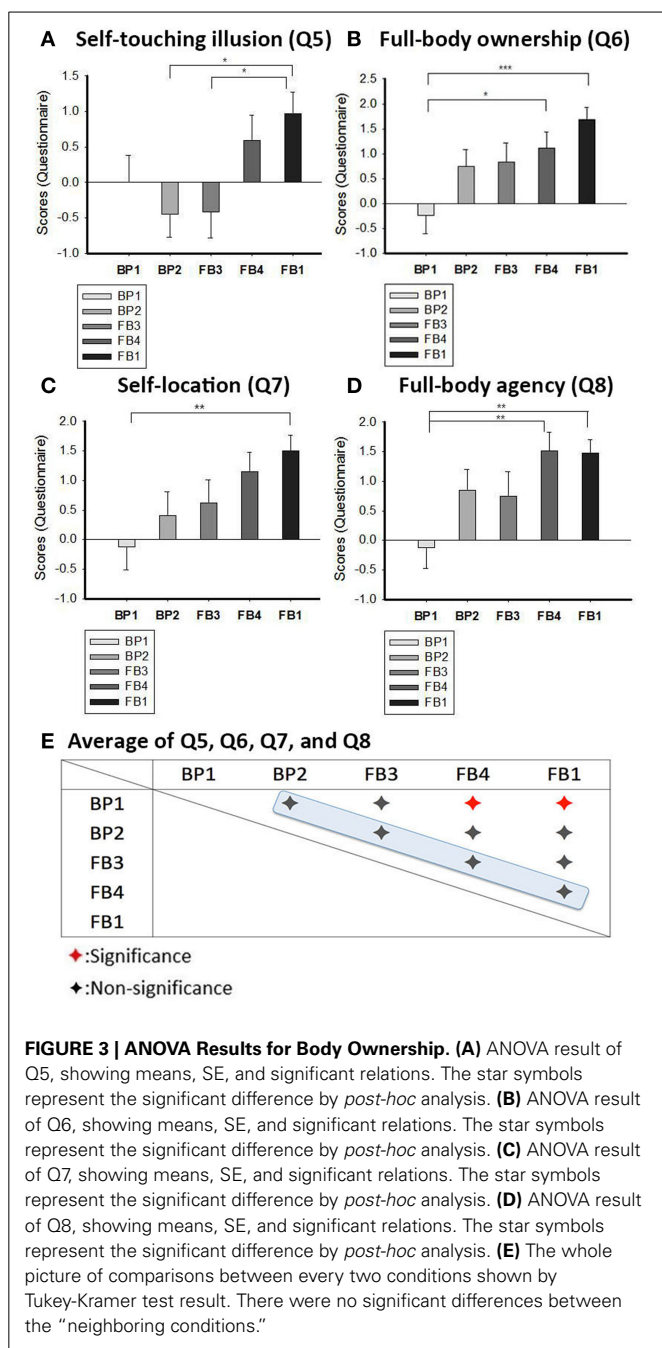
BP2 (both passive hand and active hand), FB3 (only torso), FB4 (torso and active hand) and FB1 (torso, active hand and face) (**Figures 1B,D–G**). We can see that each condition involves just one more factor than the one on its left (except for the minimum full-body condition FB3 compared with BP2) (**Table 2**). FB2 was not included in this analysis because the hand seen via the HMD was not on the same side compared with FB1 and FB4. We chose Q5~Q8 because they are all associated with the sense of full-body ownership, which was also why Q4 was not included. In addition to the hypothesis just mentioned, we also like to know whether significant differences will exist only between the two poles (or near the two poles) of the series, i.e., whether there will be no significant differences between any two conditions that appear next to each other in the series.

We conducted an ANOVA analysis on Q5~Q8 to see how the answers varied across conditions. Then we did *post-hoc* analyses to know how the significances are distributed. Regarding Q5 [$p = 0.008$, $F_{(4, 136)} = 3.625$, $\eta^2 = 0.096$, ANOVA], significant differences existed between FB1 (mean = 0.974, $SD = 1.852$) and BP2 (mean = -0.444 , $SD = 1.717$) ($p = 0.020$, Tukey-Kramer test), and between FB1 and FB3 (mean = -0.417 , $SD = 1.792$) ($p = 0.032$, Tukey-Kramer test) (**Figure 3A**). Regarding Q6 [$p = 0.001$, $F_{(4, 136)} = 5.044$, $\eta^2 = 0.129$, ANOVA], there are significant differences between FB1 (mean = 1.684, $SD = 1.378$) and BP1 (mean = -0.240 , $SD = 1.877$) ($p < 0.001$, Tukey-Kramer test), and between FB4 (mean = 1.111, $SD = 1.695$) and BP1 ($p = 0.037$, Tukey-Kramer test) (**Figure 3B**). Regarding Q7 [$p = 0.009$, $F_{(4, 136)} = 3.514$, $\eta^2 = 0.094$, ANOVA], significant differences existed only between FB1 (mean = 1.500, $SD = 1.640$) and BP1 (mean = -0.120 , $SD = 1.986$); the p -value of the Tukey-Kramer test was 0.008 (**Figure 3C**). Finally, for Q8 [$p = 0.003$, $F_{(4, 136)} = 4.219$, $\eta^2 = 0.110$, ANOVA], we observed significant differences between FB1 (mean = 1.474, $SD = 1.428$) and BP1 (mean = -0.120 , $SD = 1.787$) ($p = 0.004$, Tukey-Kramer test), and between FB4 (mean = 1.519, $SD = 1.602$) and BP1 ($p = 0.006$, Tukey-Kramer test) (**Figure 3D**). These results support our hypothesis that significant differences are observed only between the two poles (or near the two poles) of the series, i.e., there are no significant differences between any two conditions that stand next to each other in the series (**Figure 3E**).

We also did a correlation analysis on Q5~Q8 across the above five conditions, taking those conditions as a nominal variable X , and the scores of Q5~Q8 as a continuous variable Y . We found that there was a weak positive correlation between the two variables. Here are the Spearman's ρ for each of Q5~Q8: (Q5, $\rho = 0.255$) (Q6, $\rho = 0.342$) (Q7, $\rho = 0.309$) (Q8, $\rho = 0.295$). Also, the Spearman's ρ between the five conditions and the average of Q5~Q8 was low as well ($\rho = 0.341$, **Figure 4A**). All of the correlations here are significant ($p < 0.01$). Again, these results support our hypotheses that, although the illusory sense of full-body ownership gradually increases from body-part to full-body conditions, the differences between the "neighboring conditions" are not significant.

THE SENSE OF EXPERIENTIAL OWNERSHIP

In this study, two statements were designed precisely to measure the participants' sense of experiential ownership: "It was me who



felt being brushed, not someone else” (Q9), and “The one who felt being brushed was not me” (Q10). Notice that these two statements are directly opposite to each other. Also, they are not about the sense of body ownership, but about *who* felt the tactile sensations caused by brushing. We focused on FB1 and FB2, since they are supported by SCR measurements. That means the results of FB1 and FB2 revealed the participants’ subjective experiences rather than just their judgments in the questionnaires. During the experiments the participants were touched by a paintbrush, so they were indeed the subjects of those tactile sensations. This fixed the *fact* of their experiential ownership. The task was to

examine whether this fact was correctly represented by their *sense* of experiential ownership. We found that, in FB1 and FB2, the average scores on Q9 were 1.58 and 1.04 respectively (**Figure 4B**). Also, 32% of the subjects in FB1 answered (−1), (0), or (+1) on Q9, and 36% did so in FB2. The standard deviation of Q9 in FB1 was 1.50, and in FB2 it was 1.55. As an opposite statement, the average scores on Q10 in FB1 and FB2 were −1.03 and −0.50 respectively (**Figure 4B**). While 13.2% of participants in FB1 and FB2 disagreed with Q9 (i.e., they answered either −1, −2, or −3), 18.4% of them agreed with Q10. **Figure 4C** indicates that the negative correlation between these two sets of results is low (coefficient $R = -0.3278$). Later we will discuss the impact of these data on IEM and on the sense of experiential ownership.

DISCUSSION

BODY OWNERSHIP

In our experiments, the participants not only received tactile stimulations but also held a paintbrush to touch someone else’s hand. Thus, agency was clearly involved. Moreover, via the HMD, the subjects saw their own full body facing themselves (**Figures 1B–E**). This set-up was quite different from that in Lenggenhager et al. (2007) and Ehrsson (2007) where the participants watched their own virtual body from the back. Also, these previous studies did not involve agency; the participants only received visual-tactile stimulations either from the back (Lenggenhager et al., 2007) or in the chest (Ehrsson, 2007). Our set-up was more similar to one of the body swap experiments by Petkova and Ehrsson (2008), where the participant and the experimenter faced each other and squeezed each other’s hands (cf. their Figure 6). Still, our set-up differed from theirs in that we combined a subject-experimenter interaction with the RHI. This experimental strategy—incorporating elements of body-part illusions with full-body illusions—has not been used until recently (Olivé and Berthoz, 2012; van Doorn et al., 2014). We hypothesized that the sense of body-part ownership and the sense of full-body ownership are not essentially distinct from each other. This is supported by the results of our ANOVA *post-hoc* analyses and correlation analyses reported above (**Figures 3A–E, 4A**). Since we address this issue only at the psychophysical level, we do not claim that our hypothesis would automatically apply at the neurophysiological level. Still, this hypothesis is useful as it can serve as a research guide or theoretical constraint for enquiries into body mereology and the relevant neural mechanisms (Petkova et al., 2011b).

Most RHI studies agree that multisensory stimulations and integration are important for explaining the illusory sense of body ownership. According to the bottom-up approach, the RHI is caused by interactions between vision, touch and proprioception (Botvinick and Cohen, 1998; Ehrsson, 2012). In contrast, the top-down approach suggests that the synchronous multisensory stimulations and integration are necessary but not sufficient for the RHI (Tsakiris and Haggard, 2005; Tsakiris, 2010, 2011). A pre-existing representation of body is required to explain various aspects of the RHI, such as the visual form congruency, anatomical congruency, postural congruency, etc., between the viewed fake hand and the felt body-part. Tsakiris proposes a model that explains the RHI in terms of three critical comparisons: “First,

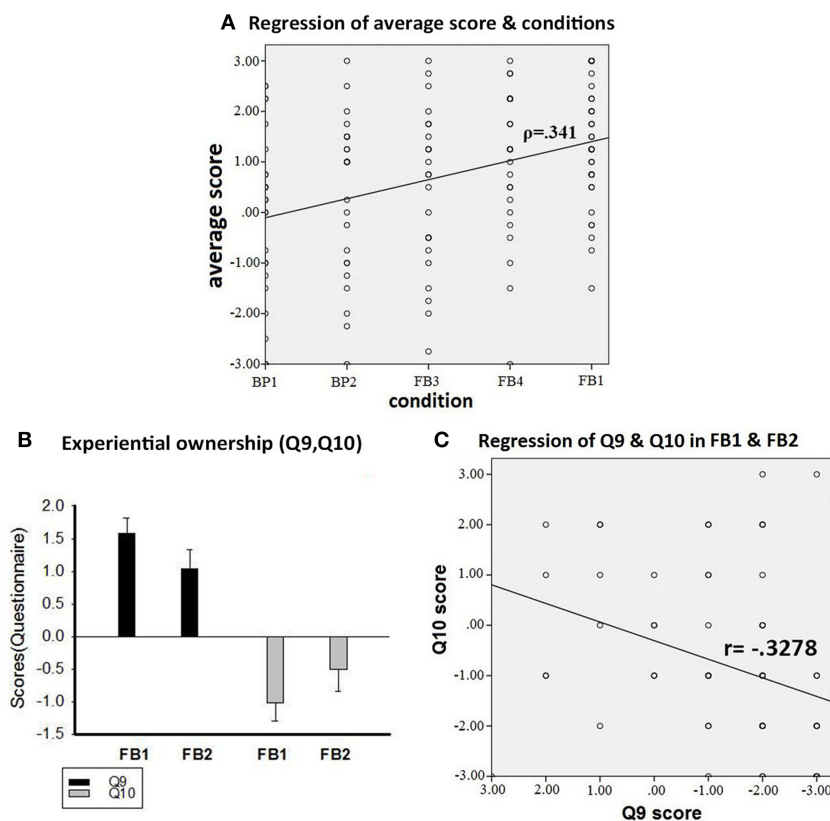


FIGURE 4 | Results of Correlation and Results of Experiential Ownership.

(A) Correlation between the average of Q5~Q8 and the condition variable, showing only a weakly positive correlation ($p = 0.341$). Spearman's ρ stands for the correlation coefficient. (B) Result of Q9 and Q10 in FB1 and FB2. Q9: "It was me who felt being brushed, not someone else." Q10: "The person

who felt being brushed was not me." The scores were not predicted by IEM. The bars represent the mean values of the physiological scale, and the error bars indicate standard error. (C) indicates that the negative correlation between the scores of Q9 and Q10 in FB1 and FB2 was low (Pearson's $R = -0.3278$). For details, see the Results section.

a pre-existing stored model of the body distinguishes between objects that may or may not be part of one's body. Second, on-line anatomical and postural representations of the body modulate the integration of multisensory information that leads to the recalibration of visual and tactile coordinate systems. Third, the resulting referral of tactile sensation will give rise to the subjective experience of body-ownership" (2010, p. 703). Not everyone agrees that there has to be a fixed body model in order to explain body ownership (Guterstam et al., 2011). We stay neutral on the debate between the bottom-up and top-down approaches, and would just like to mention that both approaches share the view that the subject's 1PP is essential for the RHI to be induced.

In contrast, the role of 1PP is more controversial in the research of full-body ownership. It has been pointed out that two rather different types of set-up have been used by researchers. As Petkova et al. (2011a, p. 2) indicate, the virtual body was either viewed by the subject from the adopted 3PP "as though looking at another individual a couple of meters in front of oneself" (Lenggenhager et al., 2007, 2009; Aspell et al., 2009), or from the adopted 1PP "as though directly looking down at one's body" (Petkova and Ehrsson, 2008; Petkova et al., 2011a; van der Hoort et al., 2011). An issue then arises concerning which set-up is more

appropriate. Petkova et al. (2011a) suggested that viewing the virtual body from the subject's (adopted) visual 1PP is absolutely crucial for full-body illusions to be induced. They made the following criticism of the 3PP set-up used by Lenggenhager et al. (2007): since the virtual body was seen from 3PP and the situation is more like recognizing oneself on a surveillance monitor, what happened to the participants could be just a visual self-recognition "without necessarily experiencing a somatic illusion of ownership in the same way as in the rubber hand illusion or in the body-swap illusion" (Petkova et al., 2011a, p. 5). That is, it is possible that the participants in Lenggenhager et al. (2007) did not really experience a genuine full-body illusion.

In our full-body conditions, the virtual body was also viewed from the adopted 3PP. But we think that the criticism by Petkova et al. (2011a) can be replied to by two aspects of our experiments. First, the questionnaires in FB1 and FB2 were supported by the SCR measurements in Experiments 1 and 2, where a kitchen knife approached the subject's (adopted) visual 1PP. Although the threat was not applied to the virtual body, the SCR data collaborate well with the second aspect: we included statements about the self-touching illusion (Q4, Q5) and the double-body effect (Q12, Q13). As reported above, both the SCR and questionnaires

show that significant differences exist between the synchronous and asynchronous conditions. Together, these data indicate that the participants in FB1 and FB2 experienced not only the illusory sense that they were brushing themselves, but also that they had two bodies. This goes beyond mere visual self-recognition and suggests that the relevant full-body illusions were genuinely induced.

In addition to tactile sensations, proprioception and visual 1PP, we think that there are two more factors which come into play: “visual form congruency” and “visual agency.” As we go back to consider the data revealed in **Figures 3A–D**, we will see that these two factors often have greater influences on the sense of body ownership than the subject’s visual 1PP.

Visual form congruency

In our full-body conditions, the participants watched their body facing themselves. In such cases, visual form congruency refers to the scope of what the subject saw via the HMD. The scope of HMD images enlarges gradually and systematically from FB3, FB4 to FB1, which positively correlates with the strength of the relevant full-body illusions (**Figures 3A–D**). Although the virtual body was always presented from the adopted 3PP, this, as we have just argued above, would not necessarily hinder the relevant full-body illusions. In our body-part conditions, visual form congruency concerns whether the hand or hands that the participants saw via the HMD looked like their own. According to the first comparison in Tsakiris’ model of body ownership, “the more the viewed object matches the structural appearance of the body-part’s form, the stronger the experience of body-ownership will be” (2010, p. 707). We agree. Seeing one’s own hand via a HMD satisfies both visual form congruency and Tsakiris’ first comparison. However, as will be discussed later, our data may challenge the second comparison of Tsakiris’ model, which concerns postural congruency. According to this comparison, “If there is incongruency between the posture of felt and seen hands, the seen hand will not be experienced as part of one’s own body” (2010, p. 708). In our experiments, when the subject saw the experimenter’s hand via a HMD, it was always presented from the subject’s adopted 1PP. On the other hand, when the subject saw his/her own hand, it was always presented from the subject’s adopted 3PP. We will soon consider whether postural incongruency can be remedied or outweighed by visual form congruency.

Visual agency

We suggest distinguishing between “body agency” and “visual agency.” Body agency refers to the subject feeling his/her own act of brushing via proprioception. It has been shown that body agency can either diminish or facilitate the RHI (Tsakiris et al., 2006; Kalckert and Ehrsson, 2012). What we would like to add is that agency can play a role in bodily self-consciousness, not only by being felt but also by being *seen*. Visual agency refers to the brushing activity that the participants *saw* through the HMD. We further suggest distinguishing between “1PP visual agency” and “3PP visual agency.” 1PP visual agency refers to the participants seeing the act of brushing by the experimenter’s hand from the adopted 1PP, while 3PP visual agency refers to the participants

seeing the act of brushing by their own hand from the adopted 3PP.

These two factors—visual form congruency and visual agency—can help explain the sense of body ownership involved in our experiments. Consider the series of conditions that we analyzed in the Result section: BP1, BP2, FB3, FB4, and FB1. These conditions involved the same tactile sensations, proprioception and body agency. What distinguished between them were visual form congruency and 3PP visual agency. Both the ANOVA and correlation analyses on Q6~Q8 showed that the illusory sense of full-body ownership gradually increased from the minimal body-part condition BP1 to the maximum full-body condition FB1 (**Figures 3B–D**). This can be nicely explained by the following comparisons: (1) compared with BP1, BP2 involves 3PP visual agency as an extra factor; (2) FB3 lacks 3PP visual agency but has a stronger visual form congruency than BP2; (3) FB4 contains not only 3PP visual agency but also a stronger visual form congruency than FB3; and (4) compared with FB4, FB1 involves an even stronger visual form congruency, i.e., seeing the subject’s own face. As mentioned above, Q5 is about self-identification, and the scores across conditions seem to form a low-group (BP1, BP2, FB3) and a high-group (FB4, FB1) (**Figure 3A**). We suspect that this indicates that both 3PP visual agency and a strong visual form congruency are required for an illusory sense of self-identification. Notice that, even so, there are no significant differences between neighboring conditions in the series (**Figure 3A**). Finally, FB4 and FB1 have almost the same scores on Q8 (**Figure 3D**). We think this is because the factor of 3PP visual agency was the same in these two conditions. Since Q8 is about full-body agency, it is expected that 3PP visual agency would be more important than visual form congruency.

To conclude this part of the discussion, we have suggested that (1) visual form congruency can sometimes outweigh postural incongruency, which implies that the second comparison in Tsakiris’ model can sometimes be violated. When there is strong visual form congruency, full-body illusions can still be induced in the face of postural incongruency; and (2) the distinction between body agency and visual agency, and the further distinction between 1PP visual agency and 3PP visual agency can help explain how body-part and full-body illusions may be hindered or facilitated.

EXPERIENTIAL OWNERSHIP

As mentioned above, the current mainstream view of the sense of experiential ownership is heavily influenced by Wittgenstein (1958) and Shoemaker’s IEM (1968). Recall that IEM is the thesis that when I am aware of a phenomenal state through first-personal access I *cannot be wrong* about whether it is I who feels it. As mentioned in the Introduction, most of its defenders consider it as a conceptual truth based on language use. There are, therefore, very few empirical discussions on IEM and on the sense of experiential ownership (Legrand, 2007; Gallagher, 2012).

One of few exceptions was by Mizumoto and Ishikawa (2005) where the authors used a full-body illusion to argue against IEM. However, the authors described that “the subject ... unanimously (all four subjects who participated in this particular experiment) reported that he ‘felt’ as if the body he was watching was his,

although he in fact knew that it was not" (2005, p. 8). They also remarked that "What we have shown is the *possibility*, not the necessity, of the subject's mistakenly reacting to the attack to the other's body, which confirms our hypothesis that they felt as if they were *there* being tapped *in* the visual frame, while in fact they were not" (2005, p. 9, the authors' italics). The problem is: in our terms, they characterized their version of full-body illusion as concerning the sense of full-body ownership (and touch referral) rather than the sense of experiential ownership. As mentioned in the Introduction, IEM is about the latter, not the former. The difference between the two was nicely illustrated by two patients described by Moro et al. (2004) who denied ownership of their left hand, in which they had no sensation, and lost their left visual field. When their left hand was moved to the right so that they could see it, they became capable of detecting tactile sensations. But despite representing themselves as the ones who felt the sensations, the two patients still denied the ownership of their left hands. This shows that it is possible to have the sense of experiential ownership without the sense of body ownership.

Here, we discuss an interdisciplinary approach that defends IEM based on the phenomenological structure of experience that we call the *Pre-reflective Account* (Gallagher, 2005, 2012; Zahavi, 2005; Legrand, 2006, 2007, 2010). According to this account, the sense of self-as-subject is not achieved through introspection, judgment or attention. At the pre-reflective level, the sense of self-as-subject is a constitutive component of conscious state rather than an intentional object of consciousness. This makes the sense of self-as-subject identification-free, i.e., it does not involve identification of self as the subject, and hence enjoys IEM (Legrand, 2007; Gallagher, 2012). When I am pre-reflectively conscious of myself-as-subject, I *cannot* be wrong about whether I am the subject of experiences. We will call this view *pre-reflective immunity*. Like Shoemaker's version of IEM, pre-reflective immunity asserts a very strong modal claim. It states that violation of IEM is not possible.

Now, based on our data reported in the Results section, we argue that the sense of self-as-subject does *not* enjoy IEM, i.e., violation of IEM is possible. It is possible for misrepresentation to occur in one's pre-reflective sense of experiential ownership. If so, pre-reflective immunity does not hold. Below we show that the data of our experiments do not lend any support to Shoemaker's IEM at all. The best interpretation suggests that misrepresentation can occur in one's sense of experiential ownership. Then we respond to a possible objection to our position from the standpoint of the Pre-reflective Account.

Part of the reason why this is a knotty issue concerns how the participants understood Q9 and Q10. For the sake of argument, we will consider different possibilities: (I) Suppose the participants understood Q9 as addressing themselves. That is, from their subjective point of view: it was *me* who felt the brushing. Then, according to IEM, no participants would commit mistakes regarding their sense of experiential ownership. One would expect that most participants would answer "strongly agree" (+3) or at least "agree" (+2) on Q9. But that is not the case. 13.2% of participants in FB1 and FB2 disagreed with Q9 (i.e., they answered either -1, -2, or -3). The average scores of Q9 were much lower than this interpretation requires (**Figure 4B**). (II) Suppose for

some reason that the participants understood Q9 as addressing someone else. That is, on their subjective experiences: it was *not me* who felt the brushing. Then, according to IEM, one would expect that most participants would answer "strongly disagree" (-3) or at least "disagree" (-2) on Q9. But this is not the case, either. This time, the average scores of Q9 are too high to fit this interpretation (**Figure 4B**). (III) Suppose that the participants did not all understand Q9 in the same way; some took it as addressing themselves, but others as addressing someone else. Then, assuming IEM holds, one would expect the participants to answer either +3 (or at least +2) or -3 (or at least -2). But, again, that is not the case. As reported in the Results section, many participants answered "slightly disagree" (-1), "not sure" (0), or "slightly agree" (+1). In fact, the standard deviation in each experiment is large (FB1 *SD* = 1.50, FB2 *SD* = 1.55), suggesting that the participants' responses to Q9 varied widely.

The point here is that none of the above interpretations can support IEM. Based on the data, it is more plausible that at least some participants in these experiments were uncertain about whether they were the subjects of the tactile sensations that they actually felt. This uncertainty could very well take place at the *pre-reflective level*. That is, the fact of receiving tactile sensations does not guarantee that the participants will necessarily have the *sense* that "I am the one who felt them." There is no empirical evidence against our position here, and that our interpretation can better accommodate why the participants did not respond to Q9 in the way that conforms to IEM. The data provide empirical evidence for the possibility that one's sense of experiential ownership can misrepresent the relevant fact of experiential ownership. Hence, IEM could potentially be falsified.

The defender of pre-reflective immunity would probably reject all the above interpretations and argue that our data can be explained in a way that does not jeopardize IEM. It might be that, due to the unusual experience of the self-touching illusion, not only did different participants understand Q9 ("It was me who felt being brushed, not someone else") differently, but also many of them were unsure about how to respond to it. The defense is that, no matter what answers the participants gave on Q9, it remains that they were the actual subjects of the tactile sensations that they felt. The variety of their answers only reveals the uncertainty of their judgments, not the uncertainty of their sense of experiential ownership or what Gallagher (2012) calls their pre-reflective 1PP. Even if some of their judgments were wrong, the mistakes were at the reflective level, not at the pre-reflective level.

Here are our responses. First, it is one thing that the participants have a pre-reflective 1PP; it is another whether they might be mistaken about that perspective. Having a pre-reflective 1PP only secures the *fact* of the participants' experiential ownership. It should not be taken for granted that this fact cannot be misrepresented by their pre-reflective *sense* of experiential ownership. Second, all the participants in our experiments were healthy subjects. There are no compelling reasons why their judgments cannot reveal their pre-reflective sense of experiential ownership. Even if they were uncertain about whom Q9 was addressing and hence were less confident about the judgments they made, this could well be an indication that at the pre-reflective level they were unsure (and hence prone to error) about who the subject

of the sensations was. Finally, in addition to Q9, we also presented Q10 (“The one who felt being brushed was not me”) in the questionnaires. The direct contrast between Q10 and Q9 was so obvious that, even if the participants felt uncertain about Q9, the contrast can still be easily recognized. So, if IEM holds, one can reasonably expect that the participants’ responses would manifest a strong “negative correlation” between Q9 and Q10. For example, if a subject answers +3 to Q9, then he/she would likely answer −3 (or at least −2) to Q10, etc. However, the data show no such correlation (**Figure 4C**).

The simple and best explanation of the data, we suggest, is that at least some of the participants were unsure or mistaken about who the subject was—even at the *pre-reflective level*. We can agree that (1) For every tactile sensation there must be a subject who experiences it; (2) Every tactile sensation is necessarily experienced from the subject’s 1PP; and (3) Every tactile sensation is experienced by the one who has the 1PP of that state. However, (1)~(3) together do not imply (4): Every tactile sensation is represented from the first-person point of view *as* experienced by the one who has the 1PP of that state. The key is that (3) and (4) are not equivalent: feeling tactile sensations is one thing, but whether one experiences oneself *as* the subject of those sensations could be another. It is empirically possible that (4) was not obeyed in FB1 and FB2. This shows that violation of IEM is indeed possible.

Another possible defense of pre-reflective immunity appeals to recent studies on the second-person perspective (2PP) and social cognition (Fuchs, 2013; Froese et al., 2014). Since the experimental set-ups in FB1 and FB2 involved two people brushing each other, perhaps the brushing experience was a *social* one. A better description of the participant’s experience would be: “It was *we* who felt being brushed by each other.” This can accommodate why some participants disagreed with Q9: although they agreed that “It was me who felt being brushed,” they disagreed with the latter part of Q9 “not someone else,” since there was indeed someone else, i.e., the experimenter, who felt being brushed as well. Thus, our data can be explained by the involvement of the participants’ 2PP rather than by misrepresentation of their pre-reflective sense of experiential ownership.

Since we focused on the mainstream view about the sense of self-as-subject and IEM, our questionnaires did not take the second-person perspective (2PP) into account. We agree that, in future studies, it would be interesting to add 2PP statements into the questionnaires to see how the subjects respond to them. Having said this, we will make the following remarks in our defense. First, suppose *some* participants disagreed with Q9 because of the 2PP considerations, this does not mean we can be certain that *all* of those who did so were due to the same reasons. Since we argue only that IEM could potentially be falsified, this stance seems to remain intact. Second, suppose some participants’ sense of experiential ownership involved a 2PP as well as a pre-reflective 1PP, and suppose that their rejection of (or uncertainty about) Q9 can be explained by the 2PP interpretation. Can we be sure that *therefore* their pre-reflective sense of experiential ownership was *necessarily* correct? Given our experiments and argument, it would require more evidence for the 2PP account to really save pre-reflective immunity from our attack. Finally, our study shows that sometimes it does make sense to ask “are you

sure it is you who feel the sensations?” We think that introducing the social question “It was *we* who felt being brushed by each other” into the investigation will make it even more significant to pursue the Wittgenstein-style questions.

To conclude this part of the discussion, we have proposed a simple account three paragraphs above—(1)~(3) do not imply (4)—that challenges the mainstream view about the sense of experiential ownership. According to this account, the fact of experiential ownership can be misrepresented by the subject’s pre-reflective sense of experiential ownership. Therefore, we believe that the current best evidence undercuts the empirical basis of pre-reflective immunity.

CONCLUSION

We have suggested that the sense of body ownership and the sense of experiential ownership are different types of bodily self-consciousness. Regarding the former, we have proposed that (1) the self-touching illusion is a solid effect; and (2) there is no fundamental difference between the sense of body-part ownership and the sense of full-body ownership. Regarding the sense of experiential ownership, we have argued that (1) the fact of experiential ownership can be misrepresented by the subject’s pre-reflective sense of experiential ownership; and (2) both Wittgenstein and Shoemaker could very well be wrong: sometimes it makes sense to ask the Wittgenstein-style questions (Q9 and Q10); it is probable that IEM as well as pre-reflective immunity fail to hold. Our study has a positive implication: not only the sense of body ownership but also the sense of experiential ownership allows and calls for interdisciplinary studies. Two important issues require further investigation. First, what is the relationship between the sense of body ownership and the sense of experiential ownership? Our current thought is that the former presupposes the latter. The idea is that when a participant experiences a body-part or a whole body as his/her own, it is relevant to consider whether the participant also represents him/herself *as* the subject of this experience of body ownership. Hence we hypothesize that the sense of experiential ownership is a constitutive component of the sense of body ownership. Further inquiries will be required to test this hypothesis. Second, what are the neural mechanisms that underlie these two types of bodily self-consciousness? Many works have been done regarding the sense of body ownership (Tsakiris, 2010; Ionta et al., 2011; Blanke, 2012; Ehrsson, 2012; Serino et al., 2013). In contrast, we currently know very little about the neural basis of the sense of experiential ownership (Christoff et al., 2011). We believe that the self-touching paradigm and the Wittgenstein-style questions that we developed can contribute to the future research on this issue.

AUTHOR CONTRIBUTIONS

Caleb Liang designed all experiments, Si-Yan Chang, Wen-Yeo Chen, Hsu-Chia Huang, and Yen-Tung Lee conducted the experiments and analyzed the data, Caleb Liang wrote the manuscript.

ACKNOWLEDGMENTS

The authors would like to thank Jakob Hohwy, Jennifer Windt and the audience at the ASSC 18 for their useful comments. We also like to thank professor Chen-gia Tsai from the Graduate Institute

of Musicology for the SCR equipment. Finally, this study was supported by Taiwan's Ministry of Science and Technology (project: MOST100-2628-H-002-133-MY3).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fpsyg.2014.01591/abstract>

REFERENCES

- 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
- Bertamini, M., Berselli, N., Bode, C., Lawson, R., and Wong, L. T. (2011). The rubber hand illusion in a mirror. *Conscious. Cogn.* 20, 1108–1119. doi: 10.1016/j.concog.2011.04.006
- 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., and Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature* 391, 756. doi: 10.1038/35784
- Christoff, K., Cosmelli, D., Legrand, D., and Thompson, E. (2011). Specifying the self for cognitive neuroscience. *Trends Cogn. Sci.* 15, 104–112. doi: 10.1016/j.tics.2011.01.001
- Coliva, A. (2006). Error through misidentification: some varieties. *J. Philos.* 103, 403–425.
- Dawson, M., Schell, A., and Filion, D. (2007). "The electrodermal system," in *The Handbook of Psychophysiology*, eds J. Cacioppo, L. Tassinary, and G. Berntson (Cambridge: Cambridge University Press), 152–191.
- Ehrsson, H. H. (2007). The experimental induction of out-of-body experiences. *Science* 317, 1048. doi: 10.1126/science.1142175
- Ehrsson, H. H. (2009). How many arms make a pair? Perceptual illusion of having an additional limb. *Perception* 38, 310–312. doi: 10.1068/p6304
- Ehrsson, H. H. (2012). "The concept of body ownership and its relation to multisensory integration," in *The New Handbook of Multisensory Processes, Chapter 43*, ed B. E. Stein (Cambridge, MA: MIT Press), 775–792.
- Froese, T., Iizuka, H., and Ikegami, T. (2014). Using minimal human-computer interfaces for studying the interactive development of social awareness. *Front. Psychol.* 5:1061. doi: 10.3389/fpsyg.2014.01061
- Fuchs, T. (2013). The phenomenology and development of social perspectives. *Phenomenol. Cogn. Sci.* 12, 655–683. doi: 10.1007/s11097-012-9267-x
- Gallagher, S. (2005). *How the Body Shapes the Mind*. Oxford: Clarendon Press.
- Gallagher, S. (2012). "First-person perspective and immunity to error through misidentification," in *Consciousness and Subjectivity*, eds S. Miguens and G. Preyer (Frankfurt: Philosophical Analysis Ontos Publisher), 187–214.
- Guterstam, A., and Ehrsson, H. H. (2012). Disowning one's seen real body during an out-of-body illusion. *Conscious. Cogn.* 21, 1037–1042. doi: 10.1016/j.concog.2012.01.018
- Guterstam, A., Petkova, V. I., and Ehrsson, H. H. (2011). The illusion of owning a third arm. *PLoS ONE* 6:e17208. doi: 10.1371/journal.pone.0017208
- Heydrich, L., Dodds, T. J., Aspell, J. E., Herbelin, B., Bühlhoff, H. H., Mohler, B. J., et al. (2013). Visual capture and the experience of having two bodies—evidence from two different virtual reality techniques. *Front. Psychol.* 4:946. doi: 10.3389/fpsyg.2013.00946
- Ionta, S., Heydrich, L., Lenggenhager, B., Mouthon, M., Fornari, E., Chapuis, D., et al. (2011). Multisensory mechanisms in temporo-parietal cortex support self-location and first-person perspective. *Neuron* 70, 363–374. doi: 10.1016/j.neuron.2011.03.009
- Kalckert, A., and Ehrsson, H. H. (2012). Moving a rubber hand that feels like your own: a dissociation of ownership and agency. *Front. Hum. Neurosci.* 6:40. doi: 10.3389/fnhum.2012.00040
- Kammers, M. P. M., de Vignemont, F., Verhagen, L., and Dijkerman, H. C. (2009). The rubber hand illusion in action. *Neuropsychologia* 47, 204–211. doi: 10.1016/j.neuropsychologia.2008.07.028
- 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
- Legrand, D. (2007). Pre-reflective self-as-subject from experiential and empirical perspectives. *Conscious. Cogn.* 16, 583–599. doi: 10.1016/j.concog.2007.04.002
- Legrand, D. (2010). "Myself with no body? Body, bodily-consciousness and self-consciousness," in *Handbook of Phenomenology and Cognitive Science*, eds D. Schmicking and S. Gallagher (Dordrecht: Springer), 180–200.
- Lenggenhager, B., Mouthon, M., and Blanke, O. (2009). Spatial aspects of bodily self-consciousness. *Conscious. Cogn.* 18, 110–117. doi: 10.1016/j.concog.2008.11.003
- Lenggenhager, B., Tadi, T., Metzinger, T., and Blanke, O. (2007). Video ergo sum: manipulating bodily self-consciousness. *Science* 317, 1096–1099. doi: 10.1126/science.1143439
- Mizumoto, M., and Ishikawa, M. (2005). Immunity to error through misidentification and the bodily illusion experiment. *J. Conscious. Stud.* 12, 3–19.
- Moro, V., Zampini, M., and Aglioti, S. M. (2004). Changes in spatial position of hands modify tactile extinction but not disownership of contralesional hand in two right brain-damaged patients. *Neurocase* 10, 437–443. doi: 10.1080/13554790490894020
- Olivé, I., and Berthoz, A. (2012). Combined induction of rubber-hand illusion and out-of-body experiences. *Front. Psychol.* 3:128. doi: 10.3389/fpsyg.2012.00128
- Petkova, V. I., Björnsdóttir, M., Gentile, G., Jonsson, T., Li, T. Q., and Ehrsson, H. H. (2011b). 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., 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., Khoshnevis, M., and Ehrsson, H. H. (2011a). The perspective matters! Multisensory integration in ego-centric reference frames determines full-body ownership. *Front. Psychol.* 2:35. doi: 10.3389/fpsyg.2011.00035
- Pfeiffer, C., Lopez, C., Schmutz, V., Duenas, J. A., Martuzzi, R., and Blanke, O. (2013). Multisensory origin of the subjective first-person perspective: visual, tactile, and vestibular mechanisms. *PLoS ONE* 8:e61751. doi: 10.1371/journal.pone.0061751
- Prosser, S., and Récanati, F. (eds.). (2012). *Immunity to Error Through Misidentification: New Essays*. Cambridge: Cambridge University Press.
- Rohde, M., Di Luca, M., and Ernst, M. O. (2011). The rubber hand illusion: feeling of ownership and proprioceptive drift do not go hand in hand. *PLoS ONE* 6:e21659. doi: 10.1371/journal.pone.0021659
- Salomon, R., Lim, M., Pfeiffer, C., Gassert, R., and Blanke, O. (2013). Full body illusion is associated with widespread skin temperature reduction. *Front. Behav. Neurosci.* 7:65. doi: 10.3389/fnbeh.2013.00065
- Serino, A., Alsmith, A., Costantini, M., Mandrigin, A., Tajadura-Jimenez, A., and Lopez, C. (2013). Bodily ownership and self-location: components of bodily self-consciousness. *Conscious. Cogn.* 22, 1239–1252. doi: 10.1016/j.concog.2013.08.013
- Shoemaker, S. (1968). Self-reference and self-awareness. *J. Philos.* 65, 555–567. doi: 10.2307/2024121
- 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. (2011). "The sense of body ownership," in *The Oxford Handbook of Self*, ed S. Gallagher (New York, NY: Oxford University Press), 180–203.
- 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., Prabhu, G., and Haggard, P. (2006). Having a body versus moving your body: how agency structures body-ownership. *Conscious. Cogn.* 15, 423–432. doi: 10.1016/j.concog.2005.09.004
- van der Hoort, B., Guterstam, A., and Ehrsson, H. H. (2011). Being Barbie: the size of one's own body determines the perceived size of the world. *PLoS ONE* 6:e20195. doi: 10.1371/journal.pone.0020195
- van Doorn, G., Hohwy, J., and Symmons, M. (2014). Can you tickle yourself if you swap bodies with someone else? *Conscious. Cogn.* 23, 1–11. doi: 10.1016/j.concog.2013.10.009

Wittgenstein, L. (1958). *The Blue and Brown Books*. New York, NY: Harper and Row Publishers.

Zahavi, D. (2005). *Subjectivity and Selfhood: Investigating the First-Person Perspective*. Cambridge, MA: MIT press.

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 September 2014; accepted: 27 December 2014; published online: 20 January 2015.

Citation: Liang C, Chang S-Y, Chen W-Y, Huang H-C and Lee Y-T (2015) Body ownership and experiential ownership in the self-touching illusion. *Front. Psychol.* 5:1591. doi: 10.3389/fpsyg.2014.01591

This article was submitted to *Consciousness Research*, a section of the journal *Frontiers in Psychology*.

Copyright © 2015 Liang, Chang, Chen, Huang and Lee. 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 Merit of Synesthesia for Consciousness Research

Tessa M. van Leeuwen^{1,2,3*}, Wolf Singer^{1,2,4} and Danko Nikolić^{1,2,4,5*}

¹ Department of Neurophysiology, Max Planck Institute for Brain Research, Frankfurt am Main, Germany, ² Ernst Strüngmann Institute for Neuroscience in Cooperation with Max Planck Society, Frankfurt am Main, Germany, ³ Centre for Cognition, Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, Netherlands, ⁴ Frankfurt Institute for Advanced Studies, Johann Wolfgang Goethe University, Frankfurt am Main, Germany, ⁵ Department of Psychology, University of Zagreb, Zagreb, Croatia

OPEN ACCESS

Edited by:

Aleksandra Mroczko-Wasowicz,
National Yang Ming University, Taiwan

Reviewed by:

Christianne Jacobs,
University of Westminster, UK
Clare Jonas,
University of East London, UK

*Correspondence:

Tessa M. van Leeuwen
tesvlee@gmail.com;
Danko Nikolić
danko.nikolic@googlemail.com

Specialty section:

This article was submitted to
Consciousness Research,
a section of the journal
Frontiers in Psychology

Received: 18 July 2015

Accepted: 15 November 2015

Published: 02 December 2015

Citation:

van Leeuwen TM, Singer W and
Nikolić D (2015) The Merit of
Synesthesia for Consciousness
Research. *Front. Psychol.* 6:1850.
doi: 10.3389/fpsyg.2015.01850

Synesthesia is a phenomenon in which additional perceptual experiences are elicited by sensory stimuli or cognitive concepts. Synesthetes possess a unique type of phenomenal experiences not directly triggered by sensory stimulation. Therefore, for better understanding of consciousness it is relevant to identify the mental and physiological processes that subserve synesthetic experience. In the present work we suggest several reasons why synesthesia has merit for research on consciousness. We first review the research on the dynamic and rapidly growing field of the studies of synesthesia. We particularly draw attention to the role of semantics in synesthesia, which is important for establishing synesthetic associations in the brain. We then propose that the interplay between semantics and sensory input in synesthesia can be helpful for the study of the neural correlates of consciousness, especially when making use of ambiguous stimuli for inducing synesthesia. Finally, synesthesia-related alterations of brain networks and functional connectivity can be of merit for the study of consciousness.

Keywords: synesthesia, consciousness, semantics, parietal cortex, effective connectivity, qualia, NCC, ambiguous stimuli

INTRODUCTION

Synesthesia is a phenomenon in which a presentation of a stimulus, or inducer, produces additional phenomenal experiences, or concurrents, for which no physical sensory inputs exist. For instance, the letter “A” may trigger an experience of red color (Hochel and Milán, 2008) even though there is nothing that can be established as objectively red in the stimulus. Rather, the experience of color red is created exclusively internally. Synesthetic mappings of the inducing stimulus and the concurrent experience are automatic and involuntary, unique for each individual, and generally stable over time (Wollen and Ruggiero, 1983; Baron-Cohen et al., 1987; but see Simner, 2012). The prevalence of synesthesia in the normal population is estimated to be around 1–2% (Simner et al., 2006).

Synesthesia is a unique subject for research particularly because of the additional phenomenal experiences, i.e., the concurrent *qualia*. How and why phenomenal experience arises is part of the “hard” problem in consciousness research (e.g., Chalmers, 1995). Many requirements that are necessary for us to be conscious of the world around us can be explained in terms of functioning, information processing, circuits or systems—for instance, the integration of information by a cognitive system, deliberate control of behavior, and the reportability of mental states (also, among others, named the “easy” problems of consciousness; Chalmers, 1995). This is the relatively easier problem. However, it has turned out much more difficult to explain the mechanisms of

subjective experience or “what things feel like.” Examples of subjective experiences are for instance the *experience of* seeing red, or the *experience of* feeling the wind on your face on a sunny day. This aspect of consciousness is considered harder to account for.

Many empirical studies have been undertaken to identify neural correlates of conscious experience (Dehaene and Naccache, 2001; Singer, 2001; Melloni et al., 2007; Aru et al., 2012). According to Crick (1994) neural correlates of consciousness is “the minimal set of neuronal events leading to subjective awareness” (Cohen and Dennett, 2011, p.358); Opinions differ widely on how and whether conscious experience can be accounted for by the physiological processes in the brain (Chalmers, 1995; Metzinger, 2000; Gray, 2005; Cohen and Dennett, 2011). For instance, there is no agreement on whether conscious experience and cognitive processing are integrated or not (e.g., Cohen and Dennett, 2011); and on whether conscious experience is established by means of global integration of information (e.g., Lamme, 2006) or by activity in e.g., localized brain areas (Zeki, 2001).

Several researchers have proposed ways in which consciousness research can benefit from the phenomenon of synesthesia. Gray (2005) has argued that synesthesia can provide a window on the hard problem of consciousness. Specifically, Gray proposes that synesthesia can make a case against functionalism—a theory stating that the mental states that constitute consciousness can be defined as interactions between different functional processes. Its prediction, according to Gray, is that for any difference in function, there should be a corresponding difference in subjective experience. The argument put forward by Gray (2005) is that in e.g., colored-hearing synesthesia, two different functions (hearing and color vision) lead to the same subjective experience of color—which is allegedly mediated in brain areas related to color vision. In the meantime, several neuroimaging studies have demonstrated that synesthetic color and veridical color perception do not necessarily share the same neuronal correlates i.e., activated brain regions (Van Leeuwen et al., 2010; Hupé et al., 2012). Moreover, synesthetic colors are often qualitatively different from veridical colors in the sense of texture, quality, and specificity of colors (Eagleman and Goodale, 2009).

Sagiv and Frith (2013) proposed that synesthesia can be used as a model problem for understanding conscious experience for several reasons. Synesthesia is phenomenologically defined while its properties can be studied in detail; there are a wide variety of types of synesthesia providing ample possibilities for testing the neural correlates of various kinds of experiences; synesthetes are normally healthy people and are much more easy to find than for instance patients with specific neurological symptoms. Sagiv et al. suggest that synesthesia should be helpful in the search for the neural correlates of consciousness. In consciousness research it is common to use paradigms in which a subjective change in perception takes place while the stimuli remain constant. Hence, synesthesia can be seen as a special case of changes in subjective experiences that can be contrasted to the conscious percepts of other people in response to the same stimuli. Another interesting point made by Sagiv et al. is with regard to the proposal by Zeki

(2001) that consciousness is mediated by “essential (localized) nodes” in the brain that are required for conscious experience. Taking color as an example, synesthesia lends support for that hypothesis because localized differences in activity level of color regions in the brain have been reported for color synesthesia (see e.g., Van Leeuwen et al., 2010; Rouw et al., 2011).

In the present work we suggest several additional reasons why synesthesia has merit for research on consciousness. We first review the research on the dynamic and rapidly growing field of the studies of synesthesia to inform the reader of the current state of affairs. We pay specific attention to the role of semantics in synesthesia, which seems important for establishing synesthetic associations in the brain. We then propose that the interplay between semantics and sensory input in synesthesia can be helpful for the studies of the neural correlates of consciousness, especially when making use of ambiguous stimuli for inducing synesthesia. Finally, alterations of functional connectivity and physical network connectivity discovered in the brains of synesthetes can be useful for the studies of consciousness.

THE ROLE OF SEMANTIC ASSOCIATIONS IN SYNESTHESIA

Synesthesia is a phenomenon that has been known for a long time, with reports dating back as far as 1812 (see Jewanski et al., 2009), but only recently has the phenomenon received due attention from researchers, leading to an explosion in research efforts, largely kicked off by Cytowic's (1993) book “The man who tasted shapes.” The most common tests to determine genuine synesthesia is the test-retest task to assess the consistency of the synesthetic experiences (e.g., Baron-Cohen et al., 1987). In addition, variants of Stroop interference tasks (Stroop, 1935) are often used to provide evidence for the automatic nature of synesthetic associations (Wollen and Ruggiero, 1983; Nikolić et al., 2011). It should be noted that for certain types of synesthesia, the automaticity has also been questioned (Mattingley, 2009; Price and Mattingley, 2013). There are even online resources that enable one to determine whether he or she is a synesthete (e.g., www.synesthete.org; Eagleman et al., 2007). Many different forms of synesthesia exist. Inducers may be letters, words, numbers, time-units (days of week, months), personal names, music, smell, taste, etc. (Day, 2014). However, there are differences in the prevalence of different forms of synesthesia. Just graphemes and time-units may account for 70–80% of all incidences of synesthesia (Day, 2014). There are some very rare forms of synesthesia such as e.g., swimming-style to color synesthesia (Nikolić et al., 2011).

The Nature of Inducers

There has been quite a controversy on what the nature of synesthesia really is and how the associations are being created. Before discussing this issue further, it is important to note that little, if anything, is known about how phenomenal experiences may come about in physiological systems. It was therefore difficult for theories of synesthesia to be created on the basis of strong empirical or theoretical foundations. The most straightforward and simple hypothesis was that the

additional experiences have to do with additional activation of neurons—presuming rather bluntly, a direct correspondence between elevated firing rates of neurons and phenomenal experience. For example, a neuron coding for red color would be activated and this would then lead, in some unspecified way, to the experience of red. The activation would then come through connections originating from a different brain area—e.g., from the grapheme area to the color area (Ramachandran and Hubbard, 2001). These theories are inspired by the connectionist approach to the brain (McClelland and Rumelhart, 1985), and can be referred to as connection-activation theories (Grossenbacher and Lovelace, 2001; Ramachandran and Hubbard, 2001).

In early days of synesthesia research the theories of aberrant sensory-to-sensory connections were the only ones discussed and studies concentrated on the mechanisms of the hypothetical activation, direct excitation or disinhibition (Grossenbacher and Lovelace, 2001; Ramachandran and Hubbard, 2001). However, later investigations of synesthetic phenomena suggested that the sensory-sensory view of synesthesia should be expanded to allow for *concepts* that can induce synesthesia. It has often been shown that it is not necessarily the sensory inputs that evoke synesthetic concurrents, but rather the extracted meaning of the stimulus. For grapheme-color synesthesia—the most common form and the most frequently studied—this semantic component has been shown reliably and from many different angles (see Chiou and Rich, 2014 for a recent review). For example, one and the same physical stimulus would evoke a different concurrent depending on how the stimulus was interpreted (e.g., the same shape can be understood as an S or as number 5; e.g., Myles et al., 2003; Dixon et al., 2006). Also, it has been shown that new synesthetic associations can be created immediately (within minutes) as new meanings are given to symbols (Mroczko et al., 2009). Indeed, it became soon clear for the most common, and thus most thoroughly studied, forms of synesthesia that they are conceptual in nature (Simner et al., 2006; Novich et al., 2011). Most obvious examples of conceptual synesthesias are days of the week that are colored by their meaning or position in the sequence (Sagiv et al., 2006) as opposed to (or simultaneous with) being colored by the colors of the letter with which the name of the weekday begins; and synesthesias for abstract representations of numerosity such as dice patterns and finger counting—despite the different surface forms, the same number elicits the same color in all cases (Ward and Sagiv, 2007; Ward et al., 2007). Dixon et al. (2000) also demonstrated that synesthesia can occur when synesthetes are merely *thinking* of the inducing stimulus.

These findings of the relevance of semantics were paralleled with failures to replicate the so-called perceptual “pop-out” that certain synesthetes reportedly experience for stimuli that normally do not induce pop-out. Initial case studies had indicated synesthetes experience pop-out (an immediate percept) during visual search (Ramachandran and Hubbard, 2001; Palmeri et al., 2002), but more elaborate (large sample size) studies revealed that this was actually not the case, at least for the majority of synesthetes (Edquist et al., 2006; Sagiv et al., 2006; Gheri et al., 2008; Laeng, 2009; Rothen and Meier, 2009). Synesthetes may show a slight benefit at visual search

compared to non-synesthetes (e.g., Palmeri et al., 2002), but for a large majority of synesthetes, synesthesia does not occur pre-attentively (Ward et al., 2010).

These developments had implications for understanding synesthesia. The notion of direct sensory-to-sensory connections as the only mechanism explaining synesthesia had to be revised, requiring researchers to incorporate semantics as a mediator in the process. This has led to the introduction of the term *ideasthesia*, meaning “sensing concepts” (*idea* is Ancient Greek for concept), as a description of the phenomenon (Nikolić, 2009). In essence, ideasthesia is an equivalent to the recently studied phenomena of semantically-mediated crossmodal correspondences (Rubinsten and Henik, 2002; Gallace and Spence, 2006), but, when applied to synesthesia, it generates a specific set of predictions and constraints. Ideasthesia suggests that synesthetes are not born with their associations, as has been suggested earlier, but that the associations have been created by an active process of assigning meaning to a stimulus (to the inducer; Mroczko-Wasowicz and Nikolić, 2014). This happens especially in situations in which synesthetes have difficulties in assigning meaning to stimuli during learning—i.e., a so-called “semantic vacuum.” The theory proposes that synesthetic associations are created to enhance the understanding of the world—to build knowledge. The theory has a strong explanatory power in accounting for the fact that letters, numbers, days of week, and months are the most common inducers; these stimuli are the first abstract concepts that a child is faced with through the educational system. A child has to build a whole new semantic network and a synesthetic child uses synesthesia to enhance this process.

Ideasthesia is consistent with the finding that synesthetes seem to “choose” the concurrent from various sources, before they internalize one of the options and stick with it for a lifetime (e.g., Simner and Bain, 2013). The “choices” for synesthetic concurrents can come from internal and external sources. Many of the crossmodal associations that exist in synesthesia follow patterns of crossmodal associations in non-synesthetes: for instance, the association of high tones with lighter colors and low tones with darker colors (Simner et al., 2005; Ward et al., 2006). Other associations are suggested directly from the external environment, such as the refrigerator-magnet letters (Witthoft and Winawer, 2006, 2013), and others are combinations such as the similarity between shapes of letters that lead to similarities in associated colors (e.g., Brang et al., 2011; Jürgens and Nikolić, 2012), or the sounds of letters (e.g., Mills et al., 2002).

Ideasthesia is not necessarily incompatible with the idea of direct sensory-to-sensory connections in some forms of synesthesias. Relatively noisy and less restricted cortical activation to new (and abstract) stimuli encountered during learning may theoretically lead to random sensory-sensory co-activation in the brain. Synesthetes may be somehow particularly vulnerable or sensitive to such cross-activations (Bargary and Mitchell, 2008; Newell and Mitchell, 2015). During learning, however, semantic processes seem to shape the representation of the (abstract) stimuli and the associated synesthetic experience either becomes incorporated in the higher-level representation of the stimulus, or not (Van Leeuwen, 2014). Newell and Mitchell

(2015) propose that although synesthesia may be predisposed in certain individuals, this does not exclude a strong influence of experience and semantics on the final phenotype of synesthesia. Ideasthesia is compatible with this view, but does pose that many forms of synesthesia are strongly dependent on semantics. Note that ideasthesia is not the same as synesthesia without physical input. It is rather the way the physical input is translated into semantics that sets ideasthesia apart.

Implications for Understanding Conscious Experience

We propose that most of the synesthetic experiences are mediated through semantics. This suggests that also other experiences should be modified by semantics. In fact, Milán et al. (2013) found that cross-modal associations of experiences are not sparse or isolated. Instead, those associations are tightly interconnected into an associative network that closely resembles the semantic network of words. A common example of cross-modal association is a relation of color to temperature (i.e., red is hot, blue is cold) or auditory pitch to visual size (low pitch for large objects and high pitch for small objects; see Spence, 2011 for a review of various crossmodal correspondences). For their study Milán et al. (2013) used Kiki-Bouba shapes (Köhler, 1947) as they found that these shapes have rich personality properties commonly shared among individuals. They conclude that everyday experiences undergo also a form of ideasthesia. This work is closely related to studies of connative meaning (e.g., Walker and Walker, 2012; Walker et al., 2012) in which it is predicted that the same set of cross-sensory correspondences should always emerge regardless of the specific sensory channel (e.g., brightness and high pitch, high pitch and sharpness, etc.) because there is overlap in the semantic interpretation of the stimuli from different modalities. In the framework proposed by Walker et al. (2012) sensory features become linked together conceptually; in Milán et al. (2013) it is also concepts (e.g., personality) that become linked in an associative network.

Research on synesthesia provides a strong case for the role of semantics in phenomenal experience, and this may be helpful in the search for the neural correlates of consciousness. In synesthetes, dependent on context, an ambiguous synesthesia-inducing stimulus (e.g., 5/S) may be accompanied by a different percept. Similar ambiguous stimuli are also available for non-synesthetes such as bi-stable percepts of the Rubin vase/faces or Necker cube; the difference being that in synesthesia, the additional synesthetic percept is *unrelated* to the actual physical sensory input. This suggests possibly a larger separation in the neural representation of the ambiguous stimulus in synesthetes, and thus an easier detection of differential activation with neuroimaging methods in this population. It should be noted, however, that in studying the neural correlates of synesthesia inducing stimuli, both the inducing stimulus and the concurrent synesthetic experience will result in changes in brain activity when the interpretation of the ambiguous stimuli changes. Cross-modality synesthesias might therefore be most suitable for consciousness paradigms, e.g., ambiguous auditory phonemes/words eliciting a color ([ph]/[f], know/no). Moreover, in synesthesia, it has been well established that the

change in percept of the ambiguous stimulus is mediated in a top-down fashion—from semantics to the percept. In the case of non-synesthetic ambiguous stimuli, it may be more difficult to disentangle changes in the percept mediated top-down from those occurring through slight differences in bottom-up processing. We therefore propose that synesthesia and especially the role of top-down semantic influences therein can be helpful in determining neural correlates of consciousness, not only with respect to localization but also with respect to the involved networks and possible dynamic interactions between higher and lower level brain areas.

THE PHYSIOLOGY OF SYNESTHESIA

Profiting from non-invasive techniques for recording brain activity and from established research paradigms of cognitive neuroscience, search for the physiological underpinnings of synesthesia has become possible. One of the goals is to discover a physiological difference between the brains of synesthetes and non-synesthetes. Another, arguably even more important, goal is to identify physiological parameters that distinguish between an activated phenomenal experience and a lack of such experience (“neural correlates of consciousness”). In the latter case, the unique properties of synesthetic minds may allow for the needed experimental controls—the presence of an additional experience. If this quest were successful, it may bring us closer to understanding the physiology of phenomenal experience. It should be noted that in synesthesia research, between-group comparisons always bear the problem of individual differences between subjects and the possibility that the synesthete group also differs from non-synesthetes on other aspects (e.g., artistic qualities, mental imagery, or personality, see e.g., Ward et al., 2008; Rouw et al., 2011; Banissy et al., 2013).

Brain Function and Structure in Synesthetes

Contrasting synesthetic experience with the absence of such an experience, fMRI studies point to excess activity in brain regions that are involved in the processing of the concurrent synesthetic experience. Examples are activity in cortical area V4/V8 when experiencing synesthetic color (e.g., Sperling et al., 2006; Van Leeuwen et al., 2010; for a review, see Rouw et al., 2011), activity in intraparietal sulci for number-form synesthesia (Tang et al., 2008), or excess activity in piriform cortex for olfactory synesthetic concurrents (Chan et al., 2014). The literature is not very consistent, however: For example, not all studies of grapheme-color synesthesia report activity in color area V4 for synesthesia (Rouw et al., 2011). One of the most common findings is excess activity in parietal regions for synesthetes, independent of the specific synesthetic subtype (e.g., Van Leeuwen et al., 2010; Rouw et al., 2011; Neufeld et al., 2012a). Thus, the results may imply a particularly important role of parietal cortex in mediating phenomenal synesthetic experiences.

Electrophysiological studies have demonstrated abnormal processing of synesthesia-inducing stimuli in early processing phases (Beeli et al., 2008; Goller et al., 2009; Brang et al., 2010), over-responsiveness to non-synesthetic parvocellular

visual stimuli (Barnett et al., 2008), and abnormal processing of synesthetic concurrent experiences (Van Leeuwen et al., 2013). Several electrophysiological studies suggest that synesthetic effects can occur very early during the processing of the inducing stimuli (105–115 ms, Brang et al., 2010; 100 ms, Goller et al., 2009; 122 ms, Beeli et al., 2008). For synesthesias clearly shown to be based on concepts, such as grapheme-color synesthesia, this would imply that semantics already plays a role early during stimulus processing.

Studies of brain structure have demonstrated both increased white matter and gray matter density (Rouw and Scholte, 2007, 2010; Hänggi et al., 2008; Jäncke et al., 2009; Weiss and Fink, 2009; Banissy et al., 2012; O'Hanlon et al., 2013; Zamm et al., 2013). Regions in which such differences were found are often related to the specific (functional) network of areas involved in the type of synesthesia at hand. For instance, Rouw and Scholte (2007) found increased white matter connectivity in temporal regions close to the grapheme areas and in parietal regions for grapheme-color synesthetes, while Hänggi et al. (2008) found anatomical differences in the auditory and gustatory areas of a tone-interval–taste synesthete. It should be noted that there are also reports of globally altered brain topology in synesthetes (Hänggi et al., 2011). However, it is important to keep in mind that the reports of structural differences in the brains of synesthetes are problematic in the sense that it is not possible to determine—in adult synesthetes—whether the synesthesia is a result of altered anatomy, or whether altered anatomy results from years of synesthetic experience.

Dynamic Connectivity Patterns in Synesthetes

Altogether, functional and structural neuroimaging studies demonstrate synesthesia-specific alterations in the brain, related to inducing stimuli as well as to concurrent synesthetic experiences. The common theme, however, that emerges from the neuroimaging literature is that communication and connectivity between brain regions appear to be altered in synesthesia. Let us take a look at functional connectivity changes in synesthesia. Changes in functional connectivity patterns in synesthetes have been reported with or without the presence of external stimulation. Tomson et al. (2013) studied networks in grapheme-color synesthetes' brains during rest, auditory grapheme stimulation, and audiovisual grapheme stimulation. Synesthetes had more significant connections during rest and auditory conditions. Investigating the connectivity between 90 anatomical regions, Tomson et al. found that synesthetes showed increased network clustering in visual regions, in line with the type of synesthesia they exhibited. It should be noted that differences in connectivity patterns were also found during rest; this was reported in another resting-state fMRI study as well (Dovern et al., 2012). Dovern et al. found that during rest, connectivity between visual and parietal networks was enhanced for grapheme-color synesthetes, and that the increase in intrinsic network connectivity correlated positively with the strength (consistency) of synesthetic experiences. This work strongly suggests that altered network function is linked to

altered conscious phenomenal experiences, even in absence of direct stimulation.

Two more studies have investigated functional connectivity during task performance. In auditory-visual synesthetes, using sounds as stimuli, Neufeld et al. (2012b) found greater connectivity between parietal regions and primary auditory and visual cortices for synesthetes compared to controls. There was no evidence of greater direct connectivity between auditory and visual regions, suggesting a strong role for intermediate areas such as parietal cortex during synesthetic experience. In Sinke et al. (2012), grapheme-color synesthetes showed greater connectivity between parietal regions and early visual cortex (but not the grapheme area). Together these two studies are in line with the resting state results reported above; functional connectivity differences appear to be related to the brain areas that are involved in the specific subtype of synesthesia, with a prominent role for parietal regions. It is important to keep in mind that alterations of functional connectivity can result from either changes in direct physical connectivity or from the alterations in the contribution of the semantic associations.

It is relevant for the study of consciousness—and for studies of the physiology of synesthesia—to consider the strong individual differences commonly detected among synesthetes. Not only do different individuals experience different forms of synesthesia, but also the nature of the concurrent experience can differ (for a review, see van Leeuwen, 2013). One of the most common distinctions in subjective experiences of the concurrent is its spatial location. For instance, in grapheme-color synesthesia the concurrent colors can be experienced either “in the mind's eye,” as if resembling an association, or “projected” into space e.g., located on the same surface as the inducing grapheme (Dixon et al., 2004; Ward et al., 2007). These differences in phenomenal experiences are highly informative for physiological studies on synesthesia: If the phenomenal experiences differ, we can also attempt to find the correlates of these differences in the brain activity or anatomy. To this end, Van Leeuwen et al. (2011) performed an effective-connectivity study, using dynamic causal modeling of fMRI signals. Van Leeuwen et al. demonstrated that the phenomenal experience of the synesthetic colors depends on the direction of information flow between brain areas involved in the phenomenon. For projectors, the data were fit best by a model in which synesthesia modulated the direct influence of the grapheme area onto color area V4. For associators, however, a different model fit the data best—one in which V4 activity was influenced indirectly via higher-order regions. This study demonstrated that the quality of phenomenal experience can depend on the route that information flow takes in the brain, even though the same brain areas are implicated in two different processes.

Two related studies have investigated anatomical changes in the brain in relation to the projector vs. associator status of the subjects. Rouw and Scholte (2007) showed that white matter structural connectivity is generally enhanced in synesthetes. More importantly, they found that white matter changes in temporal regions—those that lay near the so-called grapheme area—were more prominent for projector than for associator synesthetes. In a later study on gray matter structure and

function (Rouw and Scholte, 2010), the same authors reported that projector synesthetes had increased gray matter density compared to control subjects in areas generally responsible for perception and action (i.e., visual, auditory, and motor cortex). On the contrary, associators differed from controls in the gray matter density of hippocampus and parahippocampal gyrus, which are regions primarily involved in memory. In conclusion, even relatively subtle individual differences in phenomenal synesthetic experiences can be correlated with identifiable differences in activity and anatomy of brain regions.

THEORIES OF SYNESTHESIA

In the previous section we have discussed that concepts can induce synesthesia—i.e., synesthetic inducers can be of very abstract nature. The question is then, how do the purported physiological mechanisms of synesthesia account for the fact that synesthetic concurrents are shaped by semantic knowledge? Generally speaking, the way we understand the world exerts an influence on how we perceive it (Majid et al., 2004). Apparently, this applies also to synesthetic concurrent experiences. The mechanisms by which synesthesia is mediated by the brain are still being debated and there are roughly two groups of theories. The traditional approaches favor the activation-through-connectivity mechanisms. These presume that connections activate neurons, which then produce the concurrent experiences. The two most important theories in this group are the disinhibition or re-entrant theory (Grossenbacher and Lovelace, 2001; Smilek et al., 2001) and the cross-activation account (Ramachandran and Hubbard, 2001; Brang et al., 2010; Hubbard et al., 2011). Recently, another approach has been proposed based on a new theory of how physiological mechanisms implement semantics (Mroczko-Wasowicz and Nikolić, 2014; Nikolić, 2015). We next discuss each approach in more detail.

According to the disinhibited feedback theory, synesthesia is caused by feedback signals sent from higher-order associative regions to primary sensory regions not originally activated by the inducing stimulus (Grossenbacher and Lovelace, 2001). An example would be the activation of color area V4 via feedback from associative parietal cortex after stimulation of the grapheme area (but not color areas) by a black grapheme. This account of synesthesia allows for context-based and top-down modulation that affects synesthetic experiences via higher-order associative brain regions. The hypothesis implies that inducing stimuli are processed deeply before the conscious synesthetic experience is elicited. This is relevant for our question about the role of semantics in synesthesia: we already know that context can strongly influence synesthetic experiences (e.g., Myles et al., 2003; Dixon et al., 2006). This theory also presumes a significant role for parietal cortex in synesthesia.

The cross-activation theory (Ramachandran and Hubbard, 2001) differs from the disinhibited feedback theory in proposing that activity in the brain regions that are processing the inducing stimulus, directly results in additional activity in brain regions responsible for mediating the concurrent experience. No intermediate, higher-order processing step is included and

instead aberrant anatomical connections between the regions processing the inducing and concurrent stimuli are proposed. The lack of an intermediate processing step implies that parietal cortex is not crucial for synesthetic experience. However, in a later update of the cross-activation model, a second stage of (hyper-)binding through parietal mechanisms was added to the theory (Hubbard et al., 2011). In this way the authors acknowledged the growing evidence for an important role of parietal cortex in synesthesia (see below). The cross-activation theory accounts well for fast changes in electrophysiological signals and is consistent with the apparent evidence that synesthetes experience bottom-up perceptual pop-out in serial search task, and with anatomical differences in the brains of synesthetes. However, evidence for pop-out in synesthetes has become challenged over time, and more and more data suggested a crucial role of semantics in shaping synesthetic experiences.

The recent alternative theory of synesthesia proposed by Mroczko-Wasowicz and Nikolić (2014) is based on a novel proposal of how the brain deals with semantics, founded in the theory of *practopoiesis* (Nikolić, 2015). This view implies that the brain can quickly change its computational properties—i.e., it can make quick learning-like changes—and that the extracted meaning of a stimulus reflects those fast changes made to how the network executes its computations. Examples of quick adaptation of computational properties by the brain's network are for instance studies showing that context can affect synesthetic experiences (e.g., an S/5-shaped stimulus presented either in the context of digits or letters; Dixon et al., 2006).

The directionally changed patterns of functional connections in synesthetes found by Van Leeuwen et al. (2011) are partly consistent with the disinhibited feedback account of synesthesia—namely, the model including an indirect pathway to V4 via parietal regions that fit best for the associators—and are partly consistent with the cross-activation theory—namely the model with direct influences between the grapheme and color area that fit best for projector synesthetes. It is clear, however, that the process leading to synesthesia and its related changes in effective connectivity involves a phase of learning abstract entities for which semantics are important. During the phase in which synesthesia develops, the network that processes the inducing stimuli undergoes changes in its computational properties to accommodate the phenomenon of synesthesia. The conscious experience of the synesthetic concurrent is what results; depending on the properties of the established network. Especially relevant here is that conscious experience can depend on the direction of information flow in the network (Van Leeuwen et al., 2011).

DISCUSSION

The question of why physiological systems have qualia is arguably the most difficult problem faced by neuroscience (Chalmers, 1995; Harnad, 1995). Research on synesthesia, alone, cannot solve the big puzzle of qualia. However, due to the very nature of the phenomenon of synesthesia—i.e., the additional qualia that synesthetes experience—research efforts directed toward understanding that phenomenon may assist in identifying the

neural correlates of conscious experience. Thus, we can ask whether the progress made in the last decade or two in the research of synesthesia can be summarized in a way that is informative for consciousness research.

Synesthesia is illustrative for the importance of the extraction of meaning from stimuli for inducing phenomenal experiences. The interplay between the physical synesthesia-inducing stimulus and the way semantic associations finally shape the phenotype of synesthesia helps us to realize that semantics shape our experiences. Experiences do not exist isolated from person's understanding of the world. Semantics and understanding also play a role in the proposal that much of the problems of consciousness are determined by a social reality (Singer, 2015). Importantly, the semantic aspect of synesthesia can also be put to use in the search for the neural correlates of consciousness. Ambiguous stimuli can elicit different synesthetic concurrents of which the neural correlates might be identified. Additionally, because the ambiguity is always resolved by top-down influences, we can investigate the directionality in mediating phenomenal experiences.

On the other hand, physiological investigations of synesthesia have discovered the important role of the parietal cortex (Van Leeuwen et al., 2010; Rouw et al., 2011; Neufeld et al., 2012a) and of relating individual differences in synesthetic experiences to the directions of effective connectivity (Van Leeuwen et al., 2011).

REFERENCES

- Aru, J., Bachmann, T., Singer, W., and Melloni, L. (2012). Distilling the neural correlates of consciousness. *Neurosci. Biobehav. Rev.* 36, 737–746. doi: 10.1016/j.neubiorev.2011.12.003
- Banissy, M. J., Holle, H., Cassell, J., Annett, L., Tsakanikos, E., Walsh, V., et al. (2013). Personality traits in people with synaesthesia: do synaesthetes have an atypical personality profile? *Pers. Individ. Dif.* 54, 828–831. doi: 10.1016/j.paid.2012.12.018
- Banissy, M. J., Stewart, L., Muggleton, N. G., Griffiths, T. D., Walsh, V. Y., Ward, J., et al. (2012). Grapheme-color and tone-color synesthesia is associated with structural brain changes in visual regions implicated in color, form, and motion. *Cogn. Neurosci.* 3, 29–35. doi: 10.1080/17588928.2011.594499
- Bargary, G., and Mitchell, K. J. (2008). Synaesthesia and cortical connectivity. *Trends Neurosci.* 31, 335–342. doi: 10.1016/j.tins.2008.03.007
- Barnett, K. J., Foxe, J. J., Molholm, S., Kelly, S. P., Shalgi, S., Mitchell, K. J., et al. (2008). Differences in early sensory-perceptual processing in synesthesia: a visual evoked potential study. *Neuroimage.* 43, 605–613. doi: 10.1016/j.neuroimage.2008.07.028
- Baron-Cohen, S., Wyke, M. A., and Binnie, C. (1987). Hearing words and seeing colours: an experimental investigation of a case of synaesthesia. *Perception* 16, 761–767. doi: 10.1068/p160761
- Beeli, G., Esslen, M., and Jäncke, L. (2008). Time course of neural activity correlated with colored-hearing synesthesia. *Cereb. Cortex* 18, 379–385. doi: 10.1093/cercor/bhm072
- Brang, D., Hubbard, E. M., Coulson, S., Huang, M., and Ramachandran, V. S. (2010). Magnetoencephalography reveals early activation of V4 in grapheme-color synesthesia. *Neuroimage* 53, 268–274. doi: 10.1016/j.neuroimage.2010.06.008
- Brang, D., Rouw, R., Ramachandran, V. S., and Coulson, S. (2011). Similarly shaped letters evoke similar colors in grapheme-color synesthesia. *Neuropsychologia* 49, 1355–1358. doi: 10.1016/j.neuropsychologia.2011.01.002
- Chalmers, D. J. (1995). Facing up to the problem of consciousness. *J. Conscious. Stud.* 2, 200–219.
- Chan, J. S., Van Den Bosch, J. J. F., Theves, S., Hardt, S., Pflanz, P., Lötsch, J., et al. (2014). Synaesthesia or vivid imagery? A single case fMRI study of visually induced olfactory perception. *Multisens. Res.* 27, 225–246. doi: 10.1163/22134808-00002451
- Chiou, R., and Rich, A. N. (2014). The role of conceptual knowledge in understanding synaesthesia: evaluating contemporary findings from a “hub-and-spokes” perspective. *Front. Psychol.* 5:105. doi: 10.3389/fpsyg.2014.00105
- Cohen, M. A., and Dennett, D. C. (2011). Consciousness cannot be separated from function. *Trends Cogn. Sci.* 15, 358–364. doi: 10.1016/j.tics.2011.06.008
- Crick, F. H. C. (1994). *The Astonishing Hypothesis: The Scientific Search for the Soul*. New York, NY: Scribner.
- Cytowic, R. E. (1993). *The Man Who Tasted Shapes*. Cambridge: MIT Press.
- Day, S. A. (2014). *Types of Synesthesia*. Retrieved from: <http://www.daysyn.com/Types-of-Syn.html>
- Dehaene, S., and Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition* 79, 1–37. doi: 10.1016/S0010-0277(00)00123-2
- Dixon, M. J., Smilek, D., Cudahy, C., and Merikle, P. M. (2000). Two plus five equals yellow. *Nature* 406, 365. doi: 10.1038/35019148
- Dixon, M. J., Smilek, D., Duffy, P. L., Zanna, M. P., and Merikle, P. M. (2006). The role of meaning in grapheme-colour synaesthesia. *Cortex* 42, 243–252. doi: 10.1016/S0010-9452(08)70349-6
- Dixon, M. J., Smilek, D., and Merikle, P. M. (2004). Not all synaesthetes are created equal: projector versus associator synaesthetes. *Cogn. Affect. Behav. Neurosci.* 4, 335–343. doi: 10.3758/CABN.4.3.335
- Dovern, A., Fink, G. R., Fromme, A. C. B., Wohlschläger, A. M., Weiss, P. H., and Riedl, V. (2012). Intrinsic network connectivity reflects consistency of synesthetic experiences. *J. Neurosci.* 32, 7614–7621. doi: 10.1523/JNEUROSCI.5401-11.2012
- Eagleman, D. M., and Goodale, M. A. (2009). Why color synesthesia involves more than color. *Trends Cogn. Sci.* 13, 288–292. doi: 10.1016/j.tics.2009.03.009
- Eagleman, D. M., Kagan, A. D., Nelson, S. S., Sagaram, D., and Sarma, A. K. (2007). A standardized test battery for the study of synesthesia. *J. Neurosci. Methods* 159, 139–145. doi: 10.1016/j.jneumeth.2006.07.012
- Edquist, J., Rich, A. N., Brinkman, C., and Mattingley, J. B. (2006). Do synaesthetic colours act as unique features in visual search? *Cortex* 42, 222–231. doi: 10.1016/S0010-9452(08)70347-2

FUNDING

This review was supported by LOEWE-Neuronale Koordination Forschungsschwerpunkt Frankfurt (NeFF).

- Gallace, A., and Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Percept. Psychophys.* 68, 1191–1203. doi: 10.3758/BF03193720
- Gheri, C., Chopping, S., and Morgan, M. J. (2008). Synaesthetic colours do not camouflage form in visual search. *Proc. R. Soc. B Biol. Sci.* 275, 841–846. doi: 10.1098/rspb.2007.1457
- Goller, A. I., Otten, L. J., and Ward, J. (2009). Seeing sounds and hearing colors: an event-related potential study of auditory-visual synesthesia. *J. Cogn. Neurosci.* 21, 1869–1881. doi: 10.1162/jocn.2009.21134
- Gray, J. (2005). "Synesthesia: a window on the hard problem of consciousness," in *Synesthesia: Perspectives From Cognitive Neuroscience*, eds L. C. Robertson and N. Sagiv (New York, NY: Oxford University Press), 127–146.
- Grossenbacher, P. G., and Lovelace, C. T. (2001). Mechanisms of synesthesia: cognitive and physiological constraints. *Trends Cogn. Sci.* 5, 36–41. doi: 10.1016/S1364-6613(00)01571-0
- Hänggi, J., Beeli, G., Oechslin, M. S., and Jäncke, L. (2008). The multiple synaesthete ES - Neuroanatomical basis of interval-taste and tone-colour synaesthesia. *Neuroimage* 43, 192–203. doi: 10.1016/j.neuroimage.2008.07.018
- Hänggi, J., Wotruba, D., and Jäncke, L. (2011). Globally altered structural brain network topology in grapheme-color synesthesia. *J. Neurosci.* 31, 5816–5828. doi: 10.1523/JNEUROSCI.0964-10.2011
- Harnad, S. (1995). Why and how we are not zombies. *Philos. Explor.* 1, 164–167.
- Hochel, M., and Milán, E. G. (2008). Synaesthesia: the existing state of affairs. *Cogn. Neuropsychol.* 25, 93–117. doi: 10.1080/02643290701822815
- Hubbard, E. M., Brang, D., and Ramachandran, V. S. (2011). The cross-activation theory at 10. *J. Neuropsych.* 5, 152–177. doi: 10.1111/j.1748-6653.2011.02014.x
- Hupé, J.-M., Bordier, C., and Dojat, M. (2012). The neural bases of grapheme-color synesthesia are not localized in real color-sensitive areas. *Cereb. Cortex* 22, 1622–1633. doi: 10.1093/cercor/bhr236
- Jäncke, L., Beeli, G., Eulig, C., and Hanggi, J. (2009). The neuroanatomy of grapheme-color synesthesia. *Eur. J. Neurosci.* 29, 1287–1293. doi: 10.1111/j.1460-9568.2009.06673.x
- Jewanski, J., Day, S. A., and Ward, J. (2009). A colorful albino: the first documented case of synaesthesia, by Georg Tobias Ludwig Sachs in 1812. *J. Hist. Neurosci.* 18, 293–303. doi: 10.1080/09647040802431946
- Jürgens, U. M., and Nikolić, D. (2012). Ideesthesia: conceptual processes assign similar colours to similar shapes. *Translat. Neurosci.* 3, 22–27. doi: 10.2478/s13380-012-0010-4
- Köhler, W. (1947). *Gestalt Psychology: An Introduction to New Concepts in Modern Psychology*. New York, NY: Liveright.
- Laeng, B. (2009). Searching through synaesthetic colors. *Atten. Percept. Psychophys.* 71, 1461–1467. doi: 10.3758/APP.71.7.1461
- Lamme, V. A. (2006). Towards a true neural stance on consciousness. *Trends Cogn. Sci.* 10, 494–501. doi: 10.1016/j.tics.2006.09.001
- Majid, A., Bowerman, M., Kita, S., Haun, D. B. M., and Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends Cogn. Sci.* 8, 108–114. doi: 10.1016/j.tics.2004.01.003
- Mattingley, J. B. (2009). Attention, automaticity, and awareness in synesthesia. *Ann. N.Y. Acad. Sci.* 1156, 141–167. doi: 10.1111/j.1749-6632.2009.04422.x
- McClelland, J. L., and Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *J. Exp. Psychol. Gen.* 114, 159–188. doi: 10.1037/0096-3445.114.2.159
- Melloni, L., Molina, C., Pena, M., Torres, D., Singer, W., and Rodriguez, E. (2007). Synchronization of neural activity across cortical areas correlates with conscious perception. *J. Neurosci.* 27, 2858–2865. doi: 10.1523/JNEUROSCI.4623-06.2007
- Metzinger, T. (2000). *Neural Correlates of Consciousness: Empirical and Conceptual Questions*. Cambridge, MA: MIT Press.
- Milán, E., Iborra, O., De Cordoba, M. J., Juarez-Ramos, V., Artacho, M. A. R., and Rubio, J. L. (2013). The Kiki-Bouba effect A case of personification and ideesthesia. *J. Conscious. Stud.* 20, 84–102. Available online at: <http://www.ingentaconnect.com/content/imp/jcs/2013/00000020/F0020001/art00005>
- Mills, C. B., Viguers, M. L., Edelson, S. K., Thomas, A. T., Simon-Dack, S. L., and Innis, J. A. (2002). The color of two alphabets for a multilingual synesthete. *Perception* 31, 1371–1394. doi: 10.1068/p3429
- Mroczko, A., Metzinger, T., Singer, W., and Nikolic, D. (2009). Immediate transfer of synesthesia to a novel inducer. *J. Vis.* 9, 8. doi: 10.1167/9.12.25
- Mroczko-Wasowicz, A., and Nikolić, D. (2014). Semantic mechanisms may be responsible for developing synesthesia. *Front. Hum. Neurosci.* 8:509. doi: 10.3389/fnhum.2014.00509
- Myles, K. M., Dixon, M. J., Smilek, D., and Merikle, P. M. (2003). Seeing double: the role of meaning in alphanumeric colour synaesthesia. *Brain Cogn.* 53, 342–345. doi: 10.1016/S0278-2626(03)00139-8
- Neufeld, J., Sinke, C., Dillo, W., Emrich, H. M., Szycik, G. R., Dima, D., et al. (2012a). The neural correlates of coloured music: a functional MRI investigation of auditory-visual synaesthesia. *Neuropsychologia* 50, 85–89. doi: 10.1016/j.neuropsychologia.2011.11.001
- Neufeld, J., Sinke, C., Zedler, M., Dillo, W., Emrich, H. M., Bleich, S., et al. (2012b). Disinhibited feedback as a cause of synesthesia: evidence from a functional connectivity study on auditory-visual synesthetes. *Neuropsychologia* 50, 1471–1477. doi: 10.1016/j.neuropsychologia.2012.02.032
- Newell, F. N., and Mitchell, K. (2015). Multisensory integration and cross-modal learning in synaesthesia: a unifying model. *Neuropsychologia*. doi: 10.1016/j.neuropsychologia.2015.07.026. [Epub ahead of print].
- Nikolić, D. (2009). "Is synaesthesia actually ideesthesia? An inquiry into the nature of the phenomenon," in *Proceedings of the Third International Congress on Synaesthesia, Science and Art* (Granada), 26–29.
- Nikolić, D. (2015). Practopoiesis: or how life fosters a mind. *J. Theor. Biol.* 373, 40–61. doi: 10.1016/j.jtbi.2015.03.003
- Nikolić, D., Jürgens, U. M., Rothen, N., Meier, B., and Mroczko, A. (2011). Swimming-style synesthesia. *Cortex* 47, 874–879. doi: 10.1016/j.cortex.2011.02.008
- Novich, S., Cheng, S., and Eagleman, D. M. (2011). Is synaesthesia one condition or many? A large-scale analysis reveals subgroups. *J. Neuropsychol.* 5, 353–371. doi: 10.1111/j.1748-6653.2011.02015.x
- O'Hanlon, E., Newell, F. N., and Mitchell, K. J. (2013). Combined structural and functional imaging reveals cortical deactivations in grapheme-color synaesthesia. *Fron. Psychol.* 4:755. doi: 10.3389/fpsyg.2013.00755
- Palmeri, T. J., Blake, R., Marois, R., Flanery, M. A., and Whetsell, W. J. (2002). The perceptual reality of synesthetic colors. *Proc. Natl. Acad. Sci. U.S.A.* 99, 4127–4131. doi: 10.1073/pnas.022049399
- Price, M. C., and Mattingley, J. B. (2013). Automaticity in sequence-space synaesthesia: a critical appraisal of the evidence. *Cortex* 49, 1165–1186. doi: 10.1016/j.cortex.2012.10.013
- Ramachandran, V. S., and Hubbard, E. M. (2001). Psychophysical investigations into the neural basis of synaesthesia. *Proc. R. Soc. B Biol. Sci.* 268, 979–983. doi: 10.1098/rspb.2000.1576
- Rothen, N., and Meier, B. (2009). Do synesthetes have a general advantage in visual search and episodic memory? A case for group studies. *PLoS ONE* 4:e5037. doi: 10.1371/journal.pone.0005037
- Rouw, R., and Scholte, H. S. (2007). Increased structural connectivity in grapheme-color synesthesia. *Nat. Neurosci.* 10, 792–797. doi: 10.1038/nn1906
- Rouw, R., and Scholte, H. S. (2010). Neural basis of individual differences in synesthetic experiences. *J. Neurosci.* 30, 6205–6213. doi: 10.1523/JNEUROSCI.3444-09.2010
- Rouw, R., Scholte, H. S., and Colizoli, O. (2011). Brain areas involved in synaesthesia: a review. *J. Neuropsychol.* 5, 214–242. doi: 10.1111/j.1748-6653.2011.02006.x
- Rubinsten, O., and Henik, A. (2002). Is an ant larger than a lion? *Acta Psychol.* 111, 141–154. doi: 10.1016/S0001-6918(02)00047-1
- Sagiv, N., and Frith, C. D. (2013). "Synesthesia and consciousness," in *Oxford Handbook of Synesthesia*, eds J. Simner and E. M. Hubbard (Oxford, New York, NY: Oxford University Press), 924–940.
- Sagiv, N., Heer, J., and Robertson, L. (2006). Does binding of synesthetic color to the evoking grapheme require attention? *Cortex* 42, 232–242. doi: 10.1016/S0010-9452(08)70348-4
- Simner, J. (2012). Defining synaesthesia. *Br. J. Psychol.* 103, 1–15. doi: 10.1348/000712610X528305
- Simner, J., and Bain, A. E. (2013). A longitudinal study of grapheme-colour synaesthesia in childhood: 6/7 years to 10/11 years. *Front. Hum. Neurosci.* 7:603. doi: 10.3389/fnhum.2013.00603
- Simner, J., Mulvenna, C., Sagiv, N., Tsakanikos, E., Witherby, S. A., Fraser, C., et al. (2006). Synaesthesia: the prevalence of atypical cross-modal experiences. *Perception* 35, 1024–1033. doi: 10.1068/p5469

- Simner, J., Ward, J., Lanz, M., Jansari, A., Noonan, K., Glover, L., et al. (2005). Non-random associations of graphemes to colours in synaesthetic and non-synaesthetic populations. *Cogn. Neuropsychol.* 22, 1069–1085. doi: 10.1080/02643290500200122
- Singer, W. (2001). Consciousness and the binding problem. *Ann. N.Y. Acad. Sci.* 929, 123–146. doi: 10.1111/j.1749-6632.2001.tb05712.x
- Singer, W. (2015). “Large scale temporal coordination of cortical activity as prerequisite for conscious experience,” in *The Blackwell Companion to Consciousness*, eds M. Velmans and S. Schneider (Oxford, UK: Blackwell Publishing). doi: 10.1002/9780470751466.ch48
- Sinke, C., Neufeld, J., Emrich, H. M., Dillo, W., Bleich, S., Zedler, M., et al. (2012). Inside a synesthete's head: a functional connectivity analysis with grapheme-color synesthetes. *Neuropsychologia* 50, 3363–3369. doi: 10.1016/j.neuropsychologia.2012.09.015
- Smilek, D., Dixon, M. J., Cudahy, C., and Merikle, P. M. (2001). Synaesthetic photisms influence visual perception. *J. Cogn. Neurosci.* 13, 930–936. doi: 10.1162/089892901753165845
- Spence, C. (2011). Crossmodal correspondences: a tutorial review. *Atten. Percept. Psychophys.* 73, 971–995. doi: 10.3758/s13414-010-0073-7
- Sperling, J. M., Prvulovic, D., Linden, D. E. J., Singer, W., and Stirn, A. (2006). Neuronal correlates of colour-graphemic synaesthesia: a fMRI study. *Cortex* 42, 295–303. doi: 10.1016/S0010-9452(08)70355-1
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J. Exp. Psychol.* 18, 643–662. doi: 10.1037/h0054651
- Tang, J., Ward, J., and Butterworth, B. (2008). Number forms in the brain. *J. Cogn. Neurosci.* 20, 1547–1556. doi: 10.1162/jocn.2008.20120
- Tomson, S. N., Narayan, M., Allen, G. I., and Eagleman, D. M. (2013). Neural networks of colored sequence synesthesia. *J. Neurosci.* 33, 14098–14106. doi: 10.1523/JNEUROSCI.5131-12.2013
- van Leeuwen, T. M. (2013). “Individual differences in synesthesia,” in *Oxford Handbook of Synesthesia*, eds J. Simner and E. M. Hubbard (Oxford; New York, NY: Oxford University Press), 241–264. doi: 10.1093/oxfordhb/9780199603329.013.0013
- Van Leeuwen, T. M. (2014). Constructing priors in synesthesia. *Cogn. Neurosci.* 5, 124–126. doi: 10.1080/17588928.2014.905520
- Van Leeuwen, T. M., den Ouden, H. E. M., and Hagoort, P. (2011). Effective connectivity determines the nature of subjective experience in grapheme-color synesthesia. *J. Neurosci.* 31, 9879–9884. doi: 10.1523/JNEUROSCI.0569-11.2011
- Van Leeuwen, T. M., Hagoort, P., and Handel, B. F. (2013). Real color captures attention and overrides spatial cues in grapheme-color synesthetes but not in controls. *Neuropsychologia* 51, 1802–1813. doi: 10.1016/j.neuropsychologia.2013.06.024
- Van Leeuwen, T. M., Petersson, K. M., and Hagoort, P. (2010). Synaesthetic colour in the brain: beyond colour areas. A functional magnetic resonance imaging study of synaesthetes and matched controls. *PLoS ONE* 5:e12074. doi: 10.1371/journal.pone.0012074
- Walker, L., Walker, P., and Francis, B. (2012). A common scheme for cross-sensory correspondences across stimulus domains. *Perception* 41, 1186. doi: 10.1068/p7149
- Walker, P., and Walker, L. (2012). Size-brightness correspondence: crosstalk and congruity among dimensions of connotative meaning. *Att. Percept. Psychophys.* 74, 1226–1240. doi: 10.3758/s13414-012-0297-9
- Ward, J., Huckstep, B., and Tsakanikos, E. (2006). Sound-colour synaesthesia: to what extent does it use cross-modal mechanisms common to us all? *Cortex* 42, 264–280. doi: 10.1016/S0010-9452(08)70352-6
- Ward, J., Jonas, C., Dienes, Z., and Seth, A. (2010). Grapheme-colour synaesthesia improves detection of embedded shapes, but without pre-attentive ‘pop-out’ of synaesthetic colour. *Proc. R. Soc. B Biol. Sci.* 277, 1021–1026. doi: 10.1098/rspb.2009.1765
- Ward, J., Li, R., Salih, S., and Sagiv, N. (2007). Varieties of grapheme-colour synaesthesia: a new theory of phenomenological and behavioural differences. *Conscious. Cogn.* 16, 913–931. doi: 10.1016/j.concog.2006.09.012
- Ward, J., and Sagiv, N. (2007). Synaesthesia for finger counting and dice patterns: a case of higher synaesthesia? *Neurocase* 13, 86–93. doi: 10.1080/13554790701300518
- Ward, J., Thompson-Lake, D., Ely, R., and Kaminski, F. (2008). Synaesthesia, creativity and art: what is the link? *Br. J. Psychol.* 99, 127–141. doi: 10.1348/000712607x204164
- Weiss, P. H., and Fink, G. R. (2009). Grapheme-colour synaesthetes show increased grey matter volumes of parietal and fusiform cortex. *Brain* 132, 65–70. doi: 10.1093/brain/awn304
- Witthoft, N., and Winawer, J. (2006). Synesthetic colors determined by having colored refrigerator magnets in childhood. *Cortex* 42, 175–183. doi: 10.1016/S0010-9452(08)70342-3
- Witthoft, N., and Winawer, J. (2013). Learning, memory, and synesthesia. *Psychol. Sci.* 24, 258–265. doi: 10.1177/0956797612452573
- Wollen, K. A., and Ruggiero, F. T. (1983). Colored-letter synesthesia. *J. Ment. Imagery* 7, 83–86.
- Zamm, A., Schlaug, G., Eagleman, D. M., and Loui, P. (2013). Pathways to seeing music: enhanced structural connectivity in colored-music synesthesia. *Neuroimage* 74, 359–366. doi: 10.1016/j.neuroimage.2013.02.024
- Zeki, S. (2001). Localization and globalization in conscious vision. *Annu. Rev. Neurosci.* 24, 57–86. doi: 10.1146/annurev.neuro.24.1.57

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.

Copyright © 2015 van Leeuwen, Singer and Nikolić. 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.



Synesthesia and release phenomena in sensory and motor grounding. Cases of disinhibited embodiment?

Brian F. Gray¹ and Julia Simner^{1,2*}

¹ Department of Psychology, University of Edinburgh, Edinburgh, UK

² School of Psychology, University of Sussex, Brighton, UK

*Correspondence: j.simner@sussex.ac.uk

Edited by:

Aleksandra Mroczko-Wasowicz, National Yang Ming University, Taiwan

Reviewed by:

Berit Brogaard, University of Missouri, USA

Sean Allen-Hermanson, Florida International University, USA

Keywords: synesthesia, embodiment, embodied cognition, grounding, perception-cognition interface, motor release, alien hand sign, anarchic hand syndrome

Synesthesia is an unusual condition occurring in at least 4% of the population (Simner et al., 2006) in which certain stimuli trigger unusual perceptions which the physical properties of the stimulus alone are not sufficient to account for. For instance, in *grapheme-color synesthesia*, the sight of black-and-white printed letters or numbers triggers the experience of consistent and specific colors (e.g., A might be red, 7 yellow: Smilek et al., 2001). The associations reported by any given synesthete tend to be remarkably consistent over time (e.g., Simner et al., 2005). Hence if A is red for a certain synesthete, it will tend to always be red. There is evidence for a genetic component to the condition (Ward and Simner, 2005; Asher et al., 2009; Tomson et al., 2011), and evidence too that the brains of synesthetes are structurally and functionally different to those of non-synesthetes. For example, in a *sound-color* and *sound-taste synesthete*, Hänggi et al. (2008) found altered patterns of brain volume and functional anisotropy (FA) values (indicative of greater white matter coherence). These included increased FA values in primary auditory cortex, as well as increased gray and white matter volumes in the same region, and also structural differences in gray and white matter in visual and gustatory regions, i.e., increases in occipital regions, and increases and decreases in insular cortex (for review see Hubbard, 2012; Rouw, 2012).

It has often been suggested that synesthesia may reveal something about normal cognition (e.g., Cohen Kadosh and Henik,

2007; Wilson, 2012). Ramachandran and Hubbard (2001) have even suggested that the origins of human language may lie in synesthetic-like correspondences between speech sounds and the physical properties of objects they refer to (see Cuskley and Kirby, 2013 for review). These theories usually describe synesthesia as a “blending of the senses,” in which an experience in one *sensory* modality triggers an experience in another *sensory* modality; for example sounds triggering colors. However, synesthesia is not always, or even usually, cross-sensory: the most common variants are in fact triggered by language (e.g., colors from graphemes or words; Simner et al., 2006). In such cases, the trigger is a cognitive (linguistic) construct rather than sensory percept (for related discussions see Simner, 2007 and Mroczko-Wasowicz and Nikolić, 2014). Indeed, many types of synesthesia involve conceptual elements—often linguistic—and we will argue that these variants may reveal information about the interface between perception and cognition.

One example of how synesthesia can link cognition and perception comes in the variant known as lexical-gustatory synesthesia. In this, tastes are experienced in the mouth when synesthetes hear, speak, read, or think about words (Ward and Simner, 2003). Simner and Ward (2006) showed that these sensations are tied to word lemmas, the abstract semantic/syntactic memory traces of words that are distinct from purely visual and acoustic information about the word-form. Given this, we might conclude that the perceptual taste

sensations of these synesthetes appear to be triggered by truly abstract cognitive thought, as a case of direct interaction between cognition and perception. Our proposal here begins with the suggestion (also made by others, see below) that these kinds of interactions may occur in all people, and that synesthetes may simply differ primarily in their conscious awareness of them.

Our proposal focusses particularly on theories of embodied cognition. These theories contend that conceptual thought and language are ‘embodied’ or ‘grounded’ in neural systems for perception and action (e.g., Glenberg and Kaschak, 2002; Fischer and Zwaan, 2008). Hence, hearing a word or thinking about a concept would trigger a “simulation” of stored perceptual or motor information (Barsalou, 2003; Glenberg and Gallese, 2012). The taste of lemon, for instance, would form part of the semantics of the word “lemon,” and the implicit mental activation of that taste would be linked to its phonological representation. Whereas in most people these perceptual connections are implicit, our hypothesis, first suggested by Simner and Ward (2006), is that in lexical-gustatory synesthetes for example, these same simulations somehow attain conscious awareness, so that hearing the word “lemon” actually evokes the taste.

We suggest therefore that normal “embodied” links between perceptions (taste of lemon) and words (“lemon”) may somehow be consciously experienced by synesthetes, due to some type of

disinhibited or over-exuberant activation which is normally subdued in the average person. Of course, this proposal would only initially account for words that have an inherent taste, color, or the like (e.g., “lemon”). The problem however is that for many synesthetes, almost *all words* (e.g., “man,” “house,” “reach,” “and”) induce tastes, often without any obvious connection to their meaning at all (e.g., “reach” might taste of fruit sweets). How might this be explained and how could such widespread tastes link back to the idea of embodiment? Importantly, Simner and Haywood (2009) have shown that, in the case of lexical-gustatory synesthesia, the first words to acquire tastes in childhood synesthesia are very likely to have been food words, which acquire tastes in a semantically-direct way—so “apple” tastes of apple, “peach” tastes of peach (see also Ward et al., 2005). Simner and Haywood suggest that these tastes, or variants thereof, then spread outwards to phonologically similar words that do not have tastes of their own (e.g., “beach” tastes of peach; “reach” tastes of peach-flavored fruit sweets; Simner and Haywood, 2009). Put differently, synesthetic tastes appear to spread throughout the lexicon along the same connections that facilitate phonological priming effects: hearing “beach” might activate the word “peach,” and thereby acquire the taste of peach. The more often this word is heard and the taste experienced, the more strongly the two would become connected, so that before long, synesthetes will differ from non-synesthetes in not only having conscious tastes for food words, but also in having tastes for other words in their mental lexicon too¹.

¹Our account best fits variants of synaesthesia which, like embodiment, are influenced by word meanings (e.g., lexical-gustatory synaesthesia). Other variants such as grapheme-color synaesthesia may be less relevant to our model because colors are driven by sublexical letter units: since letters have no semantic content they cannot obviously show embodiment effects—at least not on the surface. Nonetheless, embodiment might yet hold sway even in these variants. For example, in grapheme-color synaesthesia, words become colored by their initial letter (e.g., “orate” would be white if the letter “o” is white) but some words such as color-terms are semantically colored (“orange” is often colored orange). This follows our hypothesis, and suggests that embodiment might interact with other mechanisms even in variants of synaesthesia that would seem otherwise unsusceptible to its influences.

One little-known phenomenon that provides additional evidence for a link between synesthesia and embodied cognition is *vision-touch synesthesia* (Blakemore et al., 2005), in which people report a physical sensation of being touched themselves when they observe others being touched. It has also been called *mirror-touch synesthesia* by analogy with the “mirror neurons” found initially in the monkey premotor cortex, and hypothesized to exist also in humans. These fire both when performing a certain action and when seeing that action performed (see Rizzolatti and Craighero, 2004, for a review). “Mirror-touch” is a particularly illuminating example of synesthesia, because the underlying connectivity implicated—the mirror neuron system—is not something assumed to be peculiar to these synesthetes, but is hypothesized to play a key role in the imitation, prediction, and understanding of others’ behavior in the general population (e.g., Wilson and Knoblich, 2005). The embodied simulation theory of social cognition, for example, proposes that the mirror system, by internally “simulating” the motor and somatosensory states of conspecifics, allows us to understand them (Gallese and Sinigaglia, 2011). Indeed, mirror-touch synesthetes have been found to have greater empathy than controls, which has been interpreted as evidence that mirror neurons may indeed play such a role (Banissy and Ward, 2007). Banissy and colleagues have therefore argued that the normally implicit somatosensory stimulation triggered by observing others might be experienced in an extreme form in mirror-touch synesthetes². This view is highly compatible with our own, which also raises a parallel case for other forms of synesthesia (see also Simner and Ward, 2006).

So far, we have argued that the connections between perception and cognition seen in synesthesia might be present in all people, but inhibited in the general population, consistent with embodied theories

²A link between synaesthesia and the mirror neuron system has also been made in a different way by Mroczko-Wąsowicz and Werning (2012). They describe cases of synaesthesia where colors are experienced both from movement and from observing/imagining the movement of others. They link this to mirror systems as a way to best model the inducer, and they also discuss this with reference to sensori-motor contingencies (see also Seth, 2014).

of language and cognition. However, these theories are in fact best-known as they apply to action: i.e., that action concepts (and the semantics of action words) are also embodied, this time in the motor system. This widespread view of embodiment suggests that perceiving and understanding action words (e.g., “hit”) or indeed observing the actions of others, involves mentally simulating them. So if synesthesia is the result of disinhibited simulation in the sensory system, what would be the result of disinhibition of embodied simulation in the *motor* system? (Our hypothesis thus far certainly predicts there might be such a phenomenon.) In response we point to candidates from the range of behaviors known as “motor release phenomena.”

Motor release phenomena are a variety of syndromes involving automatic motor behaviors, most commonly seen following damage to the frontal lobes from stroke, but also observed in other populations with known or suspected dysfunction of frontal control systems. These include patients with various psychiatric illnesses and children with attention deficit hyperactivity disorder (Archibald et al., 2001). They can be broadly classified into three types (Lhermitte, 1983): *disinhibition of basic reflexes* such as manual groping, when the patient’s outstretched hand will follow an object being moved; *imitation behavior*, in which patients automatically and unwillingly find themselves copying observed actions; and *utilization behavior*, where patients compulsively pick up objects placed in their view, and either toy with them or use them for their intended purpose. These behaviors are perhaps most striking when they only affect one side of the body, such that one arm appears to act independently of its owner’s will, known as the famous “alien hand sign” or “anarchic hand syndrome” (e.g., Goldberg and Bloom, 1990).

All these phenomena are thought to arise from the disinhibition of circuits which link perceptual input to motor output, possibly through the mirror neuron system (Berthier et al., 2006; McBride et al., 2013). Evidence from masked priming tasks supports the view that these links are automatically and unconsciously inhibited in healthy people (e.g., McBride et al., 2013). Crucially for theories of

embodied cognition, there is evidence that this perception-action link is conceptually or linguistically mediated. Goldberg and Bloom (1990) reported several cases of alien hand sign where the alien hand performed actions that were merely mentioned by the investigator. Schaefer et al. (2013), too, report a patient who would pick up objects with her alien hand when told to by the experimenter, but was unable to do so of her own volition. Furthermore, some patients could exercise some control over their alien hand by “talking to it”: when their alien hand gripped an object, they were able to get it to release by telling it to “let go.” These observations fit well with the predictions of embodied cognition: if understanding sentences that refer to actions involves their simulation by the motor system, then disinhibition of these simulations would indeed lead to the actions described being overtly performed (McBride et al., 2013).

As noted above, theories of embodied cognition propose that the brain's systems for perception and action are used in cognition. We have suggested that both types of simulation can be disinhibited in some people, resulting in synesthesia in the case of the sensory system (see also Simner and Ward, 2006), and release phenomena in the case of the motor system (see also McBride et al., 2013). The question might be asked why the two types of disinhibition are not more analogous: why, for instance, do we not see cases of synesthesia often arising as a result of stroke? Our first answer is that the types of inhibition involved are somewhat different. In the case of motor embodiment, it is suggested that a motor plan is initiated, but not executed. In sensory embodiment, on the other hand, perceptual information is activated, but does not, in most people, reach conscious perceptual awareness. These two forms of “inhibition” are doubtlessly underpinned by very different mechanisms. Secondly, there are in fact cases of synesthesia-like experiences occurring in non-synesthetes following brain damage (e.g., Ro et al., 2007; Brogaard et al., 2013). There are also cases where it emerged after the patient became blind late in life (Armel and Ramachandran, 1999), or under the influence of hallucinogens (see Luke

and Terhune, 2013, for review), or after long experience with meditation (Walsh, 2005). They have even been experimentally induced by post-hypnotic suggestion (Cohen Kadosh et al., 2009). While there is disagreement as to the extent to which developmental, acquired, and other forms of synesthesia might share common mechanisms (Sinke et al., 2012, but cf. Brogaard, 2013), these examples nonetheless support the hypothesis that the cognition-perception links seen in synesthesia may exist in some inhibited form in all people, and may become “released” following trauma, disease or unusual environmental interactions.

In summary, motor release phenomena have been interpreted as evidence that the motor system is involved when observing and describing actions, and in object affordances (e.g., McBride et al., 2013). We have proposed that, similarly, some types of synesthesia suggest that the sensory system, too, may play a role in language and conceptual thought. We have therefore proposed a relationship between synesthesia and release phenomena, in that each may be considered in terms of disinhibited embodiment in sensory and motor systems respectively. Overall, our arguments suggest that synesthesia may represent a case *par excellence* of the cognition-perception interface, showing an outward perceptual manifestation of implicit associations that lie at the heart of embodied cognition.

REFERENCES

- Archibald, S., Mateer, C., and Kerns, K. (2001). Utilization behavior: clinical manifestations and neurological mechanisms. *Neuropsychol. Rev.* 11, 117–130. doi: 10.1023/A:1016673807158
- Armel, K. C., and Ramachandran, V. S. (1999). Acquired synesthesia in retinitis pigmentosa. *Neurocase* 5, 293–296. doi: 10.1080/13554799908411982
- Asher, J., Lamb, J., Brocklebank, D., Cazier, J., Maestrini, E., Addis, L., et al. (2009). A whole-genome scan and fine-mapping linkage study of auditory-visual synesthesia reveals evidence of linkage to chromosomes 2q24, 5q33, 6p12, and 12p12. *Am. J. Hum. Genet.* 84, 279–285. doi: 10.1016/j.ajhg.2009.01.012
- Banissy, M. J., and Ward, J. (2007). Mirror-touch synesthesia is linked with empathy. *Nat. Neurosci.* 10, 815–816. doi: 10.1038/nn1926
- Barsalou, L. W. (2003). Situated simulation in the human conceptual system. *Lang. Cogn. Process.* 18, 1. doi: 10.1080/769813547
- Berthier, M. L., Pulvermüller, F., Green, C., and Higuera, C. (2006). Are release phenomena explained by disinhibited mirror neuron circuits? Arnold Pick's remarks on echographia and their relevance for modern cognitive neuroscience. *Aphasiology* 20, 462–480. doi: 10.1080/02687030500484004
- Blakemore, S., Bristow, D., Bird, G., Frith, C., and Ward, J. (2005). Somatosensory activations during the observation of touch and a case of vision-touch synesthesia. *Brain* 128, 1571–1583. doi: 10.1093/brain/awh500
- Brogaard, B. (2013). Serotonergic hyperactivity as a potential factor in developmental, acquired and drug-induced synesthesia. *Front. Hum. Neurosci.* 7:657. doi: 10.3389/fnhum.2013.00657
- Brogaard, B., Vanni, S., and Silvanto, J. (2013). Seeing mathematics: perceptual experience and brain activity in acquired synesthesia. *Neurocase* 19, 566–575. doi: 10.1080/13554794.2012.701646
- Cohen Kadosh, R., and Henik, A. (2007). Can synesthesia research inform cognitive science? *Trends Cogn. Sci.* 11, 177–184. doi: 10.1016/j.tics.2007.01.003
- Cohen Kadosh, R., Henik, A., Catena, A., Walsh, V., and Fuentes, L. J. (2009). Induced cross-modal synesthetic experience without abnormal neuronal connections. *Psychol. Sci.* 20, 258–265. doi: 10.1111/j.1467-9280.2009.02286.x
- Cuskley, C., and Kirby, S. (2013). “Synesthesia, cross-modality, and language evolution,” in *Oxford Handbook of Synesthesia*, eds J. Simner and E. Hubbard (Oxford: Oxford University Press), 869–899. doi: 10.1093/oxfordhb/9780199603329.013.0043
- Fischer, M. H., and Zwaan, R. A. (2008). Embodied language: a review of the role of the motor system in language comprehension. *Q. J. Exp. Psychol.* 61, 825. doi: 10.1080/17470210701623605
- 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
- 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
- Goldberg, G., and Bloom, K. K. (1990). The alien hand sign: localization, lateralization and recovery. *Am. J. Phys. Med. Rehabil.* 69, 228–238. doi: 10.1097/00002060-199010000-00002
- Hänggi, J., Beeli, G., Oechslin, M. S., and Jäncke, L. (2008). The multiple synaesthete, E. S.: neuroanatomical basis of interval-taste and tone-colour synaesthesia. *Neuroimage* 43, 192–203. doi: 10.1016/j.neuroimage.2008.07.018
- Hubbard, E. M. (2012). “Synesthesia and functional imaging,” in *The Oxford Handbook of Synesthesia*, eds J. Simner, and E. M. Hubbard (Oxford: Oxford University Press), 475–499.
- Lhermitte, F. (1983). Utilization behavior and its relation to lesions of the frontal lobes. *Brain* 106, 237–255. doi: 10.1093/brain/106.2.237
- Luke, D. P., and Terhune, D. B. (2013). The induction of synaesthesia with chemical agents: a systematic review. *Front. Psychol.* 4:753. doi: 10.3389/fpsyg.2013.00753
- McBride, J., Sumner, P., Jackson, S. R., Bajaj, N., and Husain, M. (2013). Exaggerated object

- affordance and absent automatic inhibition in alien hand syndrome. *Cortex* 49, 2040–2054. doi: 10.1016/j.cortex.2013.01.004
- Mroczko-Wąsowicz, A., and Nikolić, D. (2014). Semantic mechanisms may be responsible for developing synesthesia. *Front. Hum. Neurosci.* 8:509. doi: 10.3389/fnhum.2014.00509
- Mroczko-Wąsowicz, A., and Werning, M. (2012). Synesthesia, sensory-motor contingency, and semantic emulation: how swimming style-color synesthesia challenges the traditional view of synesthesia. *Front. Psychol.* 3:279. doi: 10.3389/fpsyg.2012.00279
- Ramachandran, V. S., and Hubbard, E. M. (2001). Synesthesia – a window into perception, thought and language. *J. Conscious. Stud.* 8, 3–34.
- Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192. doi: 10.1146/annurev.neuro.27.070203.144230
- Ro, T., Farnè, A., Johnson, R. M., Wedeen, V., Chu, Z., Wang, Z. J., et al. (2007). Feeling sounds after a thalamic lesion. *Ann. Neurol.* 62, 433–441. doi: 10.1002/ana.21219
- Rouw, R. (2012). “Synesthesia, hyper-connectivity and diffusion tensor imaging,” in *The Oxford Handbook of Synesthesia*, eds. J. Simner, and E. M. Hubbard (Oxford: Oxford University Press), 500–518.
- Schaefer, M., Heinze, H. J., and Galazky, I. (2013). Waking up the alien hand: rubber hand illusion interacts with alien hand syndrome. *Neurocase* 19, 371–376. doi: 10.1080/13554794.2012.667132
- Seth, A. K. (2014). A predictive processing theory of sensorimotor contingencies: explaining the puzzle of perceptual presence and its absence in synaesthesia. *Cogn. Neurosci.* 5, 97–118. doi: 10.1080/17588928.2013.877880
- Simner, J. (2007). Beyond perception: synaesthesia as a psycholinguistic phenomenon. *Trends Cogn. Sci.* 11, 23–29. doi: 10.1016/j.tics.2006.10.010
- Simner, J., and Haywood, S. L. (2009). Tasty non-words and neighbours: the cognitive roots of lexical-gustatory synesthesia. *Cognition* 110, 171–181. doi: 10.1016/j.cognition.2008.11.008
- Simner, J., Mulvenna, C., Sagiv, N., Tsakanikos, E., Witherby, S. A., Fraser, C., et al. (2006). Synesthesia: the prevalence of atypical cross-modal experiences. *Perception* 35, 1024–1033. doi: 10.1068/p5469
- Simner, J., and Ward, J. (2006). Synesthesia: the taste of words on the tip of the tongue. *Nature* 444, 438. doi: 10.1038/444438a
- Simner, J., Ward, J., Lanz, M., Jansari, A., Noonan, K., Glover, L., et al. (2005). Non-random associations of graphemes to colors in synesthetic and non-synesthetic populations. *Cogn. Neuropsychol.* 22, 1069–1085. doi: 10.1080/02643290500200122
- Sinke, C., Halpern, J. H., Zedler, M., Neufeld, J., Emrich, H. M., and Passie, T. (2012). Genuine and drug-induced synesthesia: a comparison. *Conscious. Cogn.* 21, 1419–1434. doi: 10.1016/j.concog.2012.03.009
- Smilek, D., Dixon, M. J., Cudahy, C., and Merikle, P. M. (2001). Synesthetic photisms influence visual perception. *J. Cogn. Neurosci.* 13, 930–936. doi: 10.1162/089892901753165845
- Tomson, S. N., Avidan, N., Lee, K., Sarma, A. K., Tushe, R., Milewicz, D. M., et al. (2011). The genetics of colored sequence synesthesia: suggestive evidence of linkage to 16q and genetic heterogeneity for the condition. *Behav. Brain Res.* 223, 48–52. doi: 10.1016/j.bbr.2011.03.071
- Walsh, R. (2005). Can synesthesia be cultivated? Indications from surveys of meditators. *J. Conscious. Stud.* 12, 5–17.
- Ward, J., and Simner, J. (2003). Lexical-gustatory synesthesia: linguistic and conceptual factors. *Cognition* 89, 237–261. doi: 10.1016/S0010-0277(03)00122-7
- Ward, J., and Simner, J. (2005). Is synesthesia an X-linked dominant trait with lethality in males? *Perception* 34, 611–623. doi: 10.1068/p5250
- Ward, J., Simner, J., and Auyeung, V. (2005). A comparison of lexical-gustatory and grapheme-color synesthesia. *Cogn. Neuropsychol.* 22, 28–41. doi: 10.1080/02643290442000022
- Wilson, M. (2012). What do the mirror system, embodied cognition, and synesthesia have to do with each other? *Cortex* 49, 2949–2950. doi: 10.1016/j.cortex.2013.07.003
- 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

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 October 2014; accepted: 12 January 2015; published online: 30 January 2015.

Citation: Gray BF and Simner J (2015) Synesthesia and release phenomena in sensory and motor grounding. *Cases of disinhibited embodiment?* *Front. Psychol.* 6:54. doi: 10.3389/fpsyg.2015.00054

This article was submitted to *Consciousness Research*, a section of the journal *Frontiers in Psychology*.

Copyright © 2015 Gray and Simner. 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.



Revisiting the empirical case against perceptual modularity

Farid Masrour*, Gregory Nirshberg, Michael Schon, Jason Leardi and Emily Barrett

Department of Philosophy, University of Wisconsin-Madison, Madison, WI, USA

OPEN ACCESS

Edited by:

Aleksandra Mroczko-Wąsowicz,
National Yang-Ming University, Taiwan

Reviewed by:

John Bickle,
Mississippi State University, USA
Marcus Rothkirch,
Charité - Universitätsmedizin Berlin,
Germany
Jacob Berger,
Idaho State University, USA

*Correspondence:

Farid Masrour
masrour@wisc.edu

Specialty section:

This article was submitted to
Consciousness Research,
a section of the journal
Frontiers in Psychology

Received: 08 May 2015

Accepted: 19 October 2015

Published: 04 November 2015

Citation:

Masrour F, Nirshberg G, Schon M,
Leardi J and Barrett E (2015)
Revisiting the empirical case against
perceptual modularity.
Front. Psychol. 6:1676.
doi: 10.3389/fpsyg.2015.01676

Some theorists hold that the human perceptual system has a component that receives input only from units lower in the perceptual hierarchy. This thesis, that we shall here refer to as the *encapsulation thesis*, has been at the center of a continuing debate for the past few decades. Those who deny the encapsulation thesis often rely on the large body of psychological findings that allegedly suggest that perception is influenced by factors such as the beliefs, desires, goals, and the expectations of the perceiver. Proponents of the encapsulation thesis, however, often argue that, when correctly interpreted, these psychological findings are compatible with the thesis. In our view, the debate over the significance and the correct interpretation of these psychological findings has reached an impasse. We hold that this impasse is due to the methodological limitations over psychophysical experiments, and it is very unlikely that such experiments, on their own, could yield results that would settle the debate. After defending this claim, we argue that integrating data from cognitive neuroscience resolves the debate in favor of those who deny the encapsulation thesis.

Keywords: perception, modularity, encapsulation, cognitive penetration, vision disorders, psychophysics

INTRODUCTION

Participants in the debate over whether cognition can influence perception could be roughly divided into two camps. One camp consists of those who emphasize that the information that is received through the senses is not sufficient to uniquely determine the correct hypothesis about its distal causes. Their proposal is that in order to solve this under-determination problem, the inputs of perception must be supplemented by more information, such as the background beliefs of the perceiver. Perceptual processes thus have access to central cognition and are susceptible to influence by cognition. The opposing camp, which we refer to as the modularists, acknowledges that the input to the perceptual system needs to be supplemented to solve the under-determination problem. Nevertheless, they hold that the additional information is localized to the perceptual module.

A central intuition behind the modularist approach is that evaluating the incoming stimulus in light of a large body of information is time consuming and costly. And since a well-designed perceptual system should enable fast responses to changing environmental conditions, it must lack access to the totality of information that a perceiver has. Imagine contemplating the possible visual tricks that someone might have played on you when it seems that a lion is about to attack you on a safari. Even if that sort of contemplation is something that *you* are prone to do, it is better if *your perceptual system* does not have this propensity, and one way to guarantee that it does not is to limit its access to what is directly relevant to its domain of input. This is one of the main insights behind Fodor's original distinction between what he calls "input analyzers" and "central cognition." Input analyzers are in the business of fast analysis of incoming data on the basis of a limited body of information, which is typically domain specific, innate, and localized.

Central cognition, in contrast, is in the business of belief formation, problem solving, contemplation, etc., on the basis of information that is often domain general, learned, and non-localized.

Our focus in this paper will be on the empirical evidence pertaining to a thesis that is at the core of the modularity debate. This is the thesis that the human perceptual system has components that are informationally encapsulated. Relying on Fodor's terminology, we call these components modules¹. As a first approximation, the claim that a module is informationally encapsulated means that the processes within the module have access only to the contents of other processes within the module as well as input from earlier units in the perceptual hierarchy². Throughout the paper, we refer to the thesis that the perceptual system has at least one informationally encapsulated module as the *generic encapsulation thesis*.

If the encapsulation thesis is correct, then the functioning of the visual module should be free from content-sensitive influence both from other modules and cognitive processes "up stream" in the visual hierarchy. However, there is a large body of research, starting in the mid-twentieth century and extending to the present date, which purports to establish that perception can be influenced by cognitive factors such as the stimulus's meaning, its familiarity, its predictability, the context it appears in, the concepts the perceiver uses to categorize it, and more (for some recent findings and reviews, see Hansen et al., 2006; Levin and Banaji, 2006; Lupyan and Spivey, 2008; Lupyan et al., 2010; Macpherson, 2012; Lupyan and Ward, 2013; Stokes, 2013; For earlier findings on the topic, see Bruner, 1957, 1973; Rock, 1983; Goldstone, 1995; and the extensive review in Pylyshyn, 1999).

The vast majority of this research comes from psychophysical studies that track potential cognitive effects on perceptual performance³. Modularists rarely deny the effects that these studies purport to show. Rather, their main strategy has been to argue that the results of these studies can be interpreted in ways that are compatible with their preferred version of the encapsulation thesis. In fact, a quick survey of the history of the debate over the empirical evidence reveals a pattern in which anti-modularists produce new results which are subsequently explained away by the modularists⁴.

We agree with the modularists that the many of the existing psychophysical findings could be explained away. However, we disagree with the modularists that this indicates that the encapsulation thesis is correct. In our view, reflection on the general methodological limitations of psychophysical

experiments shows that it is very unlikely that such experiments, on their own, could yield results that could not be explained away by the modularists. It is therefore not the truth of the thesis, but the nature of psychophysical methodology, that is the underlying cause of the impasse. Our first goal in this paper is to defend this claim. However, we do not think that one should be skeptical of the possibility of an empirical resolution to this debate. In fact, we believe that integrating data from cognitive neuroscience will very likely resolve the debate in favor of the anti-modularists. Defending this claim is the second goal of this paper.

We have defined the generic encapsulation thesis as the thesis that there is an informationally encapsulated component of the perceptual system, namely a module. However, for reasons to be discussed soon, this paper focuses on a much more specific variant of the generic encapsulation thesis. This variant, roughly characterized, holds that there is an informationally encapsulated component of the perceptual system that gives rise to access conscious representations.

Here is how we will proceed. In the next section, we further explicate the generic encapsulation thesis, distinguish among some of its interesting variants, and elaborate on the link between these variants and some broader theoretical issues surrounding the modularity debate. This discussion helps clarify and justify our choice for focusing on the specific version of the thesis in the rest of the paper. However, those readers who are interested in quickly getting to our main arguments can skip straight to the "Psychological Case Against Encapsulation" section which focuses on the psychophysical evidence. In that section, we identify three core demands that psychophysical studies must meet in order to establish the failure of encapsulation. We then argue that the nature of these demands make it very unlikely that purely psychophysical studies would be sufficient to reject the thesis. Our methodological conclusion is that the empirical approach to the debate has to draw on data from sources other than purely psychophysical studies. This is what we shall do in the "Neural Case Against Encapsulation" section where we draw on recent findings in neuroscience and argue that they militate against the version of the encapsulation thesis that we describe in the next section.

THE ENCAPSULATION THESIS

Earlier we defined the generic encapsulation thesis as the thesis that there is an informationally encapsulated component of the perceptual system, namely a module. A module is informationally encapsulated in the sense that its processes have access only to the contents of other processes within the module and the information provided by earlier units in the perceptual hierarchy. As we use the term, processes are transitions from one set of contentful states to another. A contentful state is one that can be assessed for veridicality. Under this definition, a chemical or an electrical event is not a process. When a content enters into the proximal explanation for why a process happens as it does, that process is said to access that content⁵. Suppose that in order

¹This is not to say that this property exhausts or is definitional of what a module is. Fodor (1983), for example, attributes nine properties to modules one of which is informational encapsulation.

²We hold that any information that a process receives is, by definition, input to the process. So we think that defining informational encapsulation in terms of having access only to inputs makes the thesis trivially true and thus non-substantive.

³An exception to this is Raftopoulos (2009, 2014) who engages with the evidence coming from neuroscience.

⁴See Pylyshyn (1999), Durgin et al. (2009, 2012), Raftopoulos (2009), Zeimbekis (2013), and Firestone and Scholl (2014, 2015a,b) for examples of modularist replies. For a review of some of the empirical exchange between modularists and anti-modularists see Witt et al. (2015).

⁵Note that this is different from saying that the content enters the explanation of why the process is the process that it is.

to explain why John draws the conclusion that Steve is mortal from the premise that Steve is a human being, you use a syllogism that relies on the premise that all human beings are mortal. This is a proximal explanation. Therefore, in our terminology, you are assuming that John's transition from the belief state with the content "Steve is a human being" to the state with the content "Steve is mortal" is caused by John accessing the content "All human beings are mortal." Now, suppose John formed the belief that all human beings are mortal from reading a book, and he read the book because he believed that books are useful. The belief that books are useful has thus entered into an explanation for why John transitions from the former belief state to the latter, but this explanation is not a proximal explanation. Similarly, there are non-proximal explanations of perceptual processes that involve content that do not imply that those perceptual processes access that content.

The idea behind the generic encapsulation thesis is therefore that all the transitions between contentful states in the module are proximally explained in terms of contents of states within the module or the input that the module receives from earlier units in the perceptual hierarchy. We shall say more about the perceptual hierarchy in the "Neural Case Against Encapsulation" section.

Due to its generality, the generic encapsulation thesis is too easy to satisfy and seems to be theoretically unattractive. For example, one might easily find a small neuronal assembly or a single neuron in the visual system that is informationally encapsulated. However, it is not clear how the truth of such a thesis would relate to the broader theoretical issues that are often linked to the modularity debate. Some of these theoretical issues include questions about the structure and function of the perceptual system, whether perception is a bottom-up or a top-down process, whether observation is theory-neutral, whether perceptual experience has conceptual content, and whether a foundationalist epistemology is tenable⁶. Rather than focusing on the generic encapsulation thesis, we should therefore focus on its specific variants. We can obtain such specific variants by imposing further functional constraints on the module. If these constraints are properly related to the theoretical issues surrounding the modularity debate, the resulting variants of the thesis would be more theoretically interesting.

In his canonical statement of the modularity thesis in *The Modularity of Mind*, Fodor argues that the purpose of the perceptual module is to provide fast analysis of sensory input on the basis of informationally encapsulated, domain specific, innate, and localized informational processes. The representations that result from this analysis are subsequently provided as input to higher cognitive centers and the action system.

The idea that the outputs of the perceptual modules serve as input to higher systems imposes a functional constraint on the notion of the module. Let us call a representation that could be used by higher cognitive centers and the action system without further processing a *pickup ready representation*⁷. In this case, the

functional constraint is that the outputs of the perceptual modules should be pickup ready representations. Adding this constraint to the generic encapsulation thesis would give us the following variant:

Pickup Ready Encapsulation

There is a component of the perceptual system that gives rise to pickup ready representations and is informationally encapsulated.

A module's functional role in providing input to other systems is not the only theoretical issue that is relevant to the encapsulation thesis. One reason for the earlier surge of philosophical interest in modularity has been whether observation is theory-laden, with some modularists arguing that encapsulation suggests that observation is theory-neutral (see Fodor, 1984; Raftopoulos, 2001). But obviously, not just any form of the encapsulation thesis would have this implication. Suppose, for example, that Pickup Ready Encapsulation is correct and there is a visual module that gives rise to pickup ready representations. Still, it might be the case that the outputs of this alleged module are *pre-observational* in the sense that there is more processing that has to happen over these outputs before they give rise to what can be properly regarded as observation. Since there is still the possibility that these additional processes are not encapsulated, pickup ready encapsulation and the theory-ladenness theses could both be correct at the same time.

The same observation applies to the link between the encapsulation thesis and epistemic issues. Siegel (2012), for example, has argued that some types of encapsulation failure threaten a foundationalist approach in perceptual epistemology. If perception is influenced by background beliefs, then a potentially problematic circularity threatens the idea that perception can serve as the foundation for justifying beliefs. Thus, a foundationalist who accepts Siegel's arguments might be interested in saving encapsulation. But again, for similar reasons to those that show that it does not disprove the theory-ladenness thesis, Pickup Ready Encapsulation may not serve this purpose.

We therefore need to impose further constraints on Pickup Ready Encapsulation to establish a better link with the above theoretical issues. One possible constraint is to restrict encapsulation to person-level representations. Adding this constraint would give us the following thesis.

Person-Level Encapsulation

There is a component of the perceptual system that gives rise to person-level pickup ready perceptual states and is informationally encapsulated.

The above version of the thesis gets closer to the theoretical issues surrounding encapsulation. However, those who submit

rise to what he calls "proto-objects" which are themselves the products of two sets of processes within the early visual system: those that compute features such as color, depth, luminance, and motion (Pylyshyn, 1999, p. 361–362), and those that bind these features into a single proto-object by top-down and horizontal processes within early vision. The binding of the features into a proto-object creates representations that are stable and robust enough to be compared with representations in long-term memory for use in recognition and identification (Pylyshyn, 2001, p. 145).

⁶For some discussion of the theoretical issues concerning the modularity debate (see Fodor, 1984; Churchland, 1988; Fodor and Pylyshyn, 1988; Raftopoulos, 2001; Siegel, 2012; Stokes, 2013).

⁷It is not completely obvious what could serve this function. But there are some proposals on the table. For example, on Pylyshyn's view, modules have to give

to common forms of internalist epistemology would not find this move very satisfactory. On such views, a state is capable of epistemic evaluation in so far as it belongs to a domain that a subject can access through reflection. Those who accept this form of internalism also have a tendency to demand that a state counts as an observation in so far as it is available to reflection. This suggests stronger constraints on the encapsulation thesis; constraints that would link it to the possibility of reflection. One such constraint is that the outputs of the encapsulated module have to be phenomenally conscious. This is because there is an intuitive connection between phenomenal consciousness and observation. One might even say that it is nomologically necessary that a subject, *S*, observes an object (or property) only if the subject has a phenomenally conscious representation of the object (or property). In fact, it is not uncommon to characterize the effect of cognition on perception in terms of its effect on the phenomenal content of perceptual experience (for example, see Macpherson, 2012). Adding this constraint will give us a new thesis.

Phenomenal Encapsulation

There is a component of the perceptual system that gives rise to phenomenally conscious representations and is informationally encapsulated.

However, those who think that there is a gap between phenomenal consciousness and access consciousness might question the link between phenomenal encapsulation and observation (see Block, 1995, 2005). The reason could be that in order for a representation to qualify as an observation, it should be access conscious in the sense that it has to be readily available for verbal report, voluntary control of action, and other personal-level functions⁸. In this sense, however, a phenomenally conscious representation might not be access conscious. We should therefore distinguish Phenomenal Encapsulation from another version of the generic thesis.

Access Encapsulation

There is a component of the perceptual system that gives rise to access conscious representations and is informationally encapsulated.

So far we have four versions of the encapsulation thesis. These theses are somewhat independent in that one might be false while the others are true. Are there any of these theses that should be regarded as the central encapsulation thesis? We think the answer is negative. A pluralist view according to which there are different versions of the encapsulation thesis that respond to different theoretical demands is, in our view, the correct view. Nevertheless, our focus in this paper will be on Access Encapsulation. This is not because we think perceptual states have to be access conscious. Our main reason is that in comparison to the other versions of the thesis, Access Encapsulation bears the most clear relationship

to the broader philosophical issues surrounding the modularity debate. It's rather intuitive that the failure of Access Encapsulation, even if turns out that the other versions of the theses are true would be philosophically significant. But whether the failure of other forms of encapsulation in those cases where Access Encapsulation remains intact has important philosophical implications or not is less intuitively clear. So our focus in this paper will be on access encapsulation.

The possible constraints that we have discussed so far concern the output of the module. Obviously, output is not the only feature that is relevant for determining the function of a module. One also needs to determine what serves as input to a module. It might seem natural to assume that the input to the perceptual modules should be equated with the physical energy at the level of sensory receptors. On such a view, modules start where sensory receptors start. But this is not the only option. Fodor's distinction between transducers and input analyzers can help us see why (Fodor, 1983). On Fodor's view, transducers are in the business of transforming one form of physical energy—for instance, luminance and hue—to another—for example, action potentials in the neuronal axons. According to Fodor, transducers do not perform any computation, and since perceptual modules start where computation starts, modules do not start at the boundaries of sensory organs. The same could be said about any units that transfer and relay the signal from transducers without performing any computations over them. So on a broadly Fodorian view, modules would be higher up in the perceptual hierarchy after units that transduce, transfer, and relay the signal without performing computations.

Fodor's distinction between the components of the perceptual system that perform computation and those that do not is controversial. Even if we accept it, it is not completely clear why the lower boundaries of the module have to coincide with the boundaries of computation. But the point of mentioning Fodor's distinction here is not to defend or reject it. The point is that there are at least two possible ways to draw the lower boundaries of a module. One option is to hold that the visual module starts where the boundaries of the sensory organs start. The other option is to hold that the visual module starts further upstream in the visual system. One principled way to determine where it starts would be to point to where computation starts. We thus get two new versions of the encapsulation thesis.

Primary Encapsulation

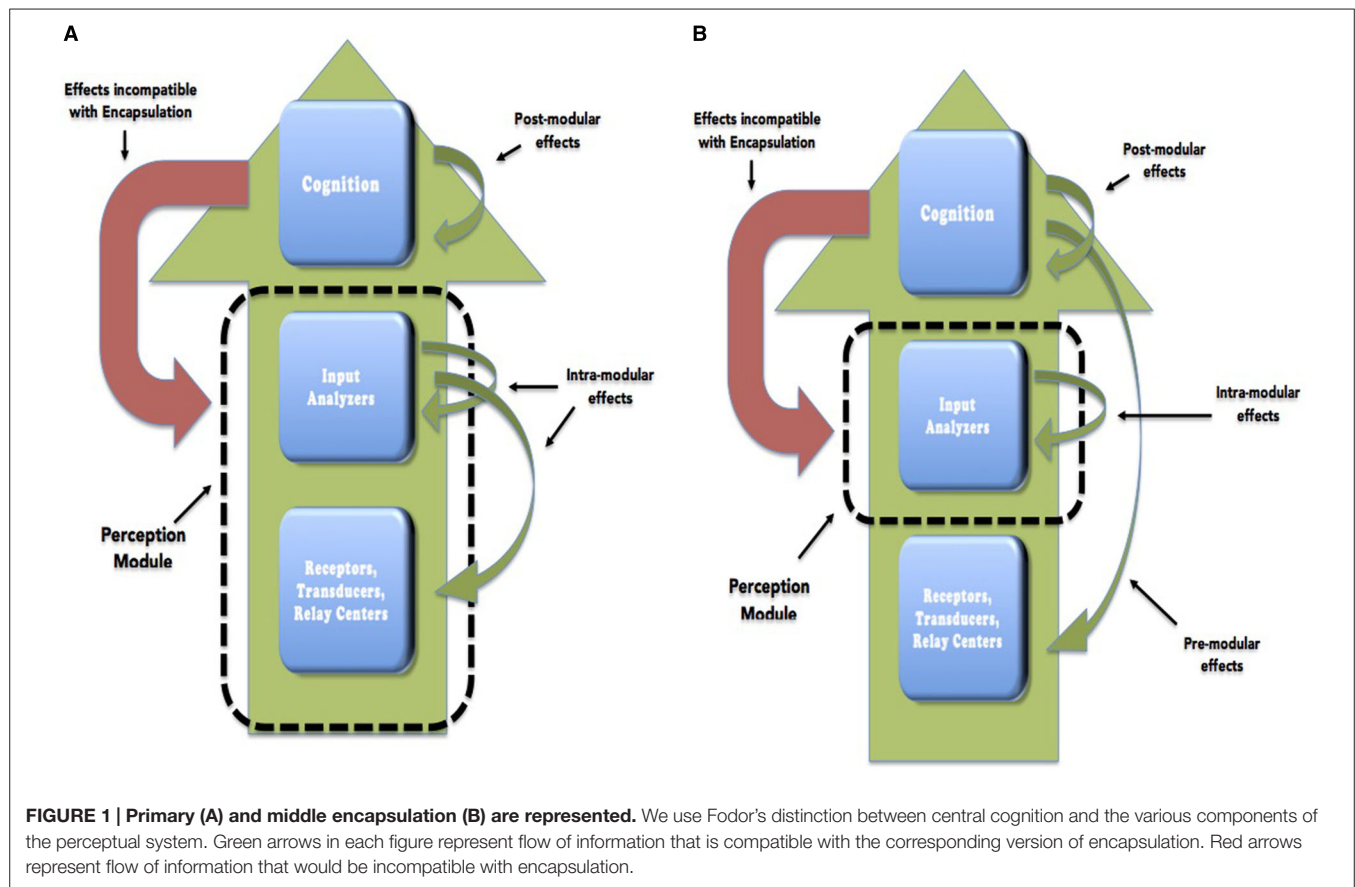
There is a component of the perceptual system whose lower boundaries coincide with the sensory receptors and is informationally encapsulated.

Middle Encapsulation

There is a component of the perceptual system whose lower boundaries are further upstream from transducers, transformers and relay centers, and is informationally encapsulated.

Obviously, the significance and the strength of the encapsulation thesis also depends on whether it is a primary form of encapsulation or a middle form. In this paper we will

⁸Note that a representation could be pickup ready without being access conscious. For example, a representation might be ready to be picked up by systems that are responsible for involuntary action without being access conscious. If, as Milner and Goodale (1995) have argued, the representations in the dorsal stream could control involuntary action without being conscious, the dorsal stream representations are pickup ready but not access conscious.



consider the empirical status of both theses (see **Figure 1** for an illustration). Unless explicitly noted otherwise, when we refer to the encapsulation thesis we will mean the disjunction of Primary Access Encapsulation and Middle Access Encapsulation. However, middle encapsulation is the dominant thesis among modularists and we will pay more attention to it. Furthermore our discussion will mainly focus on vision. Accordingly, we will be focusing on whether there is an encapsulated visual module.

One last point before we move on deserves emphasis. The encapsulation thesis is stronger than the oft-discussed thesis that perceptual modules are cognitively impenetrable (see Pylyshyn, 1999; Payne, 2001; Raftopoulos, 2001, 2014; Macpherson, 2012; Stokes, 2013). The main difference here is that informational encapsulation does not only concern cognitive contents. If the cognitive impenetrability thesis is true, then no cognitive state from outside the module can be accessed by processes inside the module. On the other hand, if the encapsulation thesis is true, then no contentful state outside the module, whether it can be regarded as cognitive or not, can be accessed by the processes within the module.

THE PSYCHOLOGICAL CASE AGAINST ENCAPSULATION

Our goal in the following sections is to show that purely psychophysical studies cannot provide sufficient evidence against

encapsulation. We do so by identifying three challenges that psychophysical studies must meet. They are the post-perceptual challenge, the intra-modular challenge and the pre-modular challenge. We then argue that no psychophysical study could jointly meet all three of them.

The Post-Perceptual Challenge

Consider an experiment in which subjects are shown images in which an individual is holding an object that is difficult to identify. Suppose that the results show that whether the subjects would report seeing a gun or a tool depends on the race of the individual holding the object (Payne, 2001). One interpretation of these findings is that implicit racial biases affect the percept, or, in other words, the way the object is seen⁹. Another interpretation is that implicit racial biases do not influence the way the object is seen, but only influence the subject's post-perceptual judgment about the identity of the object. The difference between these interpretations is relevant to the encapsulation thesis. The first interpretation is potentially incompatible with the encapsulation thesis because it shows that the processes that give rise to a percept can be influenced by factors outside the alleged visual module. However, on the second interpretation, the effect that implicit racial biases have on the subject's performance is mediated by its effect on post-perceptual judgment. Thus, this interpretation

⁹By a percept we mean an access conscious representation in the sense identified in the "Encapsulation Thesis" section.

is compatible with the encapsulation thesis. Therefore, in order to show that this finding is potentially troublesome for the encapsulation thesis we need to be able to rule out the post-perceptual interpretation.

The point of this example could be easily generalized. In the above case, the percepts and the behavioral response are mediated by judgment. However, the processes that influence the link between percepts and behavioral responses are not confined to judgment. In principle, it is possible for other mechanisms such as attention, inference, recognition, memory, various forms of response bias, etc., to intervene between the output of a module and a behavioral response. Effects that result from cognitive influences on the processes that link percepts to behavioral response should not count as evidence against the encapsulation thesis. A successful experiment has to be able to determine that the effect does not occur at a post-perceptual level. We therefore call this the post-perceptual challenge.

Meeting the post-perceptual challenge is not easy. Psychophysical studies of effects on perception have to rely on studies of perceptually-guided behavior. But effects on perceptually-guided behavior can happen in two ways: either by affecting the processes that lead to the formation of a percept or by affecting the processes that translate the percept into a behavioral response. Obviously, effects that result from cognitive impact on the latter stage should not count as evidence against encapsulation. There is, therefore, always a possible post-perceptual explanation that needs to be ruled out. Psychophysical experiments must meet this challenge if they are to provide evidence against the encapsulation thesis.

Some theorists have argued that one strategy to meet the post-perceptual challenge is to diminish the number of tasks that are needed to link a percept to a behavioral response. Stokes (2013), for example, argues that in some studies where the stimulus is present during the response phase, it would be implausible to hold that the pattern of response emerges from post-perceptual factors. For example, in Levin and Banaji (2006), subjects are asked to match the degree of luminance of a grayscale patch to the degree of luminance of pictures of faces. These studies show that the presence of labels (or typical race-indicating facial features) influence matching behavior. Faces that are labeled BLACK (or exemplify typical black facial features) are matched to darker patches than faces that are identical in luminance but are labeled WHITE (or show typical white facial features). Stokes argues that it would be implausible to explain away these results as emerging from post-perceptual judgment. The main reason is that such explanations would have the implication that although subjects have distinct phenomenal experiences of the luminance of the gray scale patch and the picture they match to it, they classify them as having the same luminance. This, in Stokes's view, renders these interpretations comparatively less plausible.

It is correct that post-perceptual explanations in the above match-to-sample experiments have a lower degree of plausibility in comparison to experiments that rely on memory, but this is clearly insufficient to show that they are implausible explanations *per se*. We should note that there could be phenomenally non-identical experiences that are nonetheless phenomenally

indistinguishable. In other words, a subject might have different experiences that she cannot distinguish from each other¹⁰. If this is plausible, then it is plausible that for any particular experience that a subject has, there is a range of phenomenally non-identical but indistinguishable experiences that the subject could match to this experience in an experimental setting. In other words, as long as two experiences are phenomenally indistinguishable, it is not implausible that a subject would classify them as identical, even when the experiences are non-identical. Perhaps Stokes thinks that post perceptual explanations, in these cases, are implausible because they imply that two phenomenally distinguishable experiences are judged to be identical. However, there is nothing in the aforementioned studies that demonstrates this. Therefore a post-perceptual explanation in these cases is not clearly implausible^{11,12}.

A second possible strategy to rule out post-perceptual explanations is to draw on the resources of signal detection theory (SDT). In fact after the advent of SDT, some psychologists quickly started to use this theory to distinguish perceptual effects from post-perceptual ones¹³. However, we believe that the perceptual vs. post-perceptual distinction that is based on SDT criteria is orthogonal to the perceptual vs. post-perceptual distinction that is operative in the debate between the modularists and the anti-modularists. Let us elaborate on this by first explaining why some have thought that SDT can help us distinguish between perceptual and post-perceptual processes.

Those who apply SDT to perception assume that our response mechanisms have the appropriate built-in structure to distinguish signal from noise. It is further assumed that this feature of response mechanisms can be used to distinguish effects on the response stage from effects on prior stages.

Imagine a task in which subjects are tasked with discriminating between pictures of dogs and non-dogs by pressing a button. Suppose subjects are more prone to classify images as dogs if they hear a story involving dogs before seeing the pictures. How can we figure out whether this is the result of effects on a perceptual detection stage during which a perceptual unit detects the presence of dogs or a post-perceptual response

¹⁰Disjunctivists are well-known for distinguishing between phenomenal distinctness and indistinguishability. But one does not have to be a disjunctivist in order to distinguish between phenomenal identity and phenomenal indistinguishability. One might independently be wedded to the view that phenomenal character is a matter of what belongs to phenomenal consciousness but indistinguishability is a matter of what a subject can distinguish under reflection on phenomenally consciousness contents. So the concepts of phenomenal identity and phenomenal indistinguishability seem to be distinct concepts. The distinction is also motivated by empirical cases such as change blindness that seem paradoxical under a view that would equate phenomenal identity with phenomenal indistinguishability. For, these cases seem to imply that phenomenal indistinguishability is not a transitive relation. Identity, however is a transitive relation. So phenomenal identity and phenomenal indistinguishability cannot be co-extensive.

¹¹Zeimbekis (2013) employs the same strategy in support of a post-perceptual interpretation of the results of color-matching experiments.

¹²Firestone and Scholl (2015a) challenge the Levin and Banaji (2006) experiments by showing that the effects disappear when we blur out race. They therefore conclude that the effects are pre-perceptual. Our goal here has been to show that the post-perceptual explanation would still be available even if Firestone and Scholl's critique of these findings fails.

¹³For a review of some of these studies see Pylyshyn (1999).

stage during which a response unit responds to the detection stage?

From the standpoint of SDT, informational connections are always noisy. Because of this noise, sometimes the detection unit is not “telling” the response unit that what it sees is a dog, but it might “sound” to the response unit that the signal is “dog.” The basic way that the response unit solves this problem is to adjust a response threshold (or a response bias). When the input exceeds the threshold the response unit will treat it as a signal, and if it is below the threshold it will be treated as noise. The central assumption behind the SDT approach is that when the input to a response unit remains constant, adjusting response thresholds is the only way to affect the behavior of this unit. So, if cognition is making you more prone to classify a picture as a dog by affecting the post-perceptual response stage, it must be doing so by lowering the response threshold to dogs.

This assumption takes us a long way. Whether an effect is a threshold effect in this sense or not can be empirically determined. This is due to an interesting feature of threshold effects. Increasing the threshold reduces the number of cases where noise is falsely treated as signal (fewer non-dogs classified as dogs). However, this has the cost of increasing the number of false negatives, namely, cases where a signal is falsely treated as noise (more dogs classified as non-dogs). Decreasing the threshold, in contrast, decreases false negatives (signal treated as noise) with the cost of an increase in false positives (noise treated as signal). Changes in threshold are therefore essentially accompanied by a coupling between false negatives and false positives. Non-threshold mechanisms, in contrast, need not give rise to a coupling pattern. So one way to find out whether an increase in correct responses to a stimulus type is the result of the adjustment of a threshold is to figure out whether there is a coupling between false negatives and false positives. And in principle, one can detect whether there is such coupling if one has a large data set of responses that can be statistically analyzed¹⁴. The upshot is that threshold effects can be empirically distinguished from non-threshold effects.

How can this help us distinguish perceptual effects from post-perceptual effects? If post-perceptual effects are threshold effects we can distinguish them from non-threshold effects. But, as was quickly noted, some perceptual effects are also threshold effects. So, finding out that an effect is a threshold effect will not tell us that it is post-perceptual. However, those who think that SDT can help us solve the problem assume that post-perceptual effects are essentially threshold effects. So finding out that an effect is not a non-threshold effect is evidence that it is not post-perceptual¹⁵.

We can now see the conceptual problem with applying SDT to the task of distinguishing perceptual from post-perceptual effects. The assumption behind this approach is that post-perceptual effects are essentially threshold effects. But it is not clear why we

should accept this assumption. There is no conceptual connection between being post-perceptual and being a threshold effect. Of course, some examples of post-perceptual effects, e.g., effects of bias, are plausibly threshold effects. But there is no reason to assume that what is true of these cases generalizes to all post-perceptual cases. Moreover, there is no reason to assume that what is true of the perceptual system does not generalize to the response system. Everyone agrees that the perceptual system can get better at figuring out what happens around us in a way that reduces false positives (or false negatives) without increasing false negatives (or false positives), but that need not involve threshold mechanisms. If this is true, then why should we assume that there could not be non-threshold improvements in how the response system figures out what the perceptual system is “telling” it? After all, there is uniformity at the neural level in that the same basic mechanisms in both the perceptual system and the response system govern the propagation of neural activity. Why should things be different at the psychological level of description?

We thus conclude that it is far more difficult to rule out post-perceptual explanations with psychological methods than opponents of encapsulation usually think. It would be wrong to think that post-perceptual explanations of “online” experiments are implausible. And there is no reason to assume that the distinction between perceptual and post-perceptual mechanisms maps to the distinction between threshold and non-threshold mechanisms. In so far as d' and other parameters are measures of the latter distinction, there is no reason to assume that they can be used to meet the post-perceptual challenge.

This is not to say that post-perceptual explanations could never be ruled out. We think that in non-borderline conditions and in the absence of confounding factors it should be uncontroversial that subjective reports about the perception of a stimulus or its detectability are good indicators of the existence of percepts. After all, it is mainly on the basis of subjective reports that everyone agrees that there is a switch between percepts during binocular rivalry. However, as we shall argue soon, there is an interesting interplay between the different challenges that makes it the case that results that are hard to explain post-perceptually are more susceptible to the pre-perceptual or intra-modular challenges¹⁶.

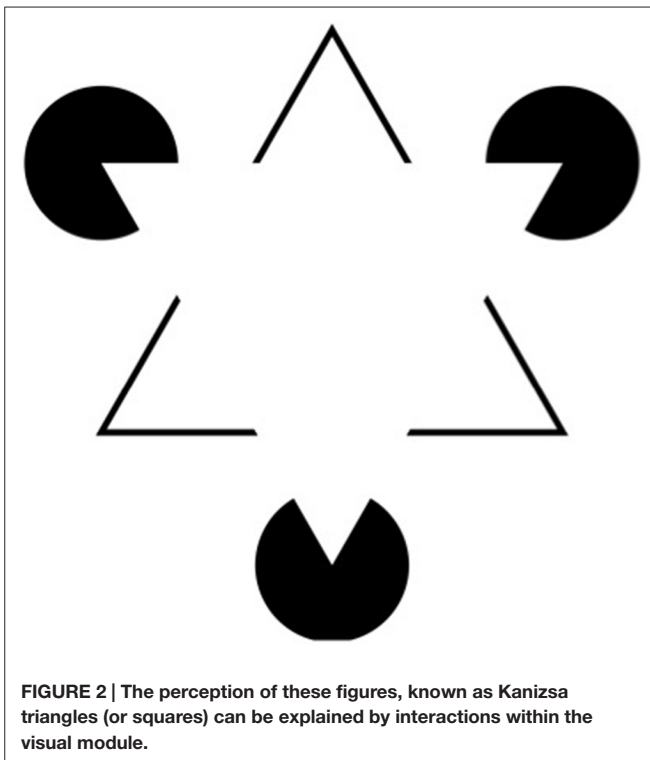
The Intra-Modular Challenge

As we saw in the previous section, one requirement a study must meet in order to provide evidence against encapsulation is to show that the locus of an effect is not post-perceptual. But this is not sufficient to refute the encapsulation thesis because it is possible that the origin of the effect is *intra-modular*. If both the locus and the origin of an effect are within the visual module, then the effect cannot count against encapsulation. In fact, putting both the locus and the origin of an effect inside the module is a common mode of explanation of what are called contextual effects. Contextual effects occur when the perception of individual elements within a visual scene are influenced by

¹⁴The two crucial parameters in SDT that were assumed to lend themselves to this task are d' (or the sensitivity parameter) and β (or the response bias parameter). The former is the distance between the means of noise and signal-plus-noise distributions. Changes in d' have been taken to indicate changes in the percept and changes in β as indicators of response bias.

¹⁵This is why changes in d' are regarded as indicative of changes in the percept.

¹⁶Wilbertz et al. (2014) and Marx and Einhauser (2015) are example of studies that demonstrate effects that cannot be easily ruled out as post-perceptual. Nevertheless, we think these studies do not sufficiently meet the pre-perceptual challenge. Thanks for an anonymous reviewer for bringing these studies to our attention.

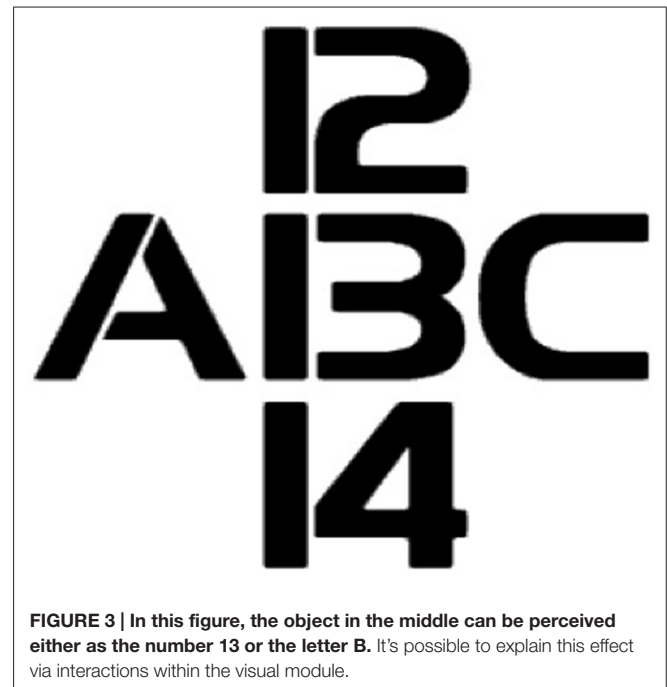


other elements within the scene. A famous example of this is the phenomenon of amodal completion where we perceive elements in the visual scene for which no direct information in the proximal stimulus is present (see **Figure 2** for an illustration of this phenomena).

The common explanation for contextual effects is that although stimuli in different areas of the visual field are processed by different and partially independent units in the visual module, these units can sometimes interact with each other through intra-modular connections. These connections embody knowledge. But this knowledge is embedded within the module, and the fact that it can influence the output of the module is compatible with the encapsulation thesis.

This observation generalizes beyond contextual effects. Some alleged effects on perception can be explained as intra-modular effects. Therefore, a second challenge for an empirical study that aims to provide evidence against the encapsulation thesis is to rule out intra-modular interpretations of the findings. We call this the *intra-modular challenge*.

Whether a study can meet this challenge partly depends on how we draw the boundaries of the visual module. Consider the experiment at the beginning of the previous section where the way that subjects categorized an ambiguous image was influenced by the race of the individual holding the object. It might seem natural to assume that categorizing an object as a gun or a tool, or categorizing individuals as belonging to different races, is a post-modular matter. But it has not been definitively established that the visual module cannot represent object categories or race categories. Consider **Figure 3** in which we can perceive the shape in the middle either as number 13 or as letter B. In our view, it is not implausible at all that the difference between perceiving the



letter as a B or as number 13 is a perceptual matter. If so, then the visual system might be able to represent categories including object categories and racial categories¹⁷. And if this is the case, effects of racial categories on how objects are categorized can be potentially explained away as intra-modular effects.

The intra-modular challenge becomes more serious if we allow, as modularists like Pylyshyn do, that the boundaries of modules are flexible and can change as a result of perceptual learning. For example, on a view like this acquiring expertise in reading written text partly consists in the automatization and encapsulation of the processes that give rise to representations of the semantic properties of a word. These processes thus become part of the visual module. So acquiring fluency in reading written text partly consists in the fact that the visual module now comes to represent semantic categories and the association between these categories and orthographic markers¹⁸. As such, an alleged effect of word meaning on visual experience of words can result from an intra-modular effect.

There is no reason to think that such an account could not be generalized beyond semantic categories. In principle, as a result of learning, many complex properties and their association with simpler visual markers such as colors and shapes can come to be represented by the visual module. If so, the effects of the representation of these properties on vision can potentially be explained away as intra-modular.

To see the significance of treating perceptual learning as a form “modularization,” consider the alleged effect of race indicative facial features on color perception (Levin and Banaji, 2006).

¹⁷For defenses of the view that perceptual experience can have a very rich representational content see Siegel (2010) and Masrour (2011).

¹⁸Of course, in such a case the label “visual module” may not be the best label anymore.

Participants in the debate normally assume that representation of race is a post-perceptual matter. If so, the origin of such effects would be outside perception, and as a result, the discussion of these effects has mostly focused on whether they can be ruled out post-perceptually. But if we allow that the boundaries of the visual module could expand with learning, the intra-modular explanation would also become an option. Perceptual learning might result in the modularization of both the representations of facial features and their association with a specific color. This would also explain why such effects are resistant to explicit beliefs to the contrary and are usually classified as effects of implicit beliefs. One could therefore explain away the effect of facial categorization on color perception as an intra-modular effect.

The idea that perceptual learning can result in the representation of new complex properties might seem incompatible with modularism. It seems plausible to assume that new complex properties come to be represented by the visual module in so far as at some point during the learning process there have been influences from outside the visual module which directed the learning process. However, one can distinguish between two types of learning, namely, Additive Learning and Revisionary Learning. Revisionary learning changes the parameters of existing processing capacities within the module. Revisionary learning, when it happens as a result of access to information outside the module is incompatible with our definition of encapsulation. Additive Learning, in contrast, occurs as a result of the addition of new processing capacities to the visual module, e.g., new capacities which often allow for the representation of new properties. This type of learning does not conflict with the encapsulation thesis as we define it¹⁹.

These observations should show that meeting the intra-modular challenge is not as easy as it initially might seem. Now let us consider a third challenge for anti-modularists.

The Pre-Modular Challenge

We have so far argued that anti-encapsulationist studies have to face the challenge of ruling out post-perceptual and intra-modular explanations of their findings. However, ruling out these explanations is not sufficient for refuting the encapsulation thesis. Suppose, for example, that it has been empirically demonstrated that expert bird-watchers are faster and more accurate in visually recognizing birds than non-experts. Let us also suppose that we have successfully ruled out post-perceptual and intra-modular explanations for this finding. Still, there is the possibility that the main cause of this difference in performance lies in the fact that experts employ more efficient visual search strategies. In short, experts know where to look. As a result, when an expert and a non-expert look at the same bird, the input that the visual module of the expert typically receives is different from the input that the visual module of the non-expert receives. It is therefore possible to explain the effect of expertise in terms of pre-perceptual differences in input. This illustrates the *pre-modular challenge*.

The pre-modular challenge is not confined to effects of expertise, and can in principle be employed to explain away

many allegedly cognitive effects on perception. For example, some priming effects on perception can be explained as pre-modular effects. Thus a third challenge for an anti-encapsulationist empirical study is to rule out pre-modular interpretations of the findings. We call this the pre-modular challenge.

Note that the breadth of the pre-modular challenge partly depends on whether we accept the primary or the middle version of the encapsulation thesis. As we noted in the “Encapsulation Thesis” section, according to Middle Encapsulation the visual module starts somewhere in the middle of the visual hierarchy. Specifically, it does not start where the retina starts. On such a view, some attentional shifts could change the inputs that a module receives by modulating the activity of the transport or relay units prior to the visual module (see O'Connor et al., 2002; Cudeiro and Sillito, 2006; McAlonan et al., 2006). Such effects would be thus compatible with middle encapsulation.

To meet the pre-modular challenge posed by Middle Encapsulation, one needs to rule out that the observed effects are effects of pre-modular attention. Primary Encapsulation, in contrast, is incompatible with attentional effects on relay centers. In order to meet the pre-modular challenge posed by Primary Encapsulation, one only needs to rule out that the observed effects result from visual search strategies (for example, changes in direction of gaze and saccadic movements). The pre-modular challenge is therefore harder to meet if our aim is provide evidence that Middle Encapsulation fails, as opposed to the Primary Encapsulation. This is so because, in addition to visual search strategies, attention may affect the relay and transport centers between the retina and the lower boundaries of the visual module.

The effects that result from where a subject looks could in principle be ruled out by controlling for factors such as eye movements (whether saccadic or deliberate), but there are other types of attention that one must rule out. Although the most frequently cited such attentional effect result from spatial attention, it is not uncommon for modularists to also appeal to feature-attention. Pre-modular feature attention occurs when a subject's attention to a specific feature changes the activity of units that relay activity pertaining to that feature.

How could we rule out attentional effects, of any type, by purely psychophysical experiments? One thought here might be that attentional effects are weak and are confined to spatial properties and simple features. So one way to meet the attentional challenge is to find robust and complex effects on perception.

However, it is not clear that the effects of attention are always weak and simple. It might be true that pre-modular attentional shifts can only cause minor changes in the input that a module receives. However, minor changes in input can result in Gestalt-like switches in the way that the input is processed. Consider, for example, **Figure 3** again, in which the ambiguous figure can be interpreted both as number 13 or as the letter B. As we noted, it is possible to interpret this effect as post-perceptual or intra-modular. But a third option is that the difference between the two cases emerges from differences in pre-modular attention. For example, perception of the shape as a 13 could result from paying more attention to the gap between the curved and the horizontal lines, and perceiving it as a B could result from moving attention

¹⁹Thanks to an anonymous reviewer whose comment helped us clarify this point.

away from the gap. Our visual module is therefore receiving two different patterns of input in the two cases. And although this difference might be minor, it may be sufficient to cause a Gestalt shift in the way that the input is processed by the visual module. Accordingly, a minor attentional change might result in a robust difference in whether the perceptual system classifies the input as a B or a 13²⁰.

It is indeed hard to see how one might be able to meet the pre-modular challenge by purely psychophysical methods. One might for example try to design experiments in which the attentional difference between experimental and control groups or within subjects during different experiment conditions are minimized. However, we are aware of no such studies. We therefore think that meeting the pre-modular challenge is especially difficult if our goal is to refute the middle encapsulation thesis.

The Interplay Between the Challenges

As we noted earlier, there is also an interesting interplay between the above three challenges in that attempting to meet one of them often makes meeting the others more difficult. One might, for example, argue that the early occurrence of an effect is a reason to rule out that it is a post-perceptual effect. However, this would obviously make ruling out intra-modular or pre-modular explanations more difficult.

Another example involves appealing to cognitive manipulability in order to rule out post-perceptual explanations. Consider the implicit bias studies that show that subjects who are otherwise completely unaware of having racial biases are more prone to report a blurry image as a gun when it's held by an African American individual (Payne, 2001). Discussion of this effect has often focused on explaining it away as a post-perceptual²¹. But now suppose that in an attempt to rule out the post-perceptual explanation, we show that the effect cannot be manipulated by, say, informing the subjects of their bias. In other words, the effect turns out to be resistant to cognitive manipulation. This would give us some reason to think that the effect is not post-perceptual. The problem is that this would also increase the likelihood that the effect is intra-modular. In general, ruling out the post-perceptual explanation of an effect by showing that it is resistant to cognitive manipulation increases the likelihood that the effect is intra-modular. So, here too our attempt to meet the post-perceptual challenge makes it less likely that we can meet the intra-modular challenge.

We have considered three challenges that psychophysical studies must meet in order to provide evidence against the encapsulation thesis. We have argued that meeting these challenges individually, and in conjunction with each other, is much more difficult than what has been often assumed. We conclude that it is unlikely that one would be able to rule out the encapsulation thesis with purely psychophysical studies.

In a forthcoming BBS target article on this issue, Firestone and Scholl take a more radical line, arguing that almost all

psychophysical findings in support of top-down influence has been debunked. Since most psychological models of perceptual processing are purely bottom up, they conclude that empirical evidence favors the encapsulation thesis.

We do not think that the claim that all psychological evidence against encapsulation has been debunked is correct. In our view, the evidence is inconclusive. But one who accepts this might still think that rather than concluding that psychophysics cannot settle the debate, the correct conclusion should be similar to Firestone and Scholl's line that is, we should embrace the encapsulation thesis.

This line of thought assumes that the encapsulation thesis is the default and has to be upheld unless there is psychophysical evidence against it. But we do not see any empirical reasons to accept the encapsulation thesis as default. It is true that many working models of psychological processes are bottom up. But that is mainly because many modelers assume the bottom-up model as an *a priori* meta-constraint on psychological theorizing. We think that given the newly emerging predictive coding models of perceptual phenomena, this pattern will gradually change (see Clark, 2013; Hohwy, 2013, for references).

Some might also think that the upshot of this conclusion should be skepticism about the empirical resolution of the debate between modularists and anti-modularists. However, we think that the proper reaction is to combine psychophysical studies with other empirical sources of evidence. In the next section we focus on one of these sources, namely, the evidence from neuroscience. As we shall argue, considering this sort of evidence, tilts the balance of the empirical evidence in favor of the anti-modularist.

THE NEURAL CASE AGAINST ENCAPSULATION

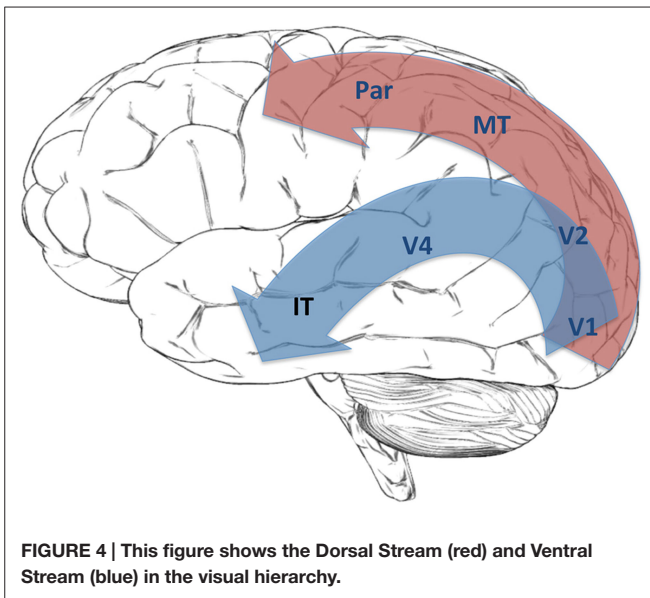
As we saw in the previous sections, the psychological evidence against the encapsulation thesis is at best inconclusive. In the following sections, we examine the plausibility of the encapsulation thesis from the perspective of neuroscience. This goes against the common approach in the recent literature that does not engage with this body of evidence. Our focus will be on the access version of the encapsulation thesis. As a reminder, this is the thesis that there is a component of the visual system that gives rise to access conscious representations and is informationally encapsulated. As before, we will simply refer to this thesis as the encapsulation thesis. We argue that recent findings about the connectivity structure and activity dynamics of the visual system militate against this thesis.

Our guiding question is whether a neural correlate of an encapsulated module could be identified in the human visual system. We consider two strategies for demarcating this alleged neural correlate. After a quick introduction to the structure of the visual system, we consider identifying the perceptual module with a neuroanatomically demarcated area of the visual system²². We argue that although an area of the cortex that would correspond to the functional profile of a module could

²⁰For evidence that deployment of spatial attention could influence subsequent perceptual categorizations see Kietzmann et al. (2011). Thanks to an anonymous reviewer for bringing this study to our attention.

²¹Raftopoulos (2009) is an example of this strategy.

²²Some of the main proponents of the encapsulation thesis associate perceptual modules with a neuroanatomically demarcatable area of the brain. This is explicit in Fodor (1984) and Pylyshyn (1999) also seems to tacitly



be roughly demarcated, empirical results show that this area is not functionally encapsulated. We then consider a strategy that identifies the neural correlate of the visual module with a temporally identified process that happens in a roughly demarcated neuroanatomical region of the cortex. We argue that although an encapsulated visual process could be identified, empirical results suggest that this process fails to give rise to access conscious representations. Lastly, we anticipate a few replies to our argument and respond to them.

The Neuroanatomical Strategy

Neuroscientific orthodoxy regards the visual system as a hierarchical structure. Activity starts at the retinal receptors and passes through the retinal ganglion cells to the lateral geniculate nucleus (LGN) in the thalamus. This is an area within the brain stem that is often regarded as a relay center for sending information to cortical areas. In the case of vision, projections from LGN connect it to the primary visual cortex (V1), after which the visual pathway divides into ventral and dorsal streams. The ventral stream includes areas V2, V4 and the temporal lobe. The dorsal stream includes areas V3, MT and the parietal cortex²³. The visual system thus manifests a fork-shaped hierarchical structure (Figure 4).

endorse it. Nevertheless, these theorists have not explicitly proposed any specific neuroanatomical demarcation of the boundaries of the module.

²³This, of course, under describes the complexity of the neural structures that underlie vision. For example, area MT receives connections from the retinal ganglion cells through the pulvinar structure in the thalamus and the superior colliculus (the tectum). These connections entirely bypass the LGN and the V1–V4 areas. So the place of area MT in the hierarchy is not completely clear. More importantly, the visual pathways also host an abundance of feedback connections that relay activity from higher areas of cortex to areas even as low as the retinal ganglion cells. So the idea that there is a simple neuroanatomical hierarchy in the visual system is somewhat questionable. Nevertheless, these complications can be accommodated within a general hierarchical framework that allows for multiple hierarchical schemes.

This neuroanatomical hierarchy also corresponds to physiologically and functionally specified hierarchies. Physiological studies show that neurons in the visual cortex could be ordered with respect to the size of their receptive fields, that is, the area of the retina that a neuron responds to. Interestingly, the ordering on the basis of receptive field sizes roughly corresponds to the position of a neuron in the neuroanatomical hierarchy; the higher the neuron in the hierarchy, the larger its receptive field. According to orthodoxy, the neuroanatomical hierarchy also roughly corresponds to a functional hierarchy. Different neurons respond to the presence of different types of stimuli in their receptive fields. For example, some neurons respond to the presence of edges, some respond to motion, some respond to colors, and some respond to complex shapes. This is often called the tuning function of a neuron. It is commonly held that tuning functions can also be ordered with respect to their complexity. For example, detecting a shape is more complex than detecting an edge. This hierarchy of functional complexity also roughly corresponds to the neuroanatomical hierarchy: neurons higher on the neuroanatomical hierarchy have more complex tuning functions.

If the boundaries of the visual module are neuroanatomical they should naturally fall somewhere within the visual hierarchy. The question is where. We shall start with the minimal working hypothesis that the visual module starts in area V1 and extends to V4 in the ventral stream. We shall call this area the lower visual system (LVS).

A few points about identifying the alleged visual module with LVS are in order. First, LVS does not include areas earlier than V1 in the visual hierarchy such as retinal receptors, ganglion cells, and LGN. The encapsulation of LVS would, therefore, correspond to the Middle Encapsulation thesis as characterized in the “Encapsulation Thesis” section. Second, we have not included the dorsal stream in LVS. The initial rationale for this is that our focus here is on Access Encapsulation and it is common to assume that the dorsal stream does not give rise to access conscious representations²⁴. Third, we have not included areas higher than V4 in LVS. The main rationale for these limitations is to simplify the structure of the discussion. After considering whether LVS is encapsulated, and arguing that it is not, we consider modifying the minimal working hypothesis by adding areas earlier than V1, the dorsal stream and areas higher than V4. The basic question to consider at this stage is whether LVS is encapsulated. We think it is not. What follows is a review of some of the main findings that support this claim.

Recent research shows that the receptive field sizes of neurons, even in the V1 area, change over time (Gilbert and Wiesel, 1989; Hupe et al., 2001; Li and Gilbert, 2002; Stettler et al., 2002; Bair et al., 2003; Angelucci and Bressloff, 2006; Gilbert and Li, 2012; for a review, see Gilbert and Li, 2013). Whereas a neuron’s early response to stimuli (<100 ms) reflects the presence of a stimulus in its classical receptive field, a neuron’s later activity (after 100 ms) is sensitive to the presence of flanking stimuli outside its receptive field. These effects are often referred to as contextual effects. The

²⁴The distinction between the two visual streams is, of course, controversial (see Schenk and McIntosh, 2009, for a review).

existence of contextual effects is relevant to encapsulation because V1 neurons do not receive direct input from the areas of retina that fall outside their classical receptive field. Therefore, if a neuron is sensitive to the presence of stimuli outside its classical receptive field, it must receive input from areas other than areas earlier in the visual hierarchy. And if these areas are outside LVS, then LVS is not encapsulated. However, modularists often hold that these contextual effects can be explained in terms of communication between neurons at the same level of neuroanatomical hierarchy (horizontal connections) or recurrent feedback from neurons higher in the visual hierarchy but still within the boundaries of the visual module. So it might be possible to explain away contextual effects in terms of connections within LVS.

However, the exact circuitry underlying contextual effects is still a matter of controversy. There are at least two camps. The first camp holds that horizontal connections are the primary carriers of contextual effects. The second camp holds that the primary carriers of contextual effects are recurrent feedback connections from areas higher than V1, including MT, which is outside LVS²⁵. So, whether contextual effects present a threat to the encapsulation of LVS is still a live issue²⁶.

A second set of findings that poses a more serious threat for the encapsulation of LVS comes from studies of the circuitry underlying attentional effects on the visual system. These effects are often classified into spatial, feature-based, and object-oriented attentional effects. There are interesting conceptual questions surrounding these distinctions, but for our purpose what matters is the following:

- (a) It has been shown that spatial attention modulates neuronal responses in V1, V2, and V4 (Mountcastle et al., 1987; Motter, 1993; Gandhi et al., 1999; Ito and Gilbert, 1999; McAdams and Maunsell, 1999; Reynolds et al., 1999; Crist et al., 2001).
- (b) Feature-based attention and task-related effects that are sometimes regarded as attentional effects modulate the activity of V4 neurons (Motter, 1994; Chelazzi et al., 2001; Reynolds and Desimone, 2003; Li et al., 2004; Zhou and Desimone, 2011; Gilbert and Li, 2013).
- (c) Object-based attention can modulate activity in areas as low as V1 (Jolicoeur et al., 1986; Roelfsema et al., 1998; Scholte et al., 1999; Lamme and Roelfsema, 2000; Roelfsema et al., 2000).
- (d) TMS studies have shown task-specific modulatory effects of expectations in V4 neurons (Morishima et al., 2009).

These findings show that the activity of neurons in LVS modulate with tasks, expectations and attention. The attentional effects are endogenous that is, attentional effects that are not induced by stimuli (exogenous attention). So the origin of these

effects lies in areas higher than LVS. Moreover, these effects are content-sensitive. It is as though the neurons inside LVS know what task a subject is performing or which aspect of the stimulus the subject is attending to. This implies that LVS is not informationally encapsulated.

We have described three sets of findings that challenge the encapsulation of LVS. To summarize, (a) there are well-established contextual effects on LVS and it is still a matter of controversy whether these effects could all be explained in terms of connections between neuronal assemblies within LVS, (b) there are well-established effects of spatial attention, feature attention, and object attention effects on LVS that originate from areas outside LVS, and (c) there are well-established task related and expectation related effects on LVS. We therefore think that LVS is not informationally encapsulated and this puts pressure on the encapsulation thesis.

It might be argued that attentional effects are not incompatible with encapsulation. Later in the paper we will argue that this response fails, but we shall discuss a more pressing question first. Could the challenge for the encapsulation thesis be simply removed by identifying the visual module with a neuroanatomical area that is different from LVS? We think the answer is negative. Let us elaborate.

One could modify the thesis that the visual module is identical with LVS by either adding areas to it, subtracting areas from it, or by a combination of these two strategies. We do not think that any of these modifications would help the modularist. For example, consider extending the alleged visual module by adding area MT to LVS. This might seem to help the modularist because under this modification the feedback connections from MT to lower areas like V4 and V2 would now count as intra-modular effects. But this move has an important cost. Now that area MT is part of the visual module, those feedback connections from higher areas that modulate the activities of MT neurons would threaten the encapsulation thesis. There is ample evidence that there are such feedback connections to MT (Treue and Maunsell, 1996; Treue and Trujillo, 1999; Ninomiya et al., 2012). Now consider the reverse strategy of shrinking the alleged module by, say, subtracting area V4 from the LVS. The benefit of this would be that modulating feedback connections to V4 would now count as post-modular. But the cost is that the well-established effects of V4 on lower areas that were originally classified as intra-modular effects would now be incompatible with the encapsulation thesis.

This problem seems to generalize to any proposal for expanding or shrinking the alleged module by adding an area to, or removing an area from, the upper boundary of LVS. Expanding the boundaries of LVS to include areas higher in the visual hierarchy might accommodate some of the aforementioned effects as intra-modular. But this risks threatening the encapsulation thesis because the higher an area in the visual hierarchy the more it is likely that it receives input from areas further up. This is due to the fact that the hierarchical organization of the cortex gradually fades away as we move up the visual hierarchy and gives way to a non-directional connectivity pattern. Subtracting areas, on the other hand, suffers from the same problem that we described above. It also risks conflicting with

²⁵For discussions of the debate concerning the circuitry of contextual effects (see Allman et al., 1985; Knierim and Van Essen, 1992; Lamme, 1995; Zipser et al., 1996; Nothdurft et al., 2000; Hupe et al., 2001; Jones et al., 2001; Angelucci et al., 2002; Cavanaugh et al., 2002; Levitt and Lund, 2002; Bair et al., 2003; Angelucci and Bressloff, 2006).

²⁶We will consider modifications of our working hypothesis that include adding MT to the module at the end of this section.

the requirement that the outputs of a module should be access conscious.

For similar reasons, it is hard to see how the other ways of expanding or shrinking the visual module, such as adding the dorsal stream, adding areas prior to V1 or subtracting areas from the lower boundary of the module, would help the modularist. We therefore conclude that there is no clear neuroanatomical strategy for demarcating an area in the visual system that is encapsulated and whose outputs are access conscious representations. There is no neuroanatomically identifiable visual module.

The Temporal Strategy

Neuroanatomical strategies do not exhaust the options for the modularists. One alternative is to partially characterize the neural correlates of the visual module in a temporal fashion. The core insight behind this strategy emerges from a deep and interesting debate over the proper way to establish a mapping between the functional and structural description of the brain. This debate is independent from the debate over modularity, but it would help to say a few words about it first. For a long time, a very influential line of thought among neuroscientists has been that there is a one to one mapping between the fine grained structure of the cortex and its functional description, especially in areas corresponding to the perceptual system. Accordingly, one could say that some V1 neurons have the single function of responding to changes in orientation in a specific area of the visual field. This is what we earlier referred to as the tuning function of a neuron or a neuronal assembly. This idea has been lately challenged by neuroscientists who argue that neurons can perform different functions at different times (see Lamme and Roelfsema, 2000, for a review). A neuron's tuning function and receptive field sizes can both change as a result of receiving input through recurrent feedback connections. For example, a V1 neuron that responds to orientation in a small area of the visual field in the first 100 ms following the presentation of the stimulus, shows sensitivity to more global and complex features after 100 ms.

Lamme and Roelfsema (2000) expand on this idea by distinguishing between two phases of activity in the early parts of the visual cortex. The first phase, the feedforward sweep, happens in the 40–100 ms window after stimulus onset. The ensemble of neurons that participates in feed forward sweep and their activation pattern is primarily determined by feed forward connections. This is simply because there has not been enough time for recurrent connections to exert their influence on these neuronal assemblies. After 100 ms, horizontal and feedback connections start modulating the activity of the neurons that participated in feed forward sweep and change their tuning functions and receptive fields.

This view, if correct, would have deep and important implications for a wide variety of issues, including the nature of attention and the neural correlates of consciousness. But for our purposes here, the point is that the view provides an attractive alternative for the modularist because it seems capable of dealing with the complications that we raised in the “Neuroanatomical Strategy” section. On this alternative, rather than identifying the neural correlates of the visual module with a specific area of

the visual cortex, we identify it with a process that takes place in a neuroanatomical area during a specific time interval. For example, one option is to identify the visual module with the feedforward sweep that takes place in LVS. This strategy seems initially promising because it guarantees the encapsulation of the visual module. Since during the feedforward sweep the activity of LVS neurons is solely determined by feedforward connections, and the neural correlate of the visual module is identified with the feedforward sweep that happens in LVS, then the visual module is encapsulated.

We can now see why some modularists, such as Raftopoulos (2009, 2014), have found the temporal strategy attractive. Raftopoulos does not identify the visual module with feedforward sweep. Rather, drawing on Lamme (2003), he divides the wave of activity after the feedforward sweep into two phases. The first phase is a local recurrent phase that culminates at 120–150 ms after stimulus onset. The second phase is a global recurrent phase that starts around 150–200 ms after stimulus onset and allows for feedback connections from higher cognitive areas. On Raftopoulos' view, the visual module (what he calls early vision) should be identified with the processes that happen in the lower areas of the visual hierarchy during the first 150 ms and include the feedforward sweep and combine the feedforward sweep with the local recurrent phase.

Despite its initial attraction, however, the temporal strategy could not help the modularist save the encapsulation thesis. Recall that our target is Access Encapsulation according to which there is an informationally encapsulated component of the visual system that gives rise to access conscious representations. But there is ample evidence that the activity in the first 150 ms after stimulus onset is not sufficient for access consciousness (Bridgeman, 1975, 1988; Kovács et al., 1995; Vogel et al., 1998; Rolls et al., 1999; Dehaene et al., 2001; Lamme et al., 2002; Sergent and Dehaene, 2004; Koivisto et al., 2006, 2009; van Aalderen-Smeets et al., 2006; Del Cul et al., 2007; Fahrenfort et al., 2007; Melloni et al., 2007; Lamy et al., 2009; Railo and Koivisto, 2009). On the dominant view, access consciousness requires the availability of representations in a global workspace which does not happen before 300 ms after stimulus onset (see Dehaene and Changeux, 2011, for a review of the relevant literature). What is controversial is whether the earlier phase of activity is sufficient for phenomenal consciousness, which is independent from issues regarding access consciousness. The temporal strategy thus fails to save Access Encapsulation for the simple reason that the activity during the early phase after stimulus onset is not sufficient for access consciousness.

Do Attentional Effects Entail Failure of Encapsulation?

We have so far argued that given the status of recent neuroscientific findings, it is very unlikely that a neural correlate for a visual module that is informationally encapsulated, and gives rise to access conscious representations, could be identified. Our strategy has, in effect, presented the modularist with a dilemma. The first horn is to identify the visual module with a neuro-anatomically demarcated area of the visual cortex. If the modularist chooses this option, she has to face the challenge

of accounting for the existence of feedback connections that modulate the activity of the neurons in early visual areas. The second horn is to identify the visual module with processes that are partially characterized temporally. But choosing this option conflicts with the requirement that the visual module should give rise to access conscious representations. Before ending the paper we want to return to a question that we brought up earlier in the “Neuroanatomical Strategy” section, namely, whether attentional effects are compatible with encapsulation. We consider three reasons for thinking that they are and argue that none of them withstand scrutiny.

We noted, in the “Pre-Modular Challenge” section, that proponents of modularity commonly hold that some attentional effects on perception are mediated by changes in the input to the module from earlier areas and are therefore compatible with the encapsulation thesis. We agreed with the modularists that pre-modular attentional effects do not count as failures of encapsulation. But it might be argued that a similar strategy could be employed against our arguments here. Accordingly, one might argue that the attention-induced modulations of the activities of the LVS neurons are mediated by effects on the units earlier than LVS. Attention-induced modulations of LVS neurons would therefore be mediated by modulations of the inputs to LVS and can thus be explained away as pre-modular.

Three problems threaten this response. First, the aforementioned studies do not mention modulations of pre-modular centers that accompany the attentional effects on LVS. Second, there is reason to think that some of the attention induced modulations in areas such as V4 and V2 could not have mediated by effects on units prior to LVS. This is because some earlier studies of attention induced modulations of areas V4, V2, and V1 could not find any modulation of V1 neurons. In fact, it was only very recently that modulations of V1 neurons have been detected (Luck et al., 1997; Gandhi et al., 1999; Ito and Gilbert, 1999; Maunsell and Cook, 2002)²⁷. Any impact of the units prior to V1 on LVS has to be mediated by V1. Therefore the absence of V1 modulation implies that the effects on higher areas, such as V4 and V2, could not have originated from direct effects on earlier units.

The third problem is that it is not clear how the pre-modular strategy could be applied to modulations induced by cases of feature and object attention. Attention induced boosting of the signal corresponding to a specific feature or object requires boosting the activity of a neuronal assembly that represents that feature or object. But it is not clear how pre-modular transducer and relay units could represent objects, or any features except for those for which there are transducers. In fact, within a broadly Fodorian framework that most modularists endorse, representations of objects and representations of most features are post-computational. Therefore, pre-modular units, which are pre-computational, could not represent features or objects. It is therefore unclear how attention induced boosting of the activities of pre-modular units could

account for the effects of object and feature attention on LVS neurons.

Attentional effects have sometimes been regarded as compatible with encapsulation because they constitute a state of “readiness.”²⁸ However, it is not clear why this should render these effects compatible with encapsulation. It is true that some attentional effects on visual areas happen before the area receives stimulus-induced activities. Such effects would be in some sense pre-perceptual, so “readiness” might be an apt label here. However, it is not clear why the fact that an effect happens prior to stimulus-induced activity makes it compatible with encapsulation. If the higher areas of the cortex could “tell” a V1 neuron that what it is about to “witness” on the left side of the visual field is important and thereby affect the way it processes the input, then the V1 neuron has access to the information in the higher areas. It does not matter whether this access happens prior or posterior to stimulus-induced activity.

A second thing that the “readiness” label might mean is that these effects are not content-sensitive. Suppose there is a perception-booster potion that boosts the readiness and thereby the post-stimulus response of all the V1 neurons to all different types of stimuli in their receptive fields. Then we agree that there is a sense in which this boosting does not qualify as a failure of encapsulation because explaining it does not require appealing to contents. This is because explaining the effect of this potion would not require attributing to the V1 neurons access to any sort of content, e.g., what is the task at hand, what is salient for the task, what should the perceiver attend to, etc. But all the attention induced effects that we have cited here, whether they qualify as cases of “readiness” or not, are content-sensitive effects.

These attentional effects cannot be explained away by saying that they are mediated by pre-modular changes in input. Nor does the fact that some of them happen before stimulus-induced activity render them compatible with encapsulation, since they are content-sensitive effects. We conclude that these attentional effects are incompatible with encapsulation.

CONCLUSION

A core thesis in the debate between modularists and their opponents is the encapsulation thesis according to which there is a component of the visual system that is informationally encapsulated from impact from areas higher in the perceptual hierarchy. The secondary literature on the encapsulation thesis has mainly focused on the implications of purely psychophysical findings for this thesis. In this paper, we have pushed against this common tendency in two ways. We have argued that due to methodological limitations, purely psychophysical studies are incapable of resolving the debate between the modularists and their opponents. This gives us some reason to look for other sources of evidence. We have also taken the first steps in this direction by arguing that findings in the past few decades about the neural structure and connectivity pattern of the visual system undermines the encapsulation thesis. We hope that our arguments

²⁷Raftopoulos (2006) also points out that according to ERP findings, some modulations due to spatial attention only affect V4 and not the previous areas.

²⁸See Raftopoulos (2009, 2014). Raftopoulos sometimes refers to this as the “rigging up” of the activities of these neurons.

here help move the debate to the direction of taking the neural data more seriously.

Throughout the discussion, we have also distinguished between different versions of the encapsulation thesis and analyzed its

relation to the broader context of theoretical disputes surrounding the modularity debate. These distinctions helped structure and clarify the discussion that followed, and we hope that they will be of service to the continuing debate on the topic.

REFERENCES

- Allman, J., Miezin, F., and McGuinness, E. (1985). Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons. *Annu. Rev. Neurosci.* 8, 407–430. doi: 10.1146/annurev.ne.08.030185.002203
- Angelucci, A., and Bressloff, P. C. (2006). Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons. *Prog. Brain Res.* 154, 93–120. doi: 10.1016/S0079-6123(06)54005-1
- Angelucci, A., Levitt, J. B., Walton, E. J., Hupe, J. M., Bullier, J., and Lund, J. S. (2002). Circuits for local and global signal integration in primary visual cortex. *J. Neurosci.* 22, 8633–8646.
- Bair, W., Cavanaugh, J. R., and Movshon, J. A. (2003). Time course and time-distance relationships for surround suppression in macaque V1 neurons. *J. Neurosci.* 23, 7690–7701.
- Block, N. (1995). On a confusion about a function of consciousness. *Behav. Brain Sci.* 18, 227–247. doi: 10.1017/S0140525X00038188
- Block, N. (2005). Two neural correlates of consciousness. *Trends Cogn. Sci.* 9, 46–52. doi: 10.1016/j.tics.2004.12.006
- Bridgeman, B. (1975). Correlates of metacontrast in single cells of the cat visual system. *Vision Res.* 15, 91–99. doi: 10.1016/0042-6989(75)90065-6
- Bridgeman, B. (1988). Visual evoked potentials: concomitants of metacontrast in late components. *Attent. Percept. Psychophys.* 43, 401–403. doi: 10.3758/BF03208812
- Bruner, J. S. (1957). *Going Beyond the Information Given*. New York: Norton.
- Bruner, J. S. (1973). *The Relevance of Education*. New York: Norton.
- Cavanaugh, J. R., Bair, W., and Movshon, J. A. (2002). Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. *J. Neurophysiol.* 88, 2530–2546. doi: 10.1152/jn.00692.2001
- Chelazzi, L., Miller, E. K., Duncan, J., and Desimone, R. (2001). Responses of neurons in macaque area V4 during memory-guided visual search. *Cereb. Cortex* 11, 761–772. doi: 10.1093/cercor/11.8.761
- Churchland, P. M. (1988). Perceptual plasticity and theoretical neutrality: a reply to Jerry Fodor. *Philos. Sci.* 55, 167–187. doi: 10.1086/289425
- 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
- Crist, R. E., Li, W., and Gilbert, C. D. (2001). Learning to see: experience and attention in primary visual cortex. *Nat. Neurosci.* 4, 519–525. doi: 10.1038/87470
- Cudeiro, J., and Sillito, A. (2006). Looking back: corticothalamic feedback and early visual processing. *Trends Neurosci.* 29, 298–306. doi: 10.1016/j.tins.2006.05.002
- Dehaene, S., and Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron* 70, 200–227. doi: 10.1016/j.neuron.2011.03.018
- Dehaene, S., Naccache, L., Cohen, L., Le Bihan, D., Mangin, J. F., Poline, J. B., et al. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nat. Neurosci.* 4, 752–758. doi: 10.1038/89551
- Del Cul, A., Baillet, S., and Dehaene, S. (2007). Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biol.* 5:e260. doi: 10.1371/journal.pbio.0050260
- Durgin, F. H., Baird, J. A., Greenburg, M., Russell, R., Shaughnessy, K., and Waymouth, S. (2009). Who is being deceived? The experimental demands of wearing a backpack. *Psychon. Bull. Rev.* 16, 964–969. doi: 10.3758/PBR.16.5.964
- Durgin, F. H., Klein, B., Spiegel, A., Strawser, C. J., and Williams, M. (2012). The social psychology of perception experiments: hills, backpacks, glucose and the problem of generalizability. *J. Exp. Psychol. Hum. Percept. Perform.* 38, 1582–1595. doi: 10.1037/a0027805
- Fahrenfort, J., Scholte, H., and Lamme, V. (2007). Masking disrupts reentrant processing in human visual cortex. *J. Cogn. Neurosci.* 19, 1488–1497. doi: 10.1162/jocn.2007.19.9.1488
- Firestone, C., and Scholl, B. J. (2014). “Top-down” effects where none should be found: the El Greco fallacy in perception research. *Psychol. Sci.* 25, 38–46. doi: 10.1177/0956797613485092
- Firestone, C., and Scholl, B. J. (2015a). Can you experience top-down effects on perception? The case of race categories and perceived lightness. *Psychon. Bull. Rev.* 22, 694–700. doi: 10.3758/s13423-014-0711-5
- Firestone, C., and Scholl, B. J. (2015b). Enhanced visual awareness for morality and pajamas? Perception vs. memory in ‘top-down’ effects. *Cognition* 136, 409–416. doi: 10.1016/j.cognition.2014.10.014
- Fodor, J. (1984). Observation reconsidered. *Philos. Sci.* 52, 23–43. doi: 10.1086/289162
- Fodor, J. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press.
- Fodor, J., and Pylyshyn, Z. (1988). Connectionism and cognitive architecture: a critical analysis. *Cognition* 28, 3–71. doi: 10.1016/0010-0277(88)90031-5
- Gandhi, S. P., Heeger, D. J., and Boynton, G. M. (1999). Spatial attention affects brain activity in human primary visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 96, 3314–3319. doi: 10.1073/pnas.96.6.3314
- Gilbert, C. D., and Li, W. (2012). Adult visual cortical plasticity. *Neuron* 75, 250–264. doi: 10.1016/j.neuron.2012.06.030
- Gilbert, C. D., and Li, W. (2013). Top-down influences on visual processing. *Nat. Rev. Neurosci.* 14, 350–363. doi: 10.1038/nrn3476
- Gilbert, C. D., and Wiesel, T. N. (1989). Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *J. Neurosci.* 9, 2432–2442.
- Goldstone, R. L. (1995). Effects of categorization on color perception. *Psychol. Sci.* 6, 298–230. doi: 10.1111/j.1467-9280.1995.tb00514.x
- Hansen, T., Olkkonen, M., Walter, S., and Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nat. Neurosci.* 9, 1367–1368. doi: 10.1038/nn1794
- Hohwy, J. (2013). *The Predictive Mind*. Oxford: Oxford University Press.
- Hupe, J. M., James, A. C., Girard, P., Lomber, S. G., Payne, B. R., and Bullier, J. (2001). Feedback connections act on the early part of the responses in monkey visual cortex. *J. Neurophysiol.* 85, 134–145.
- Ito, M., and Gilbert, C. D. (1999). Attention modulates contextual influences in the primary visual cortex of alert monkeys. *Neuron* 22, 593–604. doi: 10.1016/S0896-6273(00)80713-8
- Jolicoeur, P., Ullman, S., and Mackay, M. (1986). Curve tracing: a possible basic operation in the perception of spatial relations. *Mem. Cogn.* 14, 129–140. doi: 10.3758/BF03198373
- Jones, H. E., Grieve, K. L., Wang, W., and Sillito, A. M. (2001). Surround suppression in primate V1. *J. Neurophysiol.* 86, 2011–2028.
- Kietzmann, T., Geuter, S., and König, P. (2011). Overt visual attention as a causal factor of perceptual awareness. *PLoS ONE* 6:e22614. doi: 10.1371/journal.pone.0022614
- Knierim, J. J., and Van Essen, D. C. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *J. Neurophysiol.* 67, 961–980.
- Koivisto, M., Kainulainen, P., and Revonsuo, A. (2009). The relationship between awareness and attention: evidence from ERP responses. *Neuropsychologia* 47, 2891–2899. doi: 10.1016/j.neuropsychologia.2009.06.016
- Koivisto, M., Revonsuo, A., and Lehtonen, M. (2006). Independence of visual awareness from the scope of attention: an electrophysiological study. *Cereb. Cortex* 16, 415–424. doi: 10.1093/cercor/bhi121
- Kovács, G., Vogels, R., and Orban, G. A. (1995). Cortical correlate of pattern backward masking. *Proc. Natl. Acad. Sci. U.S.A.* 92, 5587–5591. doi: 10.1073/pnas.92.12.5587
- Lamme, V. A. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *J. Neurosci.* 15, 1605–1615.
- Lamme, V. (2003). Why visual attention and awareness are different. *Trends Cogn. Sci.* 7, 12–18. doi: 10.1016/S1364-6613(02)00013-X
- Lamme, V. A., and Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* 23, 571–579. doi: 10.1016/S0166-2236(00)01657-X

- Lamme, V., Zipser, K., and Spekreijse, H. (2002). Masking interrupts figure-ground signals in V1. *J. Cogn. Neurosci.* 14, 1044–1053. doi: 10.1162/089892902320474490
- Lamy, D., Salti, M., and Bar-Haim, Y. (2009). Neural correlates of subjective awareness and unconscious processing: an ERP study. *J. Cogn. Neurosci.* 21, 1435–1446. doi: 10.1162/jocn.2009.21064
- Levin, D., and Banaji, M. (2006). Distortions in the perceived lightness of faces: the role of race categories. *J. Exp. Psychol. Gen.* 135, 501–512. doi: 10.1037/0096-3445.135.4.501
- Levitt, J. B., and Lund, J. S. (2002). The spatial extent over which neurons in macaque striate cortex pool visual signals. *Vis. Neurosci.* 19, 439–452. doi: 10.1017/S0952523802194065
- Li, W., and Gilbert, C. D. (2002). Global contour saliency and local colinear interactions. *J. Neurophysiol.* 88, 2846–2856. doi: 10.1152/jn.00289.2002
- Li, W., Piëch, V., and Gilbert, C. D. (2004). Perceptual learning and top-down influences in primary visual cortex. *Nat. Neurosci.* 7, 651–657. doi: 10.1038/nn1255
- Luck, S. J., Chelazzi, L., Hillyard, S. A., and Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J. Neurophysiol.* 77, 24–42.
- Lupyan, G., and Spivey, M. J. (2008). Ascribing meaning to unfamiliar items facilitates visual processing. *Curr. Biol.* 18, R410–R412. doi: 10.1016/j.cub.2008.02.073
- Lupyan, G., Thompson-Schill, S. L., and Swingle, D. (2010). Conceptual penetration of visual processing. *Psychol. Sci.* 21, 682–691. doi: 10.1177/0956797610366099
- Lupyan, G., and Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proc. Natl. Acad. Sci. U.S.A.* 110, 1419–1201. doi: 10.1073/pnas.1303312110
- Macpherson, F. (2012). Cognitive penetration of colour experience: rethinking the issue in light of an indirect mechanism. *Philos. Phenomenol. Res.* 84, 24–62. doi: 10.1111/j.1933-1592.2010.00481.x
- Marx, S., and Einhauser, W. (2015). Reward modulates perception in binocular rivalry. *J. Vis.* 15, 1–13. doi: 10.1167/15.1.11
- Masrour, F. (2011). Is perceptual phenomenology thin? *Philos. Phenomenol. Res.* 83, 366–397. doi: 10.1111/j.1933-1592.2010.00443.x
- Maunsell, J. H., and Cook, E. P. (2002). The role of attention in visual processing. *Philos. Trans. R. Soc. B Biol. Sci.* 357, 1063–1072. doi: 10.1098/rstb.2002.1107
- McAlonan, K., Cavanaugh, J., and Wurtz, R. H. (2006). Attentional modulation of thalamic reticular neurons. *J. Neurosci.* 26, 4444–4450. doi: 10.1523/JNEUROSCI.5602-05.2006
- McAdams, C. J., and Maunsell, J. H. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J. Neurosci.* 19, 431–441.
- Melloni, L., Molina, C., Pena, M., Torres, D., Singer, W., and Rodriguez, E. (2007). Synchronization of neural activity across cortical areas correlates with conscious perception. *J. Neurosci.* 27, 2858–2865. doi: 10.1523/JNEUROSCI.4623-06.2007
- Milner, A. D., and Goodale, M. A. (1995). *The Visual Brain in Action*, Vol. 27. New York, NY: Oxford University Press.
- Motter, B. C. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *J. Neurophysiol.* 70, 909–919.
- Morishima, Y., Akaishi, R., Yamada, Y., Okuda, J., Toma, K., and Sakai, K. (2009). Task-specific signal transmission from prefrontal cortex in visual selective attention. *Nat. Neurosci.* 12, 85–91. doi: 10.1038/nn.2237
- Motter, B. C. (1994). Neural correlates of attentive selection for color or luminance in extrastriate area V4. *J. Neurosci.* 14, 2178–2189.
- Mountcastle, V. B., Motter, B. C., Steinmetz, M. A., and Sestokas, A. K. (1987). Common and differential effects of attentive fixation on the excitability of parietal and prestriate (V4) cortical visual neurons in the macaque monkey. *J. Neurosci.* 7, 2239–2255.
- Ninomiya, T., Sawamura, H., Inoue, K. I., and Takada, M. (2012). Segregated pathways carrying frontally derived top-down signals to visual areas MT and V4 in macaques. *J. Neurosci.* 32, 6851–6858. doi: 10.1523/JNEUROSCI.6295-11.2012
- Nothdurft, H. C., Gallant, J. L., and Van Essen, D. C. (2000). Response profiles to texture border patterns in area V1. *Vis. Neurosci.* 17, 421–436. doi: 10.1017/S0952523800173092
- O'Connor, D. H., Fukui, M. M., Pinsk, M. A., and Kastner, S. (2002). Attention modulates responses in the human lateral geniculate nucleus. *Nat. Neurosci.* 5, 1203–1209. doi: 10.1038/nn957
- Payne, B. K. (2001). Prejudice and perception: the role of automatic and controlled processes in misperceiving a weapon. *J. Pers. Soc. Psychol.* 81, 181–192. doi: 10.1037/0022-3514.81.2.181
- Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behav. Brain Sci.* 22, 341–365. doi: 10.1017/S0140525X99002022
- Pylyshyn, Z. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition* 80, 127–158. doi: 10.1016/S0010-0277(00)00156-6
- Raftopoulos, A. (2001). Is perception informationally encapsulated? The issue of the theory-ladenness of perception. *Cogn. Sci.* 25, 423–451. doi: 10.1207/s15516709cog2503_4
- Raftopoulos, A. (2006). Defending realism on the proper grounds. *Philos. Psychol.* 19, 47–77. doi: 10.1080/09515080500462370
- Raftopoulos, A. (2009). *Cognition and Perception: How Do Psychology and Neural Science Inform Philosophy?* Cambridge, MA: Bradford Books.
- Raftopoulos, A. (2014). The cognitive impenetrability of the content of early vision is a necessary and sufficient condition for purely nonconceptual content. *Philos. Psychol.* 27, 601–620. doi: 10.1080/09515089.2012.729486
- Railo, H., and Koivisto, M. (2009). The electrophysiological correlates of stimulus visibility and metacontrast masking. *Conscious. Cogn.* 18, 794–803. doi: 10.1016/j.concog.2009.01.006
- Reynolds, J. H., Chelazzi, L., and Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *J. Neurosci.* 19, 1736–1753.
- Reynolds, J. H., and Desimone, R. (2003). Interacting roles of attention and visual salience in V4. *Neuron* 37, 853–863. doi: 10.1016/S0896-6273(03)00097-7
- Rock, I. (1983). *The Logic of Perception*. Cambridge, MA: MIT Press.
- Roelfsema, P. R., Lamme, V. A., and Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature* 395, 376–381. doi: 10.1038/26475
- Roelfsema, P. R., Lamme, V. A., and Spekreijse, H. (2000). The implementation of visual routines. *Vision Res.* 40, 1385–1411. doi: 10.1016/S0042-6989(00)00004-3
- Rolls, E. T., Tové, M. J., and Panzeri, S. (1999). The neurophysiology of backward visual masking: information analysis. *J. Cogn. Neurosci.* 11, 300–311. doi: 10.1162/089892999563409
- Schenk, T., and McIntosh, R. (2009). Do we have independent visual streams for perception and action? *Cogn. Neurosci.* 1, 1–15. doi: 10.1080/17588920903388950
- Scholte, H. S., Lamme, V. A. F., Spekreijse, H., and Roelfsema, P. R. (1999). Attention in mental curve tracing task is not zoom-lens or spotlight based. *Invest. Ophthalmol. Vis. Sci.* 40, 272–232.
- Sergent, C., and Dehaene, S. (2004). Is consciousness a gradual phenomenon? Evidence for an all-or-none bifurcation during the attentional blink. *Psychol. Sci.* 15, 720–728. doi: 10.1111/j.0956-7976.2004.00748.x
- Siegel, S. (2010). *The Contents of Visual Experience*. Oxford: Oxford University Press.
- Siegel, S. (2012). Cognitive penetrability and perceptual justification. *Noûs* 46, 201–222. doi: 10.1111/j.1468-0068.2010.00786.x
- Stettler, D. D., Das, A., Bennett, J., and Gilbert, C. D. (2002). Lateral connectivity and contextual interactions in macaque primary visual cortex. *Neuron* 36, 739–750. doi: 10.1016/S0896-6273(02)01029-2
- Stokes, D. (2013). Cognitive penetrability of perception. *Philos. Compass* 8, 646–663. doi: 10.1111/phc3.12043
- Treue, S., and Maunsell, J. H. (1996). Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature* 382, 539–541. doi: 10.1038/382539a0
- Treue, S., and Trujillo, J. C. M. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* 399, 575–579. doi: 10.1038/21176
- van Aaldereen-Smeets, S. I., Oostenveld, R., and Schwarzbach, J. (2006). Investigating neurophysiological correlates of metacontrast masking with magnetoencephalography. *Adv. Cogn. Psychol.* 2, 21–35. doi: 10.2478/v10053-008-0042-z
- Vogel, E. K., Luck, S. J., and Shapiro, K. L. (1998). Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. *J. Exp. Psychol. Hum. Percept. Perform.* 24, 1656. doi: 10.1037/0096-1523.24.6.1656
- Wilbertz, G., van Slooten, J., and Sterzer, P. (2014). Reinforcement of perceptual inference: reward and punishment alter conscious visual perception during binocular rivalry. *Front. Psychol.* 5:1377. doi: 10.3389/fpsyg.2014.01377

- Witt, J. K., Linkenauger, S., and Wickens, C. (2015). Action-specific effects in perception and their potential applications. *J. Appl. Res. Mem. Cogn.* (in press).
- Zeimbekis, J. (2013). Color and cognitive penetrability. *Philos. Stud.* 165, 167–175. doi: 10.1007/s11098-012-9928-1
- Zhou, H., and Desimone, R. (2011). Feature-based attention in the frontal eye field and area V4 during visual search. *Neuron* 70, 1205–1217. doi: 10.1016/j.neuron.2011.04.032
- Zipser, K., Lamme, V. A., and Schiller, P. H. (1996). Contextual modulation in primary visual cortex. *J. Neurosci.* 16, 7376–7389.

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.

Copyright © 2015 Masrour, Nirshberg, Schon, Leardi and Barrett. 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.

Cognitive penetrability and emotion recognition in human facial expressions

Francesco Marchi* and Albert Newen

Department of Philosophy II, Ruhr University Bochum, Bochum, Germany

OPEN ACCESS

Edited by:

Aleksandra Mroczko-Wąsowicz,
National Yang-Ming University, Taiwan

Reviewed by:

Tom Froese,
Universidad Nacional Autónoma
de México, Mexico

Dustin Stokes,
University of Utah, USA

*Correspondence:

Francesco Marchi,
Department of Philosophy II, Ruhr
University Bochum,
Universitätsstrasse 150,
44780 Bochum, Germany
francesco.marchi@rub.de

Specialty section:

This article was submitted to
Consciousness Research,
a section of the journal
Frontiers in Psychology

Received: 03 February 2015

Accepted: 01 June 2015

Published: 19 June 2015

Citation:

Marchi F and Newen A (2015)
Cognitive penetrability and emotion
recognition in human facial
expressions.
Front. Psychol. 6:828.
doi: 10.3389/fpsyg.2015.00828

Do our background beliefs, desires, and mental images influence our perceptual experience of the emotions of others? In this paper, we will address the possibility of cognitive penetration (CP) of perceptual experience in the domain of social cognition. In particular, we focus on emotion recognition based on the visual experience of facial expressions. After introducing the current debate on CP, we review examples of perceptual adaptation for facial expressions of emotion. This evidence supports the idea that facial expressions are perceptually processed as *wholes*. That is, the perceptual system integrates lower-level facial features, such as eyebrow orientation, mouth angle etc., into facial compounds. We then present additional experimental evidence showing that in some cases, emotion recognition on the basis of facial expression is sensitive to and modified by the background knowledge of the subject. We argue that such sensitivity is best explained as a difference in the visual experience of the facial expression, not just as a modification of the judgment based on this experience. The difference in experience is characterized as the result of the interference of background knowledge with the perceptual integration process for faces. Thus, according to the best explanation, we have to accept CP in some cases of emotion recognition. Finally, we discuss a recently proposed mechanism for CP in the face-based recognition of emotion.

Keywords: cognitive penetrability, emotion recognition, adaptation, facial expressions, social perception

Introduction: What is Cognitive Penetration? Does it Really Occur?

Cognitive penetrability is a phenomenon that occurs if higher-level cognitive states, such as beliefs, desires, intentions, etc., can directly influence perceptual experience. In other words, if cognitive penetration (CP) takes place, what one, believes, desires, intends, etc., may alter what one sees, hears, etc. It is currently a matter of debate whether such a phenomenon occurs and, if it does, under which circumstances it is to be expected and how it is to be characterized. A definition of CP is offered by Stokes (2013):

(CP) A perceptual experience *E* is cognitively penetrated if and only if (1) *E* is causally dependent on some cognitive state *C* and (2) the causal link between *E* and *C* is internal and mental. (Stokes, 2013, p. 650)

We shall adopt this definition as a starting point. The main advantage of the definition is that by emphasizing that the relevant link between *C* and *E* must be *internal and mental*, it clearly excludes

instances of bodily movement and changes in non-mental bodily states from the domain of CP. In this section, we introduce the current debate on CP and review some of the reasons for thinking that such a phenomenon occurs, before exploring its possible consequences for the realm of social cognition.

In the twentieth century, the possibility of CP was the core idea behind the *new look* movement in psychology, which studied several alleged cases, albeit without appeal to the precise notion of CP (Bruner and Goodman, 1947; Bruner and Postman, 1949). Later, the idea was almost abandoned in the light of severe criticisms from Fodor (1984, 1988) and Pylyshyn (1984, 1999), who were concerned with the characterization of a reliable visual system that is capable of representing the world adequately, i.e., of delivering some true information. Fodor (1984, 1988) and Pylyshyn (1984, 1999), who introduced the current terminology of penetrability, think of vision as a serial bottom-up process that, roughly, encompasses stimulus onset to categorization. Accordingly, they present several arguments against the possibility of CP. One famous example is Fodor's argument about the impenetrability of visual illusions such as the Müller-Lyer illusion (see below). Driven by the consideration that in order to function quickly and reliably, part of the visual system must work independently of any other cognitive subsystem and domain, Pylyshyn (1999) describes a functionally characterized early visual system that he calls *early vision* (EV), and he reviews several forms of psychological evidence motivating the proposed move. Raftopoulos (2014) has argued for EV on neurophysiological grounds, offering a temporal characterization of EV as the first 100 ms of visual processing. He is led by the observation that there is as yet very little evidence of any top-down modulation of the visual system from areas higher in the brain's cortical hierarchy during this time period. Hence, according to Fodor, Pylyshyn, and Raftopoulos, a significant part of the visual system, and, by extension, its counterparts in other sensory modalities must be considered to be modular in a strong sense. Part of the visual system is domain-specific, an inborn system that can only be influenced by inner-sensory information. It follows from this last point that the processes of the primary visual system cannot be influenced by non-perceptual information. This is especially the case with regard to higher-level cognitive information like background beliefs or mental images. This is the core idea of cognitive impenetrability.

As previously mentioned, one central observation offered in support of the impenetrability thesis is the Müller-Lyer illusion: even if we know that the two arrows have the same length, we continue to perceive one as being shorter than the other. Our perceptual experience seems to be impenetrable to our knowledge of the line's length. However, some researchers have recently challenged the impenetrability claim, observing that in some cultures the illusion does not arise (MacCauley and Henrich, 2006). How can we account for this? One could describe this as a case of long-term adaptation, or of perceptual learning effects that remain intra-perceptual. But how could this modification of perceptual processing take place? The reasoning behind the objection to the impenetrability argument is, roughly, that people who live in highly carpentered environments may develop a form of implicit perceptual knowledge about edges, corners,

and relative distances of geometrical displays that determine the illusion, since the phenomenon is not observed (or is observed to a lesser degree) in cultures that live in non-carpentered environments. Such implicit knowledge is connected to development and long term perceptual interaction between subjects and their environment and, as such, may be relatively stable and not easily overwritten by the currently entertained belief that the two lines are equal. However, if it is indeed a form of knowledge that determines the illusion, under certain assumptions the Müller-Lyer case can be considered evidence of long-term (diachronic) CP.

Pylyshyn also allows for two kinds of interactions between perception and cognition that are compatible with his impenetrability claim. Specifically, higher-level cognitive information may either influence attention, thereby modifying the input to the visual system, or modify the output of the primary visual system after EV has done its work. Both alternatives leave EV impenetrable. Pylyshyn writes:

"Our hypothesis is that cognition intervenes in determining the nature of perception at only two loci. In other words, the influence of cognition upon vision is constrained in how and where it can operate. These two loci are: (a) in the allocation of attention to certain locations or certain properties prior to the operation of early vision [...] (b) in the decisions involved in recognizing and identifying patterns after the operation of early vision. Such a stage may (or in some cases must) access background knowledge as it pertains to the interpretation of a particular stimulus." (Pylyshyn, 1999, p. 344)

Therefore, in arguing against the impenetrability view, the principal challenge is to present convincing cases where the influence cannot be explained with reference to either of the strategies proposed by Pylyshyn, and to show that cognitive information modifies the primary visual system.

In the last decade, there has been a substantial increase in the literature describing in detail those aspects of brain architecture that are compatible with CP. Hard-wired bottom-up mechanisms are not found in the brain: perception is much more interactive and far-reaching in several respects: (i) concerning connectivity, there are many more feedback connections from higher cognitive areas to the primary visual cortex than feedforward connections to higher cognitive areas (e.g., Salin and Bullier, 1995); (ii) concerning timing, there is evidence to suggest that the timing allows for an early activation of brain areas that, if the bottom-up processing view were correct, should only be activated later. The time course of visual processes in V1 and V2 is such that we cannot presuppose simple serial feedforward processing. For example, in the processing of images eliciting perception of illusory contours, the activation of V1 caused by illusory contours emerges 100 ms after stimulus onset in the superficial layers of V1, and at around 120–190 ms in the deep layers of V1. However, in V2, the illusory contour response begins earlier, occurring at 70 ms in the superficial layers and at 95 ms in the deep layers (Lee and Nguyen, 2001). Thus, we must presuppose an interactive temporal dynamics. Furthermore, Bar (2003, 2009) argues that the prefrontal cortex can be activated very early in the processing of

a stimulus and its context, and that it can interact top-down with the visual processing of that stimulus before its completion. Thus, a purely bottom-up view of visual processing is not correct if we adhere to classical views about visual areas like V1 to V5, and nor can such a view adequately account for the relation between ventral visual areas and the prefrontal cortex. We will come back to this issue when speculating about the mechanism of CP. The available evidence so far indicates that CP is a physiological possibility. Having established that CP is physiologically possible, we will now present evidence from empirical studies that cannot be adequately explained without relying on CP.

Core Examples in the Debate

Macpherson (2012) reviews an experiment (Levin and Banaji, 2006)¹ in which knowledge and expectations about the race and skin color of human faces biases a perceptual color-matching task. There are four versions of the experiment. We cannot present all of these in detail (but see Macpherson, 2012; Stokes and Bergeron, 2015, for further discussion). For present purposes, we will focus on the second version, which we take to be the least controversial. Subjects had to adjust a uniform patch of gray to the color of a computer generated target face, which was averaged to display ambiguous facial traits between prototypical African-American and Caucasian faces. The ambiguous face was presented next to either a prototypical African-American or Caucasian face, and all the stimuli were adjusted to the exact same color (surface lightness). The African-American faces and the Caucasian faces were labeled, respectively, “BLACK” and “WHITE,” while the ambiguous face was labeled either “BLACK” or “WHITE,” depending upon whether it was presented next to the Caucasian or African-American face respectively. The experimenters found that even when subjects were presented with the same target stimulus, namely the ambiguous face, they adjusted the patch of gray to a darker shade when it was labeled “BLACK” and to a lighter shade when labeled “WHITE.” The take of the experimenters on this result was that the subject’s knowledge and expectations about the skin-tone associated with a certain race, as triggered by the label, altered their perceptual experience of the color of the target ambiguous face. This experiment has the advantage of requiring the subjects to perform an on-line perceptual-matching task, i.e., the results are not based on subjects’ reports or introspections. This methodology aims to rule out several alternative explanations to CP, such as cognitive influences on the subject’s post-perceptual judgments or pre-perceptual attentional shifts.

Stokes (2014) reports an experiment with a very similar methodology performed by Witzel et al. (2011). Experimenters found that when strongly color-biasing shapes (e.g., a Smurf or a Coca-Cola icon) were presented in a random target-color and had to be adjusted for color to match a gray background, subjects chose a matching shade of gray in the opposite hue-range to the thematic-color. To give an example, a subject may adjust a randomly colored Smurf slightly in the yellow hue, which is the

opposite hue to the thematic color of the smurf (blue). This result is to be expected if the subject sees the randomly colored Smurf as *bluish*. Such an effect did not occur in the control condition, where the same procedure was applied when color-neutral shapes (e.g., a sock or a golf ball) were presented. Importantly, subjects in the experimental condition did not choose a shade of gray in the opposite hue-range to the random target-color of the biasing shapes, but in that of the thematic-color (the usual color of that object). Accordingly, experimenters concluded that the subjects’ knowledge of the thematic color slightly altered their perceptual experience of the target-color. Such results provide support for the idea that CP actually occurs in color perception. However, as Stokes rightly points out, the literature in this field is in its infancy, and few experiments have employed the methodology of on-line perceptual matching. It is plausible that as the literature develops, more evidence for CP in different domains of perceptual experience will emerge. Further evidence of CP includes the evaluation of steepness of slopes (Bhalla and Proffitt, 1999; Durgin et al., 2009) and spatial perception (Stefanucci and Geuss, 2009)².

Another experiment demonstrating the online-influence of concepts on perception was carried out by Winawer et al. (2007). They presented Russian and English speakers with color samples of different shades of blue. The experiment was based on different ways of categorizing shades of “blue” in the two languages: Russian speakers lexicalize the “blue” category by means of two basic level terms: “siniy” for darker blues and “goluboy” for lighter blues. In contrast, English speakers have just one basic-level term (“blue”). The students were asked to decide as quickly as possible whether a color presented at the top matched a color on its left or its right exactly. While all the shades presented were in the same category of “blue” for English speakers, the colors fell under two different basic categories for the Russians. Winawer et al. (2007) found that the Russians—but not the English—had slower reaction times (RTs) in same-color trials (comparing a darker and a lighter shade of blue) than in between-colors trials (comparing a light blue and green).

In addition to the RT results presented above, Carruthers (2015) reviews an analog experiment (Mo et al., 2011) done using EEG-data. The experiment relies on mismatch negativity, measured after 150 ms, indicating the online-influence of early visual processes. Mo et al. (2011) reported mismatch negativity in native speakers of Mandarin, who distinguish two shades of green but not of blue:

“Subjects were required to fixate on a central cross flanked by two colored squares, and were asked to respond as swiftly as possible whenever the cross changed to a circle. The squares were positioned so that the one on the left would be represented initially in the right hemisphere whereas the one on the right would be represented initially in the left (linguistic) hemisphere. As expected, both hemispheres showed a mismatch negativity response to changes in the presented color. But in the right hemisphere there was no difference in the amplitude of the response to changes of color within a category (one shade of green changed to

¹This experiment will be of particular importance in the later sections of the paper. Therefore, we present it in somewhat greater detail than the others mentioned.

²Some of this evidence has been criticized (see, for example Firestone and Scholl, 2014, 2015) and is currently a matter of debate.

another shade of green) versus across categories (a shade of green changed to a shade of blue). However, in the left (linguistic/conceptual) hemisphere there was a significant difference, with a much larger effect for cross-category changes.” (Report taken from Carruthers, 2015)

Finally, Lupyan (2012, 2015) provides further evidence that this experiment cannot be interpreted as involving modular processing of the primary visual cortex. In addition, he offers an alternative model of how inferential processes produce online-modifications in perceptual experience, and provides further examples of CP that are especially related to the interactions between perception and language processing.

Thus, given the available evidence, which does not involve core dimensions of social cognition (except for the aspect of race), it is plausible to accept CP, in principle, for cases of object perception and color perception. But what about social perception? Can we plausibly extend the discussion of CP to this area? In this paper, we aim to show that cognitive penetrability also shapes our perception of socially relevant information. We focus on a clear case of perceptual recognition of socially relevant information, and, specifically, on face-based recognition of basic emotions.

Before proceeding further, however, we must point out that the claim that our perceptual experience of another person’s emotion (the “emotion-percept”) is influenced by memorized images or background beliefs is not entirely new. One line of argument, mainly inspired by phenomenology, supports the idea of CP of emotion recognition by arguing that recognizing the emotions of others is primarily a *direct perceptual* achievement (Gallagher, 2008; Zahavi, 2011; Krueger, 2012; Stout, 2012). Although we sympathize with the direct perception claim with respect to basic emotion recognition (see below and Newen et al., 2015), we want to develop our argument in this article in such a way as to be acceptable even for those who deny direct perception. If we cannot presuppose that the content of a percept is rich, i.e., that it involves rich images as well as conceptual information, it becomes much more difficult to argue that obvious changes in the recognition of emotion rely on a change of the percept, instead of a change of judgment alone. Furthermore, our main claim converges with the position that emotion, cognition and perception cannot be neatly separated into distinct modules (Pessoa, 2013; Colombetti, 2014), which draws support from emotion science. But it is important to note that the debate about CP would be empty if one were to hold the view that cognition and perception could not be separated at all. Thus, we are presupposing a minimally clear separation of the *perceptual experience* (be it conceptual or non-conceptual), and the *judgment* based on this perceptual experience.

Perceptual Adaptation and the Experience of Facial Expressions

Given the complex nature and extreme relevance of human faces in our perceptual life, it is an interesting question whether recognition of an emotion in a human face is achieved through a judgment made on the basis of perceptual experience, a purely perceptual automatic process, or an interaction between both that admits some degrees of CP. In order to argue for the third of these

options, we start with the question of whether we can perceive facial expressions as wholes, or whether the evaluation of a facial expression depends on post-perceptual processes.

The structure of our argument, presented in more detail, runs as follows: in the first step, we argue for a process of feature integration in the case of facial expressions of emotions, and claim that this is a perceptual process. The integration process we have in mind consists in the gradual combination of facial features and cues into complex compounds. By discussing perceptual adaptation to facial expressions of emotions, we show that there are reasons to think that the resulting compounds, i.e., whole facial expressions, have to be considered as perceptual states. Secondly, we argue that such perceptual integration processes can be influenced by contextual background knowledge, such that we have to accept that the social perception of emotion involves CP.

Human faces are complex stimuli. They are arguably one of the richest and most reliable sources of information available to us in our everyday lives. Two of many examples of phenomena based on face perception that constitute a significant subset of the perceptual development of a healthy human subject include gaze following and joint attention. According to some researchers (Dunbar, 1998; Adams and Kveraga, 2015), the enormous amount of information conveyed by human faces is of such relevance for behavior and social interaction that it is plausible to think that humans have evolved a dedicated perceptual sub-system for quickly integrating the various social cues conveyed by a face into meaningful compounds.

The phenomenon of pareidolia (e.g., Hadjikhani et al., 2009) provides further evidence for the existence and relevance of an integration mechanism for faces: we tend, for instance, to see faces in natural collections of sand or in cloud formations, because the integrated patterns are extremely important for humans and can be easily activated in various situations. Furthermore, the widely accepted empirical model of face-based recognition of emotion proposed by Haxby et al. (2000) and Haxby and Gobbini (2011) involves the following two-step process: (1) the construction of facial identity and (2) the recognition of facial expressions. The latter, extended part of recognizing a facial expression, is supposed to involve such an integration process of core facial features. Furthermore, recent models analyze normal object perception as involving Bayesian processes of cue integration and cue combination (Ernst and Bühlhoff, 2004), and it is very plausible that the principles of perception remain the same in the case of non-social objects and in the case of perceiving emotions in faces (Newen et al., 2015). Thus, it is very plausible to accept a feature integration process in the case of recognizing the expression of an emotion in a face, or recognizing a face in certain perceptual configurations. However, it is not clear whether faces and facial expressions as wholes are perceptually processed or not. In fact, it may be that even if there is a feature integration process at play, facial expressions are only recognized post-perceptually on the basis of certain perceptual arrays of lower-level features. Why should we take this integration process and its results to be perceptual?

In the present section, we present the first step of our argument as outlined above. In particular, relying on evidence recently reviewed in Block (2014), we show that in some cases the

proposed result of the integration process, i.e., a whole facial expression, shows perceptual adaptation. Under the assumption that adaptation is a perceptual process, and that only perceptual states/contents may adapt, it follows that since facial expressions as wholes show adaptation, facial expressions as wholes are perceptually processed. In the next section, we show that the perceptual integration process of facial expressions may be influenced by contextual background knowledge and the subject's beliefs.

Perceptual adaptation consists in a process where being exposed to a certain perceptual feature (or set of features), either repeatedly or for a long time, makes that feature less likely to be detected in other stimuli. One explanation for this adaptation is that the firing threshold of the neurons that code such feature in the perceptual system is raised by prolonged exposure. Block (2014)³ addresses this phenomenon in the case of facial expressions of emotion, focusing on the problem of whether the nature of certain adaptation effects is perceptual or cognitive.⁴ Block reviews an experiment by Butler et al. (2008). In this study, experimenters found that whether a still picture of a face displaying an emotional expression, ambiguous between anger and fear, was more or less likely to be perceived as expressing anger or fear depending upon previous exposure to a clearly fearful or clearly angry face. Most importantly, the effect was found to persist when the low-level features of the face were varied, as long as the emotion expressed was kept constant. This seems to be a clear case of perceptual adaptation. The exposure to a clearly angry face raises the threshold for detecting anger-related features in the subsequently presented ambiguous face, and the opposite happens in the case of exposure to a clearly fearful face, which is then perceived as expressing fear.

Concerning this case, Block writes:

"[...] can we be sure from introspection that those "looks"-[fearful/angry] - are really perceptual, as opposed to primarily the "cognitive phenomenology" of a conceptual overlay on perception, that is, partly or wholly a matter of a conscious episode of perceptual judgment rather than pure perception?" (Block, 2014, p. 7)

Providing an answer to this question is difficult, but Block thinks that there is reason to reply in the affirmative, and thus to consider adaptation to facial expression to be a perceptual phenomenon. The preliminary reason for this conclusion, according to Block, is that concepts are in general much more resilient to adaptation than percepts. In particular, Block argues that in cases of ambiguous pictures, we find a form of multi-stable perception in which two percepts are alternatively perceived, and that this switching is the result of perceptual adaptation.

The alternation of the two percepts works according to the three properties of exclusivity (only one percept at a time), inevitability (the alternation will surely happen at some point), and randomness (there is no function of duration for each percept).

³Block's case is framed in the context of Burge's (2010) discussion of perceptual attributives. In this paper, we shall try to phrase the discussion in more general terms.

⁴Here, "cognitive" is used in the sense of Block, as concerning conceptual and propositional states.

Block assumes that correspondent judgments and beliefs are not subject to an alternation that works according to the same three properties even in highly conceptually ambiguous situations, and concludes that there is no such thing as conceptual adaptation.⁵

As further evidence, Block considers an experiment by Schwiedrzik et al. (2014), in which subjects were first exposed to a clearly oriented (either 90° or 0°) grid-like stimulus, and had to report the orientation. Afterward, they had to evaluate the direction of tilt of an ambiguously oriented grid-like stimulus of the same kind. The experimenters found that there was an adaptation effect in the reports of the orientation of the second stimulus that depended upon the *objective tilt* of the first stimulus, not its *reported tilt*. In other words, when there was a discrepancy between the objective and the reported tilt of the first stimulus, the subsequent adaptation effect was consistent with the former, not the latter. According to Block, this means that subjects showed an adaptation effect that depended exclusively on what they actually saw, not on what they thought they saw. Therefore, adaptation effects have to be considered to be purely perceptual phenomena.

For present purposes, it is very important to note that Schwiedrzik et al. (2014) investigated adaptation and the different phenomenon of priming, in the same experiment, as two opposite effects. Priming is basically the facilitation of detecting a certain perceptual feature (or set of features) as triggered by a briefly presented previous stimulus, called the prime. While adaptation is exclusively triggered by prolonged exposure to a perceptual stimulus, priming can be triggered by a prime of the same or similar perceptual kind as the target, or by a prime that is semantically related to the target, i.e., a word. Schwiedrzik et al. (2014) monitored the cortical activity of the subjects and, consistent with what has just been said, found that adaptation involved only areas V1 and V2, while priming involved a wider range of cortical areas. This data shows that adaptation is largely independent of the subject's judgment about their experience, and that the locus of adaptation is mainly in the visual cortical areas, lending further support to the idea that adaptation must be considered a purely perceptual phenomenon.

Facial expressions of emotions are complex stimuli, constituted by specific arrangements of lower-level facial cues like eyebrow orientation, mouth shape, etc. Hence, if facial expressions of emotions as a whole show adaptation and, conversely, if a perceptual system can adapt to facial expressions as a whole, this means that such a system is capable of detecting lower-level facial features and integrating them into meaningful compounds,⁶ even before corresponding judgments about the emotion expressed by the faces are formed. If this is correct, it is clear that the integration-process we just described is sensitive to and is directly affected by different factors such as lower-level feature saliency and different kinds of attention. In addition, as we aim to show,

⁵On this topic it is worth noting that on Mroczko-Wąsowicz (2015) construal, "Phenomenal adaptation" is a broader notion that may include non-sensory states. As she points out (p. 2), however, such a notion is quite different from the uncontroversial physiological notion of a perceptual adaptation, which is the one Block employs. We remain neutral with respect to the broader phenomenon of phenomenal adaptation. Nevertheless, following Block, we hold that the more constrained phenomenon of perceptual-adaptation does not involve non-sensory states, which suffices for our argument here.

⁶This idea is proposed by Adams and Kveraga (2015), see section below.

the perceptual integration process can be influenced by previously formed expectations and beliefs. We will now present a case study in which one finds an effect that seems to be just such a case, where the integration process is influenced by contextual background knowledge.

Face-Based Recognition of Emotion is Sensitive to Background Knowledge

The upshot of Block's argument is that it is plausible to think that facial expressions of emotions are processed as compounds that are largely the result of a feature integration process belonging to perception, insofar as it shows adaptation. The reasons, as we have seen, are that adaptation to facial expression is at least partly independent of the lower-level features constituting the expression, and that concepts and other cognitive features do not adapt the way percepts do, thus ruling out the possibility that adaptation depends on higher-level cognitive features. What adaptation shows is that if the perceptual system is exposed over a prolonged period to a facial expression of emotion *x*, the exposure will affect the integration process such that it will be less likely that *x* is recognized as being expressed by a subsequent similar facial expression. In other words, the integration process that gives rise to the emotionally meaningful perceptual compound associated with *x* is sensitive to stimulus familiarity. In this section, we will present a case in which the same perceptual integration process seems to be sensitive to the background knowledge of the subject. We argue that if this is the case, then we are dealing with the clear and direct influence of knowledge on perceptual processing and, plausibly, on the corresponding perceptual experience. If this is correct, such a case qualifies as an instance of CP in social perception.

The experiment of Butler et al. (2008) reviewed by Block shows that perceptual experience of facial expressions, expressions of emotion in particular, is sensitive to adaptor stimuli that bias the interpretation toward a different emotion. Moreover, Block's discussion points to the fact that this phenomenon may plausibly be considered purely perceptual. Our case study presents a very similar effect on the facial expression of emotion, in which different emotions are recognized as being expressed by the same face. The experimental condition is actually very similar to Butler et al. (2008). The main difference between the two studies is that in the case we report, what triggers the shift in the integration process is not a perceptual adaptor (like another facial expression, as in the above case), but a subject's expectations, which are driven by her background knowledge and activated by a form of conceptual priming.

The experiment we will discuss was carried out by Carroll and Russell (1996). The participants had to evaluate the emotion expressed by a human face. Subjects were presented with combinations of faces and situations. The target stimuli were still photographs of posed facial expressions, selected from among the prototypical facial expressions of fear, anger, or sadness, as collected in Ekman and Friesen (1976). Such prototypical facial expressions have the peculiar characteristic of being reliably evaluated as expressing the same emotion across different subjects and cultures (Keltner et al., 2003), in cases where no additional information is available. Situations were provided in the form of

short stories concerning the persons depicted in the stimuli. Such stories were designed to trigger an emotional response of fear, anger, or disgust. Subjects were first told the story, and then shown the picture. They then had to evaluate the emotion expressed by the face by choosing one of six possible emotion labels.

Carroll and Russell addressed the possibility that providing contextual information to subjects may alter which emotion is recognized as being signaled by the prototypical facial expressions. For simplicity, we present only the pairing of an anger-situation with a fearful face. The situation was provided in the form of the following story:

This is a story of a woman who wanted to treat her sister to the most expensive, exclusive restaurant in their city. Months ahead, she made a reservation. When she and her sister arrived, they were told by the maitre that their table would be ready in 45 minutes. An hour passed, and still no table. Other groups arrived and were seated after a short wait. The woman went to the maitre and reminded him of her reservation. He said that he'd do his best. Ten minutes later, a local celebrity and his date arrived and were immediately shown to a table. Another couple arrived and were seated immediately. The woman went to the maitre, who said that all the tables were now full and that it might be another hour before anything was available.⁷

The researchers found that when presented with such contextual information, the vast majority of subjects evaluated the face as signaling anger. When the contextual information was not presented, however, subjects evaluated the same face as expressing fear, in accordance with Ekman's earlier findings.

Can we be sure that this effect demonstrates the influence of background knowledge on perceptual processes, and that it is not only a product of modifying our perception-based judgment?⁸ Assuming, for the reasons discussed above, that the perceptual system is capable of integrating different low-level facial cues into meaningful compounds, it is clearly possible that in the present case, the background knowledge (based on *conceptual semantic priming*) provided by the story actually interferes with such an integration process.⁹ There are two possible positions that may be taken in response to this. According to the previously mentioned approaches inspired by continental phenomenology, emotions are always directly perceptible in visual experience. If this is the case, however, the possibility that emotion recognition on the basis of

⁷Carroll and Russell (1996, p. 208).

⁸Our notion of Judgment is neutral on how judgments are to be understood. To be clear, we do not think of judgments as necessarily explicit propositional states. Rather, we allow for the possibility of implicit and automatic perceptual judgments.

⁹This interaction should work in the same way as in the Butler et al. (2008) case, albeit in the opposite direction. Adaptation and priming can, in some sense, be thought of as two sides of the same coin. As Block points out, the former makes certain things harder to perceptually process, while the latter makes them easier. If we have a perceptual integration process that binds together lower-level features in order to create emotionally meaningful compounds, different factors can make some of these compounds harder or easier to construct, as in, respectively, the adaptation and priming cases. Hence, our account has the advantage of providing a straightforward and unified explanation of both cases.

facial expressions is the upshot of a cognitive inferential process of judgment [i.e., judgment shift (JS)] seems to be excluded.¹⁰ On the other hand, if we accept that emotion recognition may be the result of a cognitive inferential process, the question that arises is whether, under certain conditions, the perceptual experience that underlies such process may be modified by a subject's background knowledge or some other of his cognitive states. We will not discuss the motivations for adopting either of these positions here. Instead, we will argue that even if emotions are not directly perceivable, there are reasons to think that the perceptual process that leads to emotion recognition on the basis of facial expressions is penetrated by higher-level cognitive states.

Emotion Recognition: Perceptual Categories and Judgments

Even if one accepts *priming* in the case of the facial expression of emotions, one can still doubt that the evidence provided above constitutes a clear case of the conceptual priming of perceptual experience, as opposed to a case of the conceptual priming of perceptual judgment. We will now propose some additional reasons to support the perceptual (as opposed to conceptual) nature of the effect of background information on the recognition of emotion expressions. Our argument takes the form of an inference to the best explanation, intended to show that the CP of perceptual experience provides a better explanation for shifts in emotion attribution, as compared to the alternative explanation that involves perceptual judgments.

The phenomenon to be explained is the recognitional shift that subjects are ready to make when provided with additional information about the emotion expressed by a face, where that face is otherwise reliably taken to signal a specific emotion. Our argument takes the form of an inference to the best explanation,¹¹ so we need to put two competing explanations on the table. The two alternatives we shall consider are CP and JS:

CP: Subjects recognize two different emotions as expressed by the same face on the basis of two different perceptual experiences of that face.

JS: Subjects recognize the same face as expressing two different emotions by forming two different perceptual judgments on the basis of the very same perceptual experience.¹²

There are several things to consider here. First of all, it is a widely studied phenomenon that, taken out of context, certain human facial expressions tend to signal one specific emotion and not others very reliably.¹³ Secondly, it is known

that contextual information may alter the kind of emotion that the face is taken to signal. This happens both in cases of a change/enrichment of perceptual context (for the visual case, see Aviezer et al., 2008; Hassin et al., 2013) and in cases of conceptual priming, as described above. The most important point, however, is that shifts in emotion recognition do not happen arbitrarily. Even if a prototypical facial expression of fear can be taken to signal anger under certain conditions, there are some constraints that make it highly unlikely that such a prototypical expression of fear could ever be taken to signal a radically different emotion, such as joy.¹⁴ We shall argue that these constraints are best explained as perceptual constraints. That is to say, the different possible emotions that subjects are ready to recognize as expressed by a particular face depend on the perceptual integration of different low-level features of the face itself, like mouth shape, eyebrow orientation, gaze, and so on. We shall call such features facial cues. According to JS, a subject may recognize a prototypical facial expression of fear as expressing anger by forming different judgments on the basis of the same perceptual experience of a fearful face. If this were the case, however, we do not see how constraints on emotion recognition could be introduced in a principled way. If recognizing an emotion were only a matter of judgment, it would seem possible, regardless of the epistemic confidence of the subject, to provide enough background information for the subject to revise his judgment from one of recognizing fear to one of recognizing joy. This, as per our assumption, cannot be the case. One might argue that there are indeed such cases of radical JSs. For example, if someone were to tell you that the person in the target picture has a rare dysfunction in her facial muscles that forces her to adopt a fearful expression whenever she is joyful (and *vice versa*), you might in the end come to the correct evaluation of an expression of joy in the fearful face. This illustrates that we can adapt our judgments, but only at a later stage. We need to presuppose that—at least at the beginning of noticing such a special case—the face is rightly recognized as expressing fear and only subsequently evaluated as expressing joy, on the basis of background information. After the initially correct recognition of fear, subsequent judgments that associate the face with a different emotion can be made without constraint. But if JS were true, even the initial recognition judgment would be subject to such unconstrained flexibility, which is implausible in the light of the strong reliability of emotion recognition. Therefore, we do not see how a principled way of constraining emotion recognition can be introduced at the level of pure judgment. This is not to say that it is in general impossible to introduce such constraints, only that, as we shall see, it is much more straightforward and empirically more plausible that the required constraints work at the level of perception.

Here, one might try to reinforce JS by taking into account similarity of stimuli, and say that if we are right, then our argument

¹⁰See, for example Froese and Leavens (2014) for a discussion of the interaction between perceptual experience of various physical features (including facial expressions) and conceptual categories from the perspective of the direct perception hypothesis.

¹¹This argument echoes some of the considerations above concerning perceptual adaptation.

¹²By saying that two experiences could be the same or different, we mean that they could be token-experiences of the same type or of a different type.

¹³Here, we do not need to take a stance in the debate between dimensionalist views of emotion and views that posit basic emotions. For a theory of emotion that fits nicely with our proposal, see Barlassina and Newen (2014).

¹⁴We do not inquire which specific shifts are allowed and which are not; for the present argument, it is sufficient that emotion recognition changes on the basis of background information do not happen arbitrarily. However, Carroll and Russell (1996) review previous findings (e.g., Tomkins, 1962, 1963) showing that not all background information leads to such a shift. Specifically, the shift does not happen in the case of joy-related information and an anger signaling face (p. 17).

should apply to a whole lot of different cases of perception-based judgment. For example, one might come up with the following case:¹⁵ there is a picture depicting my very similar-looking twin (but who is noticeably different in some matters of detail) wearing a red coat. If one sees the picture and knows that I like to wear red coats, one might mistakenly recognize me in the picture instead of my twin. However, the counter argument goes on, this seems to be a clear case of a mistaken perceptual judgment that requires no difference in the perceptual experience of the subject. Why cannot the case above be explained along the same lines? We argue that the consequences of such an account are less plausible than our alternative explanation. The problem that the JS explanation faces comes in the form of a dilemma. The defender of JS might either (1) propose that the two kinds of stimuli of fearful faces and angry faces are very similar to each other and (both) very different from joyful faces, or (2) claim that they are not so similar.

If one goes with (1), and proposes that such stimuli are similar, then one could say that the similarity and ambiguity between fearful and angry faces, which they do not share with joyful ones, could explain why, on the basis of the very same fearful-face experience, subjects are allowed to activate fear judgments and anger judgments but not joy judgments: so far, so good. However, in this case, one faces the serious problem of how to account for the high reliability of emotion recognition across different subjects and cultures. Even if one does not buy into the original basic-emotion framework, the studies conducted by Ekman and colleagues provide quite compelling reasons to think that the overwhelming majority of subjects¹⁶ are at least capable of making very clear perceptual discriminations between different facial expressions of the basic emotions: people of different cultures can reliably distinguish between anger, fear, disgust, sadness, and surprise, and can reliably combine the judgment with the facial expression, given a selection of basic emotions. How can a defender of a JS explanation account for such reliability? If some of the target faces for basic emotions of fear and anger are supposed to be very similar, we would expect a higher rate of mistakes from subjects evaluating which face expresses which emotion.

If, on the other hand, one goes with (2) and claims that the stimuli are not similar, one needs to accept that, in order for the judgment to shift from anger to fear, almost all the perceptual information conveyed by the target fearful face must be disregarded. But, if this were the case, then the judgment would no longer be perception-based. Moreover, if the evidence is disregarded, nothing prevents additional background information shifting the judgment even further to a radically different emotion, thus generating the problem of how to constrain possible judgments discussed above. Thus, if JS fails to adequately account for the relevant constraints, we need to see whether CP fares any better.

We want to highlight that with CP, we have the possibility of collocating the required constraints at the lower perceptual level of facial-cues. In fact, a straightforward way of accounting for these constraints is to think of them as a range of shared possible values of lower-level facial cues for different emotions. According to this

view, in order to explain why anger is recognized in a prototypical fearful face, one needs only suppose that the integration process in the target case highlights the relevance of the shared features. Such features are selected on the basis of background information and expectations, and bound together into an anger-signaling compound. Hence, we have two distinct perceptual compounds, a fear-compound in the case of no conceptual priming, and an anger-compound in the case of conceptual priming. Most importantly, by explaining the difference on the basis of two different compounds, we avoid the dilemma depicted above for the defender of JS. If the integration process is affected before a compound is formed, we can easily understand the possibility that only some relevant perceptual information conveyed by the face is disregarded or given increased saliency. This is precisely what allows two different compounds to be formed. Hence, the recognition process need not disregard the whole information conveyed by the final compound. At the same time, we need not assume that facial expressions for different emotions need to be largely similar. In previous sections, we argued that such compounds are integrated at the level of perception. We therefore hold that different compounds give rise to different experiences,¹⁷ and that on the basis of these different experiences, two different emotions are recognized.¹⁸

Hence, CP provides a natural way of explaining why certain recognition outputs are allowed and certain others are not. Which emotion can be recognized in a facial expression depends on the nature, number, and relevance of shared features across different facial expressions and on the integration process. Different outputs of the integration process in turn give rise to different perceptual experiences. Therefore, CP constitutes a better explanation than JS for both the reliability and the (limited) unreliability of emotion recognition across different subjects, insofar as it provides a principled way of constraining the results to be expected. Thus, we conclude that Carroll and Russell (1996) provide a case of CP of perceptual experience, and, more generally, that the perceptual experience of facial expression of emotions is sensitive to background knowledge and expectations. In the next section, we briefly present a recently developed neuro-functional mechanism that supports our view of emotion recognition. If we are correct so far, it seems that CP fares better than JS in accounting for the constraints on possible emotion recognition on the basis of the same stimulus. In addition, we will present further evidence offering independently support for CP over JS. Our strategy is to show that emotion recognition—at least in the case of basic emotions—can be carried out in large part by the perceptual system alone. Therefore, since we presented evidence of particular cases in which background beliefs and knowledge can influence emotion recognition, that influence must be exerted

¹⁵We are grateful to Peter Brössel for this example.

¹⁶See Ekman and Friesen (1971).

¹⁷Whether such difference in the experience is best characterized as a difference in content or as a difference in the phenomenal character of the two experiences (or both) is an important open question. However, it goes beyond the scope of the present paper.

¹⁸As Jackendoff (1987) and Prinz (2012) argue, further support for this claim comes from introspection. Introspectively, we have experiences of integrated objects (including faces) and not of unbound low-level features. Therefore, we should situate the locus of conscious perceptual experience after some sort of integration process has taken place, not before.

at the level of perception as CP describes, not at the level of post-perceptual cognitive judgments described by JS.¹⁹

Emotion recognition is a complex process that may involve several perceptual and cognitive mechanisms (see Adolphs, 2002, for an extensive review). However, there is reason to think, at least in the case of basic emotions such as fear, anger, joy, etc., that a large part of the process is carried out by the perceptual system alone. First of all, if an organism's perceptual system were capable of quickly and automatically processing critical social stimuli and reliably associating these with appropriate behavioral responses and other key features such as non-verbal sounds and lexical labels, this would provide a clear adaptive advantage for the organism. Evidence for this possibility in the case of facial expressions of emotions comes from several sources. One example is research into primates' facial expressions (Redican, 1982), which shows that in a comparison of new world monkeys (prevalently arboreal) and old world monkeys (prevalently terrestrial), only the latter, which can rely on visual contact with conspecifics, have developed a complex system of facial expressions. This supports both the close connection between facial expressions of emotions and vision and the social value of perceptual integration of facial expressions of emotions (Adams and Kveraga, 2015).

Further interesting evidence for the perceptual nature of emotion recognition comes from computer models (discussed in Adolphs, 2002) designed to achieve comparable performance to humans in evaluating when two facial expressions belong to a different emotional category (even when the structure of the two stimuli is very similar), but that cannot rely on any form of conceptual knowledge about emotions. Moreover, evidence from perceptual priming studies (Carroll and Young, 2005) shows that facilitation effects on emotion recognition are sensitive to the emotional category of the primes (e.g., anger vs. disgust), not only to the positive or negative valence of the emotions. In combination, the evidence discussed here provides support for a quick and reliable perceptual process of emotion recognition that relies on clearly separated perceptual categories that may not always need conceptual knowledge. Hence, if emotion recognition is achieved on the basis of a quick process that relies on discrete perceptual categories, this undermines the claim that cognitive judgment plays a strong role in emotion recognition. Now, if emotions are categorized at the level of perception, shifts in categorization that depend on contextual information (such as those discussed in the previous section) seem to be plausibly explained as special cases, in which background knowledge interferes directly with the perceptual process that leads from feature detection to perceptual categorization, in accordance with CP.

A further consideration in favor of CP is that of explanatory parsimony. If one accepts CP in color perception (Levin and Banaji, 2006), an explanation of the form of CP needs to already be available. Critically, the color case has many relevant similarities with Carroll and Russell (1996). In both studies, target stimuli were of the same broad perceptual kind, namely human faces.

In both studies, relevant background knowledge was triggered by conceptual information (a story and a verbal label respectively). However, recall that in Levin and Banaji (2006), subjects were required to perform a perceptual matching task, which rules out the possibility that the influence of racial categories could have been exerted at the level of judgment. Hence, it seems that a CP explanation could account for both cases, whereas JS could account only for the emotion study. If we admit that background knowledge can interfere with the perceptual processing of certain facial features, such as skin color, why should we not favor the same line of explanation (CP) in the case of perceptual processing of other facial features, such as expressions of emotion?²⁰

To conclude this section, we wish to examine a final worry based on the claim that the phenomenon described Carroll and Russell (1996) depends on a shift in the subject's attention, and that it is therefore not a case of CP. This strategy is the one adopted by Pylyshyn to rule out most cases of CP. We need to show that it does not apply in the present case. Pylyshyn (1999) thought that attention shifts exclude CP because the functional role of attention is basically to select (or gate) a subset of the available perceptual information as an input to EV. If this were always the case, a shift in attention would be a pre-perceptual effect amounting to a shift in the input, similar to looking in a different direction in order to gather more information about a stimulus. The resulting perceptual experience would still be different, but it would be causally dependent on such input shift, and this would not be an interesting case of CP. However, we now know that attention shifts can have different effects while the input remains stable.

Here, we have two things to say to counter Pylyshyn's view. First, it is questionable whether the role that Pylyshyn assigns to attention is the correct or the only possible one. Views of attention differ significantly in terms of the functional role they assign to attention and its underlying processes.²¹ Therefore, it is not so clear that the scope of attentional modulation of perception can be constrained in such a way as to rule out the possibility that attention affects the whole scope of visual processing, including EV. Second, we have seen that if we accept that facial expressions as wholes are perceptually integrated into complex compounds from lower-level facial cues, this must happen after the lower-level cues that constitute such compounds have been processed. Hence, an attentional shift on a facial expression can either affect how the features are integrated, or how the resulting compound is processed. In both cases, it would be an effect that alters perceptual processing itself, not a pre-perceptual effect that changes the input, as Pylyshyn conceived of it. Thus, even if one wishes to call this an attentional shift,²² it is nevertheless a shift that happens within perceptual processing, not before. Hence, the case does not meet Pylyshyn's requirement of attention changing the input to perception. Consequently, it does not undermine CP.²³

²⁰We know from the previous section that facial expressions are perceptually processed as wholes.

²¹See Mole (2011) for a radically different view of attention, and see Mole (2015) and Stokes (2014) for a discussion of attention and its relation to cognitive penetrability.

²²More on this below.

²³We would just like to mention that a CP explanation is consistent with very recent models of emotion recognition and facial expressions such as Carruthers (2015) and Haxby and Gobbini (2011).

¹⁹The evidence we present below is in line with a form of direct perception for basic emotions.

The Mechanism: Neural Shortcuts, Compound Cues Integration, and Social Vision

So far, we have proposed two reasons for taking the experiment conducted by Carroll and Russell (1996) as evidence for the cognitive penetrability of perceptual experience. The first is that facial expressions of emotion show adaptation, and should therefore be considered as perceptually integrated compounds. The second is that CP is a better explanation for the constrained shifts that can happen in emotion recognition on the basis of background knowledge. However, we have not yet proposed a plausible candidate mechanism that supports such penetration effects.

Before discussing a candidate, we should outline the framework for the search for such a mechanism. It is an open question whether there is only one mechanism that accounts for top-down influences on perceptual integration processes. We have argued elsewhere that we need to distinguish different types of CP (Vetter and Newen, 2014) that may reasonably be assumed to have different underlying mechanisms. We want to describe two routes of top-down influences that are not the preferential candidates for explaining our core example, before outlining a plausible candidate. Top-down influences on perceptual processes may be produced because newly activated beliefs shift our attention and thus relevantly modify the sensory input. Although, as we mentioned above, if attention is conceived differently from Pylyshyn's account, it may sometimes be a possible mediator of CP, this does not seem to be what happens in the case of contextual background stories (see above). The important candidates as mechanisms of top-down attention modulation are reviewed in Baluch and Itti (2011). A second consideration is that background knowledge is conceptual, and needs to be transformed into a perceptual format before it can causally influence purely perceptual processes. Macpherson (2012) proposes that the top-down modulation of perceptual processes can only be indirect, modulated by activating the relevant imagery. This, however, would only be true if conceptual representations were absolutely separated from imagery and sensory representations. This traditional view of concepts as purely cognitive has been radically called into question by recent data and theories, including embodied concept formation (Barsalou, 1999; Pulvermüller, 2003; Pulvermüller and Fadiga, 2010). Thus, it remains a reasonable option to look for a mechanism that involves direct causal top-down-influences and that may not be purely attentional.

Fortunately for us, there is already a theory available that posits such a top-down mechanism in the case of stimuli that have relevance for social interaction, a paradigmatic class of which is human faces. Moreover, this theory has both a functional component and a neurophysiological model of implementation. The model in question is that of compound social-cues integration (Adams et al., 2010; Adams and Nelson, 2011; Adams and Kveraga, 2015), which relies on the studies of Bar (2003, 2009). According to this view, the anatomy of the visual system supports quick recruitment of higher-level cognitive areas, such as the orbitofrontal cortex (OFC),

before a visual stimulus is recognized.²⁴ This is possible because the retinal projection of a visual stimulus activates a specific "neural-shortcut," the magnocellular-pathway (M-pathway), mostly identifiable with the dorsal visual stream.²⁵ The M-pathway is known to quickly²⁶ project coarse information about the stimulus to the associative areas of OFC. OFC, in turn, presents feedback projections to areas in the ventral stream, including recognition areas in the infero-temporal cortex (IT). Of course, we cannot make inferences from neuroanatomical to functional mechanisms easily. Nevertheless, the existence of many specific and very quick feedback connections in the brain shows at least that nothing in neuroanatomy prevents the occurrence of a process of CP such as the one described above. Moreover, the feedback loop from prefrontal areas (typically associated with reasoning and conceptual knowledge) to visual areas seems to be a plausible preliminary candidate for a neural correlate of CP.

Provided that neuroanatomical characteristics of the brain support the idea of a modulation of perceptual integration exerted by background knowledge, Adams and Kveraga (2015) argue that different social cues, such as gender, age, posture, etc., are relevant to such perceptual integration processes, which they call social vision. In previous sections, we have already provided a sketch of their model, which claims that one of the main tasks of vision is precisely to deliver such integrated meaningful compounds. According to these authors, the plausibility of the idea is supported by evolutionary and everyday considerations about the importance for human beings and other animals of being able to quickly integrate as much socially relevant information as possible. For the purposes of the present paper, however, we need not delve into much detail about the social-vision view. It suffices for our argument that facial-cues, such as eyebrow orientation, mouth shape, gaze direction, and perhaps other facially evident cues such as gender and age, are perceptually integrated together in order to form meaningful emotion-signaling compounds.

If one admits that such integration is possible at the level of the face, then our considerations concerning adaptation and principled constraints on emotion recognition should be enough to show that under certain conditions, the integration process is sensitive to background knowledge, expectations and, possibly, to other high-level cognitive features. We are aware that this is a somewhat unusual way of arguing for CP. We think, however, that perception is a much more dynamic and integrative process that it is described to be in the traditional modular model, and that the evidence we have presented here supports this view. Hence, we conclude that the boundary between perception and cognition should be at least partially blurred.

Conclusion and Outlook

Cognitive penetration is not only a plausible claim about the perception of objects and physical scenes, but also about the social

²⁴See also Bar (2003, 2009) and Kveraga et al. (2009, 2011).

²⁵See Milner and Goodale (1995).

²⁶As quick as 80 ms.

perception of emotion. The results presented here indicate that we should even go further, and start to investigate the extent to which the perceptual recognition of other social and mental phenomena is shaped by CP. We suggest that face-based recognition of emotion is only one basic component of the most important integration process for humans, namely the integration on the level of person perception (Macrae and Quadflieg, 2010). Person perception is accompanied by an impression formation that should also be explained by a systematic interaction of bottom-up and top-down processes, constituting a person impression

(Newen, 2015). Thus, we suggest future work investigating whether CP also holds for the formation of a complex person impression based on perception. One further interesting upshot of this line of investigation is that perceptual processes may essentially rely on the same type of bottom-up and top-down mechanisms, despite the fact that physical objects like trees and social objects like human faces provide us with radically different inputs, and despite the observation that some social stimuli are processed in highly functionally specialized brain areas, like FFA (fusiform face-area) for faces.

References

- Adams, R. B., Franklin, R. G., Nelson, A. J., and Stevenson, M. T. (2010). "Compound social cues in human face processing," in *The Science of Social Vision*, eds R. B. Jr. Adams, N. Ambady, K. Nakayama, and S. Shimojo (New York, NY: Oxford University Press), 90–107.
- Adams, R. B., and Kveraga, K. (2015). Social vision: functional forecasting and the integration of compound social cues. *Rev. Philos. Psychol.* doi: 10.1007/s13164-015-0256-1 [Epub ahead of print].
- Adams, R. B., and Nelson, A. J. (2011). "Intersecting identities and expressions: the compound nature of social perception," in *The Handbook of Social Neuroscience*, eds J. Decety and J. Cacioppo (New York, NY: Oxford University Press), 394–403.
- Adolphs, R. (2002). Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behav. Cogn. Neurosci. Rev.* 1, 21–62. doi: 10.1177/1534582302001001003
- Aviezer, H., Hassin, R. R., Bentin, S., and Trope, Y. (2008). "Putting facial expressions into context," in *First Impressions*, eds N. Ambady and J. Skowronski (New York, NY: Guilford Press), 255–286.
- Baluch, F., and Itti, L. (2011). Mechanisms of top-down attention. *Trends Neurosci.* 34, 210–224. doi: 10.1016/j.tins.2011.02.003
- Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *J. Cogn. Neurosci.* 15, 600–609. doi: 10.1162/089892903321662976
- Bar, M. (2009). The proactive brain: memory for predictions. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1235–1243. doi: 10.1098/rstb.2008.0310
- Barlassina, L., and Newen, A. (2014). The role of bodily perception in emotion: in defense of an impure somatic theory. *Philos. Phenomenol. Res.* 89, 637–678. doi: 10.1111/phpr.12041
- Barsalou, L. W. (1999). Perceptual symbol system. *Behav. Brain Sci.* 22, 637–660. doi: 10.1017/S0140525X99532147
- 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
- Block, N. (2014). Seeing-as in the light of vision science. *Philos. Phenomenol. Res.* 89, 560–572. doi: 10.1111/phpr.12135
- Bruner, J. S., and Goodman, C. C. (1947). Value and need as organizing factors in perception. *J. Abnorm. Soc. Psychol.* 42, 33–44. doi: 10.1037/h0058484
- Bruner, J. S., and Postman, L. (1949). On the perception of incongruity: a paradigm. *J. Pers.* 18, 206–223. doi: 10.1111/j.1467-6494.1949.tb01241.x
- Burge, T. (2010). *Origins of Objectivity*. New York, NY: Oxford University Press.
- Butler, A., Oruc, I., Fox, C. J., and Barton, J. J. S. (2008). Factors contributing to the adaptation aftereffects of facial expression. *Brain Res.* 1191, 116–126. doi: 10.1016/j.brainres.2007.10.101
- Carroll, J. M., and Russell, J. A. (1996). Do facial expression signal specific emotions? Judging emotions from the face in context. *J. Pers. Soc. Psychol.* 70, 205–218. doi: 10.1037/0022-3514.70.2.205
- Carroll, N. C., and Young, A. W. (2005). Priming of emotion recognition. *Q. J. Exp. Psychol.* 58A, 1173–1197. doi: 10.1080/02724980443000539
- Carruthers, P. (2015). Perceiving mental states. *Conscious. Cogn.* doi: 10.1016/j.concog.2015.04.009 [Epub ahead of print].
- Colombetti, G. (2014). *The Feeling Body: Affective Science Meets the Enactive Mind*. Cambridge, MA: The MIT Press.
- Dunbar, R. I. M. (1998). The social brain hypothesis. *Evol. Anthropol.* 6, 178–190. doi: 10.1002/(SICI)1520-6505(1998)6:5<178::AID-EVAN5>3.0.CO;2-8
- Durgin, F. H., Baird, J. A., Greenburg, M., Russell, R., Shaughnessy, K., and Waymouth, S. (2009). Who is being deceived? The experimental demands of wearing a back-pack. *Psychon. Bull. Rev.* 16, 964–969. doi: 10.3758/PBR.16.5.964
- Ekman, P., and Friesen, W. V. (1971). Constants across cultures in the face and emotion. *J. Pers. Soc. Psychol.* 17, 124–129. doi: 10.1037/h0030377
- Ekman, P., and Friesen, W. V. (1976). *Pictures of Facial Affect*. Palo Alto, CA: Consulting Psychologists Press.
- Ernst, M. O., and Bühlhoff, H. H. (2004). Merging the senses into a robust percept. *Trends Cogn. Sci.* 8, 162–169. doi: 10.1016/j.tics.2004.02.002
- Firestone, C., and Scholl, B. J. (2014). Top down effects where none should be found. *Psychol. Sci.* 25, 38–46. doi: 10.1177/0956797613485092
- Firestone, C., and Scholl, B. J. (2015). Can you experience top-down effects on perception? *Psychon. Bull. Rev.* 22, 694–700. doi: 10.3758/s13423-014-0711-5
- Fodor, J. (1984). Observation reconsidered. *Philos. Sci.* 51, 23–43. doi: 10.1086/289162
- Fodor, J. (1988). A reply to churchland's 'perceptual plasticity and theoretical neutrality'. *Philos. Sci.* 55, 188–198. doi: 10.1086/289426
- Froese, T., and Leavens, D. A. (2014). The direct perception hypothesis: perceiving the intention of another's action hinders its precise imitation. *Front. Psychol.* 5:65. doi: 10.3389/fpsyg.2014.00065
- Gallagher, S. (2008). Direct perception in the intersubjective context. *Conscious. Cogn.* 17, 535–543. doi: 10.1016/j.concog.2008.03.003
- Hadjikhani, N., Kveraga, K., Naik, P., and Ahlfors, S. P. (2009). Early (N170) activation of face-specific cortex by face-like objects. *Neuroreport* 20, 403–407. doi: 10.1097/WNR.0b013e328325a8e1
- Hassin, R., Aviezer, H., and Bentin, S. (2013). Inherently ambiguous: facial expressions of emotions, in context. *Emot. Rev.* 5, 60–65. doi: 10.1177/1754073912451331
- Haxby, J. V., and Gobbini, M. I. (2011). "Distributed neural systems for face perception," in *The Oxford Handbook of Face Perception*, eds A. J. Calder, G. Rhodes, M. H. Johnson, and J. V. Haxby (New York, NY: Oxford University Press), 93–110.
- Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends Cogn. Sci.* 4, 223–233. doi: 10.1016/S1364-6613(00)01482-0
- Jackendoff, R. (1987). *Consciousness and the Computational Mind*. Cambridge: The MIT Press.
- Keltner, D., Ekman, P., Gonzaga, G. C., and Beer, J. (2003). "Facial expression of emotion," in *Handbook of Affective Sciences*, eds R. J. Davidson, K. R. Scherer, and H. H. Goldsmith (New York, NY: Oxford University Press), 415–431.
- Krueger, J. (2012). Seeing mind in action. *Phenomenol. Cogn. Sci.* 11, 149–173. doi: 10.1007/s11097-011-9226-y
- Kveraga, K., Boshyan, J., and Bar, M. (2009). "The proactive brain: using memory-based predictions in visual recognition," in *Object Categorization: Computer and Human Vision Perspectives*, eds S. Dickinson, M. Tarr, A. Leonardis, and B. Schiele (Cambridge: Cambridge University Press), 384–400.
- Kveraga, K., Ghuman, A. S., Kassam, K. S., Aminoff, E., Hamalainen, M. S., Chaumon, M., et al. (2011). Early onset of neural synchronization in the contextual associations network. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3389–3394. doi: 10.1073/pnas.1013760108
- Lee, T. S., and Nguyen, M. (2001). Dynamics of subjective contour formation in the early visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 98, 1907–1911. doi: 10.1073/pnas.98.4.1907

- Levin, D., and Banaji, M. (2006). Distortions in the perceived lightness of faces: the role of race categories. *J. Exp. Psychol. Gen.* 135, 501–512. doi: 10.1037/0096-3445.135.4.501
- Lupyan, G. (2012). Linguistically modulated perception and cognition: the label-feedback hypothesis. *Front. Psychol.* 3:54. doi: 10.3389/fpsyg.2012.00054
- Lupyan, G. (2015). “Cognitive penetrability of perception in the age of prediction,” in *Cognitive Penetration of Perception, Review of Philosophy and Psychology*, eds S. Siegel and Z. Jenkins (forthcoming).
- MacCauley, R. N., and Henrich, J. (2006). Susceptibility to the Müller-Lyer illusion, theory neutral observation, and the diachronic cognitive penetrability of the visual input system. *Philos. Psychol.* 19, 79–101. doi: 10.1080/09515080500462347
- Macpherson, F. (2012). Cognitive penetration of colour experience. Rethinking the issue in light of an indirect mechanism. *Philos. Phenomenol. Res.* 84, 24–62. doi: 10.1111/j.1933-1592.2010.00481.x
- Macrae, C. N., and Quadflieg, S. (2010). “Perceiving people,” in *Handbook of Social Psychology*, 5th Edn, Vol. 1, eds S. T. Fiske, D. T. Gilbert, and G. Lindzey (Hoboken, NJ: John Wiley & Sons), 428–463.
- Milner, A. D., and Goodale, M. A. (1995). *The Visual Brain in Action*. New York, NY: Oxford University Press.
- Mo, L., Xui, G., Kay, P., and Tan, L. H. (2011). Evidence for the left-lateralized effect of language on preattentive categorical perception of color. *Proc. Natl. Acad. Sci. U.S.A.* 108, 14026–14030. doi: 10.1073/pnas.1111860108
- Mole, C. (2011). *Attention is Cognitive Unison*. New York, NY: Oxford University Press.
- Mole, C. (2015). “Attention and cognitive penetration,” in *The Cognitive Penetrability of Perception*, eds A. Raftopoulos and J. Zeimbekis (New York, NY: Oxford University Press), 218–238.
- Mroczo-Wąsowicz, A. (2015). “What can sensorimotor enactivism learn from studies on phenomenal adaptation in atypical perceptual conditions? A commentary on rick grush and colleagues,” in *Open MIND: 16*, eds T. Metzinger and J. M. Windt (Frankfurt am Main: MIND Group), 1–17. doi: 10.15502/9783958570283
- Newen, A. (2015). “Understanding others—the person model theory,” in *Open MIND: 26*, eds T. Metzinger and J. M. Windt (Frankfurt am Main: MIND Group), 1–28. doi: 10.15502/9783958570320
- Newen, A., Welpinghus, A., and Juckel, G. (2015). Emotion recognition as pattern recognition. The relevance of perception. *Mind Lang.* 30, 187–208. doi: 10.1111/mila.12077
- Pessoa, L. (2013). *The Cognitive-Emotional Brain: From Interactions to Integration*. Cambridge, MA: The MIT Press.
- Prinz, J. (2012). *The Conscious Brain: How Attention Engenders Experience*. New York, NY: Oxford University Press.
- Pulvermüller, F. (2003). *The Neuroscience of Language: On Brain Circuits of Words and Serial Order*. Cambridge: Cambridge University Press.
- Pulvermüller, F., and Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nat. Rev. Neurosci.* 11, 351–360. doi: 10.1038/nrn2811
- Pylyshyn, Z. W. (1984). *Computation and Cognition: Toward a Foundation for Cognitive Science*. Cambridge, MA: The MIT Press.
- Pylyshyn, Z. W. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behav. Brain Sci.* 22, 341–365. doi: 10.1017/S0140525X99002022
- Raftopoulos, A. (2014). The cognitive impenetrability of the content of early vision is a necessary and sufficient condition for purely nonconceptual content. *Philos. Psychol.* 27, 601–620. doi: 10.1080/09515089.2012.729486
- Redican, W. K. (1982). “An evolutionary perspective on human facial displays,” in *Emotion in the Human Face*, 2nd Edn, ed. P. Ekman (Cambridge, MA: Cambridge University Press), 212–280.
- Salin, P. A., and Bullier, J. (1995). Corticocortical connections in the visual system: structure and function. *Physiol. Rev.* 75, 107–154.
- Schwiedrzik, C. M., Ruff, C. C., Lazar, A. A., Leitner, F. C., Singer, W., and Melloni, L. (2014). Untangling perceptual memory: hysteresis and adaptation map into separate cortical networks. *Cereb. Cortex* 24, 1152–1164. doi: 10.1093/cercor/bhs396
- Stefanucci, J. K., and Geuss, M. N. (2009). Big people, little world: the body influences size perception. *Perception* 38, 1782–1795. doi: 10.1068/p6437
- Stokes, D. (2013). Cognitive penetrability of perception. *Philos. Comp.* 8, 646–663. doi: 10.1111/phc3.12043
- Stokes, D. (2014). Cognitive penetration and the perception of art. *Dialectica* 68, 1–34. doi: 10.1111/1746-8361.12049
- Stokes, D., and Bergeron, V. (2015). Modular architectures and informational encapsulation: a dilemma. *Eur. J. Philos. Sci.* doi: 10.1007/s13194-015-0107-z [Epub ahead of print].
- Stout, R. (2012). What someone’s behavior must be like if we are to be aware of their emotions in it. *Phenomenol. Cogn. Sci.* 11, 135–148. doi: 10.1007/s11097-011-9224-0
- Tomkins, S. S. (1962). *Affect, Imagery, Consciousness*, Vol. 1. New York, NY: Springer.
- Tomkins, S. S. (1963). *Affect, Imagery, Consciousness*, Vol. 2. New York, NY: Springer.
- Vetter, P., and Newen, A. (2014). Varieties of cognitive penetration in visual perception. *Conscious. Cogn.* 27, 62–75. doi: 10.1016/j.concog.2014.04.007
- Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., and Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proc. Natl. Acad. Sci. U.S.A.* 104, 7780–7785. doi: 10.1073/pnas.0701644104
- Witzel, C., Valkova, H., Hansen, T., and Gegenfurthner, K. (2011). Object knowledge modulates colour appearance. *Iperception* 2, 13–49. doi: 10.1068/i0396
- Zahavi, D. (2011). Empathy and direct social perception: a phenomenological proposal. *Rev. Philos. Psychol.* 2, 541–558. doi: 10.1007/s13164-011-0070-3

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.

Copyright © 2015 Marchi and Newen. 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.



Cognitive penetration and the gallery of indiscernibles

Bence Nanay^{1,2*}

¹ Centre for Philosophical Psychology, University of Antwerp, Belgium

² Peterhouse, University of Cambridge, Cambridge, UK

*Correspondence: bn206@cam.ac.uk; bence.nanay@ua.ac.be

Edited by:

Aleksandra Mroczko-Wąsowicz, National Yang-Ming University, Taiwan

Reviewed by:

Gary Lupyan, University of Wisconsin - Madison, USA

Keywords: perception, cognitive penetration, aesthetic value, indiscernibles, attention

Here is Danto's gallery of indiscernibles thought experiment (Danto, 1981, p.1)—a thought experiment that radically transformed the kind of questions aesthetics and the philosophy of art asks today. Imagine a gallery of indiscernible canvases that are all monochrome red of the same shade and of the same size. While the observable properties of all these artworks are the same, their “meaning” and aesthetic value can be very different: if one of the paintings, made by a counterrevolutionary Russian émigré is called “Red Square” and the other one is called “The Israelites crossing the Red Sea,” then these two paintings, in spite of being indistinguishable, will have very different aesthetic value. Thus, aesthetic value is only loosely (if at all) related to perception.

Let us examine this argument more closely. First, I need to introduce a bit of terminology: I call a property “aesthetically relevant” if attending to this property changes one's aesthetic evaluation (Nanay, 2015). Aesthetically relevant properties are abundant. If I attend to the arrangement of the small red patches in a Corot landscape, it can change my experience and assessment of the balance of the picture. And if I attend to the second violin in a string quartet, it can also change my experience of the entire movement. Danto's argument is supposed to establish that there is only a loose connection between aesthetically relevant properties and perception.

Danto oscillates between two arguments in his exposition of the gallery of indiscernibles, one weaker and unproblematic (and somewhat trivial), the other stronger and problematic. Here is the weaker one (I take P1 to be the painting

émigré, “Red Square” and P2 to be the painting called “The Israelites crossing the Red Sea”):

- (1*) The observable properties of P1 are the same as the observable properties of P2.
- (2*) The aesthetically relevant properties of P1 are different from the aesthetically relevant properties of P2.
-
- (3*) Aesthetically relevant properties do not supervene on observable properties.

I take it that no-one would want to deny (1*), (2*) or (3*). Nor should any of these claims strike anyone as particularly surprising or novel. Everyone but really strict formalists would accept that at least some aesthetically relevant properties do not supervene on observable properties. Danto must have meant something stronger. In fact, he did mean something stronger, namely, the following:

- (1) The perceptual experience of P1 is the same as the perceptual experience of P2.
- (2) The aesthetically relevant properties we attribute to P1 are different from the ones we attribute to P2.
-
- (3) The attribution of aesthetically relevant properties does not supervene on our perceptual experience.

Note the difference between this argument for (3) and the one for (3*) above. (3*) is about the logical relation between aesthetically relevant properties and observable properties, whereas (3) is about the logical

relation between two kinds of mental states (the attribution of aesthetically relevant properties and perceptual experiences). Very different claims (and very different arguments) indeed.

We have good reasons to hold (2). And (3) clearly follows from (1) and (2). The problem with Danto's argument for (3) is premise (1). I will argue that premise (1) is false. But, again, (1*) is true:

- (1*) The observable properties of P1 are the same as the observable properties of P2.

The observable properties of the two paintings are, by supposition, exactly the same. In fact, Danto goes further and says that even all physical properties of the two pictures are identical. But (1*) does not imply (1). Two objects may have the very same observable properties, nonetheless, one's perceptual experience of them may be very different. (1*) only implies (1) if we add a further premise, one that Danto took for granted (see esp. Danto, 2001a,b; see also Fodor, 1993):

- (1**) Perceptual experiences are not cognitively penetrable.

Danto's argument only goes through if we add this extra premise (see also Wollheim, 1993; Margolis, 1998, 2000; Lamarque, 2010; Nanay, 2015). If we block (1**), we have no reason to hold (1) and then we have no reason to hold (3). Here is the reason why blocking (1**) jeopardizes the whole argument. If perceptual experiences are cognitively penetrable (i.e., if (1**) is false), then the difference in the title of the pictures can and will bring about a perceptual difference. As a result of my

non-perceptual state of reading the title, my perceptual experience will be different. But then (1) is false, which means that there is no reason to hold (3).

And, as it turns out, there is very strong empirical evidence against (1**)—both as a general claim and as a claim in this specific context (Goldstone, 1995; Hansen et al., 2006; Lupyan and Spivey, 2008; Lupyan et al., 2010; Lupyan and Ward, 2013; Nanay, 2013a,b, but see also Firestone and Scholl, 2014 for a critical analysis). There are top-down processes that influence perceptual processing as early as the primary visual cortex (Gandhi et al., 1999) or the thalamus (O'Connor et al., 2002). Here is an old experiment, very much known at the time when Danto gave this argument (Delk and Fillenbaum, 1965): if we have to match the color of a picture of an orange heart to color samples, we match it differently (closer to the red end of the spectrum) from the way we match the color of a picture of some other, orange shapes. This shows that our recognition of the object in question (the heart) influences the color we experience it as having. In general, one's experience is not determined in a bottom-up manner by the perceptual stimulus: it depends on language, attention, the contrast classes and one's expectations (see Hansen et al., 2006; Lupyan and Ward, 2013).

The defender of Danto's claim would need to deny that we have any reasons to question (1**). The concept of "cognitive penetrability" has been severely debated in the last decades and depending on how one defines this concept (see Siegel, 2011; Macpherson, 2012 for summaries), it may not be too farfetched to retain *some* sense in which perceptual experiences are not cognitively penetrable—in which case, we can salvage (1**) and with it Danto's argument. While it may indeed be true that there may be some sense in which perceptual experiences are cognitively impenetrable (Pylyshyn, 1999), the sense of cognitive impenetrability that would be required for Danto's argument to go through is not one of these.

Here is why (Levin and Banaji, 2006): Two pictures of identical (mixed race) faces were shown to subjects—the only difference between them was that under one the subjects read the word "white" and under the other they read "black."

When they had to match the color of the face, subjects chose a significantly darker color for the face with the label "black"¹.

This experiment has the exact same structure as Danto's thought experiment: the two visual stimuli share all their observable properties—just like the two canvases. But, crucially, our experience of the two stimuli are different—we see one as being darker than the other. Similarly, when we see the painting called Red Square, our perceptual experience of the painting may be colored by our previous exposure of the soviet red flag, for example—something that is missing from our perceptual experience of the other painting. While (1*) is true, (1) is false. But if (1) is false, then we have no reason to hold (3).

The gallery of indiscernibles thought experiment is based on an empirically inadequate way of thinking about perception. On any empirically adequate ways of thinking about perception, we have no reason to take the gallery of indiscernibles seriously.

The structure of my argument was that we can bypass the thorny question of what counts as cognitive penetration because whatever sense it is in which our perception of the faces in the Levin and Banaji experiment is penetrable, it is exactly the same sense in which our perception of the artworks in Danto's thought experiment is penetrable. But it is important to highlight that this is a very weak sense of cognitive penetrability—so much so that it wouldn't even count as cognitive penetrability under many formulation of cognitive penetrability (e.g., Siegel, 2011, p. 204) because all it implies is that our perceptual experience is subject to top-down attentional

¹There has been some controversy about the Levin and Banaji (2006) findings, especially about their first experiment (e.g., Firestone and Scholl, 2014). But the experiment I want to use here is not their first but their second experiment (where two faces are identical in all respects except for the label under them). While there are some methodological issues about this experiment as well (about whether the label influences our experience or merely the matching task performed, see Lupyan, in press, footnote 4), I want to bracket these for the purposes of this discussion. If the reader is not fully convinced by this experiment, she can use some of the other, less debated empirical findings, (see Goldstone, 1995; Hansen et al., 2006; Lupyan and Spivey, 2008; Lupyan et al., 2010; Lupyan and Ward, 2013).

influences—something even those who deny the cognitive penetrability of perception would accept (see Pylyshyn, 1999).

Taking the painting to be about Russia or about the Red Sea influences what properties of the picture we are attending to. But as the inattentional blindness findings show, what properties we are attending to very much influences our perceptual phenomenology (and we also know that attention can modulate even the earliest stages of visual processing, the primary visual cortex, see Gandhi et al., 1999). But then our perceptual experience of the two paintings in the Gallery of Indiscernibles thought experiment is very different because we are attending to them very differently.

Crucially, the difference in attention brings about a difference in *perceptual* phenomenology. To see this, it may be helpful to consider the following example (e.g., Nickel, 2007; Nanay, 2010): You are looking at a 3 × 3 grid of squares against a white background. First experience: you are attending to the corner and the center squares. Second experience: you are attending to the remaining four squares. The two experiences are phenomenally different—different squares seem prominent.

In other words, different ways of attending to very simple figures of this kind changes our perceptual phenomenology. But then presumably different ways of attending to Danto's indistinguishable canvases would also make our perceptual phenomenology of these experiences different. It is important to emphasize that this phenomenal difference is a difference in perceptual phenomenology. There are some more controversial cases, like the duck-rabbit illusion, where it is also true that attention to the rabbit vs. attention to the duck very much influences our phenomenology. But in the duck-rabbit case one could object that what changes is not something perceptual: that it is the interpretation of the scene that changes (Brewer, 2007, p. 93).

In the case of the 3 × 3 grid, however, the phenomenal differences (e.g., salience) are clearly properties that are perceptually experienced (even by the thinnest accounts of perceptual experience). And we can run the same explanatory scheme for why our perceptual experience of P1 is different

from our perceptual experience of P2—in one, but not in the other, our experience of the red of the canvas is colored by the association of the red of the Soviet flag, for example. But then the difference in the attribution of aesthetically relevant properties is accompanied by a difference in our perceptual experience. Danto's argument from the Gallery of Indiscernibles fails.

We can now conclude that the attribution of aesthetically relevant properties, while it does not have to be a perceptual attribution, very much supervenes on one's perceptual experience: if there is a difference in the attribution of aesthetically relevant properties, there must also be a difference in one's perceptual experience. This restores the nice and tight connection between aesthetically relevant properties and perception: while not all aesthetically relevant properties are perceived, they all have very serious perceptual consequences.

ACKNOWLEDGMENTS

This work was supported by the EU FP7 CIG grant PCIG09-GA-2011-293818 and the FWO Odysseus grant G.0020.12N.

REFERENCES

- Brewer, B. (2007). Perception and its objects. *Philos. Stud.* 132, 87–97. doi: 10.1007/s11098-006-9051-2
- Danto, A. C. (1981). *The Transfiguration of Commonplace*. Cambridge, MA: Harvard University Press.
- Danto, A. C. (2001a). Seeing and showing. *J. Aesthetics Art. Criticism* 59, 1–9. doi: 10.1111/0021-8529.00001
- Danto, A. C. (2001b). The pigeon within us all. *J. Aesthetics Art. Criticism* 59, 39–44. doi: 10.1111/0021-8529.00005
- Delk, J. L., and Fillenbaum, S. (1965). Differences in perceived color as a function of characteristic color. *Am. J. Psychol.* 78, 290–293.
- Firestone, C., and Scholl, B. J. (2014). “Top-down” effects where none should be found: the El Greco fallacy in perception research. *Psychol. Sci.* 25, 38–46. doi: 10.1177/0956797613485092
- Fodor, J. A. (1993). “Déjà vu all over again: how Danto's aesthetics recapitulates the philosophy of mind,” in *Danto and his Critics*, ed M. Rollins (Oxford: Blackwell), 41–54.
- Gandhi, S. P., Heeger, D. J., and Boynton, G. M. (1999). Spatial attention affects brain activity in human primary visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 96, 3314–3319.
- Goldstone, R. L. (1995). Effects of categorization on color-perception. *Psychol. Sci.* 6, 298–304.
- Hansen, T., Olkkonen, M., Walter, S., and Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nat. Neurosci.* 9, 1367–1368. doi: 10.1038/nn1794
- Levin, D. T., and Banaji, M. R. (2006). Distortions in the perceived lightness of faces: the role of race categories. *J. Exp. Psychol. Gen.* 135, 501–512. doi: 10.1037/0096-3445.135.4.501
- Lamarque, P. (2010). *Work and Object*. Oxford: Oxford University Press.
- Lupyan, G. (in press). Cognitive penetrability of perception in the age of prediction: predictive systems are penetrable systems. *Rev. Philos. Psychol.*
- Lupyan, G., and Spivey, M. J. (2008). Perceptual processing is facilitated by ascribing meaning to novel stimuli. *Curr. Biol.* 18, R410–R412. doi: 10.1016/j.cub.2008.02.073
- Lupyan, G., Thompson-Schill, S. L., and Swingle, D. (2010). Conceptual penetration of visual processing. *Psychol. Sci.* 21, 682–691. doi: 10.1177/0956797610366099
- Lupyan, G., and Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proc. Natl. Acad. Sci. U.S.A.* 110, 14196–14201. doi: 10.1073/pnas.1303312110
- Macpherson, F. (2012). Cognitive penetration of colour experience. *Philos. Phenomenol. Res.* 84, 24–62. doi: 10.1111/j.1933-1592.2010.00481.x
- Margolis, J. (1998). Farewell to Danto and Goodman. *Br. J. Aesthetics* 38, 353–374.
- Margolis, J. (2000). A closer look at Danto's account of art and perception. *Br. J. Aesthetics* 40, 326–339. doi: 10.1093/bjaesthetics/40.3.326
- Nanay, B. (2010). Attention and perceptual content. *Analysis* 70, 263–270. doi: 10.1093/analysis/70.3.263
- Nanay, B. (2013a). *Between Perception and Action*. Oxford: Oxford University Press.
- Nanay, B. (2013b). “Is action-guiding vision cognitively impenetrable?,” in *Proceedings of the 35th Annual Conference of the Cognitive Science Society (CogSci 2013)* (Hillsdale, NJ: Lawrence Erlbaum), 1055–1060.
- Nanay, B. (2015). *Aesthetics as Philosophy of Perception*. Oxford: Oxford University Press.
- Nickel, B. (2007). Against intentionalism. *Philos. Stud.* 136, 279–304. doi: 10.1007/s11098-005-2013-2
- O'Connor, D. H., Fukui, M. M., Pinsk, M. A., and Kastner, S. (2002). Attention modulates responses in the human lateral geniculate nucleus. *Nat. Neurosci.* 5, 1203–1209. doi: 10.1038/nn957
- Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behav. Brain Sci.* 22, 341–423.
- Siegel, S. (2011). Cognitive penetrability and perceptual justification. *Nous* 46, 201–222. doi: 10.1111/j.1468-0068.2010.00786.x
- Wollheim, R. (1993). “Danto's gallery of indiscernibles,” in *Danto and his Critics*, ed M. Rollins (Oxford: Blackwell), 28–38.

Conflict of Interest Statement: The author declares 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: 10 September 2014; accepted: 10 December 2014; published online: 08 January 2015.

Citation: Nanay B (2015) Cognitive penetration and the gallery of indiscernibles. *Front. Psychol.* 5:1527. doi: 10.3389/fpsyg.2014.01527

This article was submitted to *Consciousness Research*, a section of the journal *Frontiers in Psychology*.

Copyright © 2015 Nanay. 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.



Do intentions for action penetrate visual experience?

Robert E. Briscoe*

Philosophy, Ohio University, Athens, OH, USA

*Correspondence: rbriscoe@gmail.com

Edited by:

Aleksandra Mroczko-Wasowicz, National Yang Ming University, Taiwan

Reviewed by:

Joshua Shepherd, Florida State University, USA

Keywords: cognitive penetration of perception, two visual systems hypothesis, intention, consciousness, motor control

A commentary on

Planning to reach for an object changes how the reacher perceives it

by Vishton, P. M., Stephens, N. J., Nelson, L. A., Morra, S. E., Brunick, K. L., and Stevens, J. A. (2007). *Psychol. Sci.* 18, 713–719. doi: 10.1111/j.1467-9280.2007.01965.x

A now-famous study by Aglioti et al. (1995) involves a graspable version of the Ebbinghaus illusion (**Figure 1**). Aglioti and colleagues constructed a 3D version of the illusion, using thin solid disks. Subjects were asked to pick up the central disk on the left if the two central disks appeared identical in size, and to pick up the central disk on the right if they appeared different in size. The experimenters varied the relative sizes of the two target disks randomly so that in some trials physically different disks appeared perceptually identical in size, while in other trials physically identical disks appeared perceptually different in size. In selecting a disk in either trial condition, Milner and Goodale observe,

“subjects indicated their susceptibility to the visual illusion” (1995/2006, p. 168): that is, their *choice* of which disk to pick up was determined by its apparent size rather than its real one. Nonetheless, the effect of the illusion was significantly less pronounced with respect to action, as measured by maximum grip aperture (MGA) in prehension, than with respect to conscious perceptual estimation (PE), as measured by the distance between thumb and forefinger in a manual estimate of disk size. Although the disk surrounded by small circles in the illusion display typically *looks* about 10% larger than the disk surrounded by large circles, the increase in MGA when reaching for the former disk exhibited a magnitude of around only 6%.

According to proponents of the dual systems model of visual processing (Milner and Goodale, 1995/2006), the illusion has a different effect on visual awareness than on visually guided grasping because the former makes use of different sources of visuospatial information than the latter. On this model, how the size of an object

appears in conscious vision should not influence grip aperture, and, conversely, how the size of the object is represented by motor systems that guide grasping should not influence representation of its size in conscious vision.

At variance with this idea, however, Vishton et al. (2007) (experiment 3) found that the act of reaching for a disk in a 3D version of Ebbinghaus illusion significantly diminished the magnitude of the effect on subsequent PE for several minutes after reaching trials had ended (5.74% for PE vs. 6.10% for grasping). Strikingly, they also found (experiment 2) that when subjects were *merely informed* prior to engaging in PE trials that they would subsequently be required to grasp the disk that appeared larger, the effect of the illusion on PE was significantly diminished (6.18% for PE vs. 5.54% for grasping). “Simply listening to a description of a reaching task,” Vishton and co-authors write, “seems to affect size perception” (Vishton et al., 2007, p. 718).

These findings suggest that the phenomenal contents of visual experience can be cognitively penetrated: high-level information originating outside of the visual system seems to modulate the way an object’s size visually appears. There are different possible mechanisms whereby such penetration might occur. Vishton and co-authors propose that “intending to reach for a target changes how the reacher perceives it” and that “action choice changes the nature of visual size perception” (p. 718). But how does action selection have this effect? One possibility (a) is that an abstract, high-level intention to act—either a “distal” or “proximal” intention in the sense of Pacherie (2008)—somehow exerts a direct influence on PE, say, by

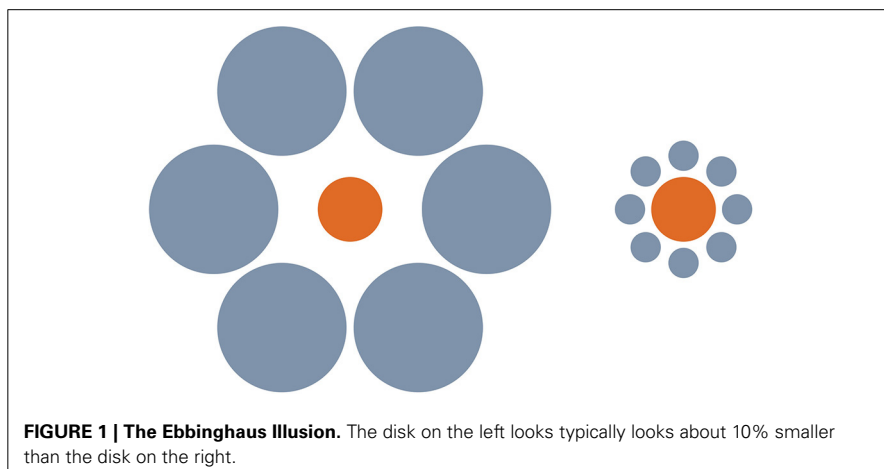


FIGURE 1 | The Ebbinghaus Illusion. The disk on the left looks typically looks about 10% smaller than the disk on the right.

changing the relative weightings assigned by the visual system to sources of depth information such as binocular disparity, vergence, accommodation, and relative size. Since size estimation depends, in part, on perceived distance in depth, this could explain the influence of intention on perception. A second possibility (b) is that the relevant effect is brought about via lower-level motor representations that implement and provide kinematic and dynamical specification for the subject's high-level intention. This would arguably still count as a case of cognitive penetration if the lower-level, action-specifying motor representations carried information from the subject's high-level intention that influenced relative cue weighting or other visual computations. As Wu (2013) writes, "The key [to cognitive penetration of vision by intention] is not directness of link but (internal) informational transfer of an appropriate kind" (p. 662). A third possibility (c) looks to motor imagery elicited in the course of both experiments for the source of penetration. Possibility (c), however, is not entirely distinct from (a) and (b), since there is evidence that internally rehearsing the performance of an action activates representations at all

levels in the motor processing hierarchy (for reviews, see Decety and Grèzes, 2006; Jeannerod, 2006). A final possibility (d) is that the effect is not due to motor representations at all, but rather to the subject's *beliefs* concerning the action that she has been requested to perform.¹ Future studies will have to investigate which, if any, of these four explanations best accounts for the intriguing effects that Vishton and his co-authors have reported.

ACKNOWLEDGMENTS

I am grateful to the reviewer for comments that resulted in significant improvements to this commentary.

REFERENCES

- Aglioti, S., DeSouza, J., 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
- Decety, J., and Grèzes, J. (2006). The power of simulation: imagining one's own and other's behavior. *Brain Res.* 1079, 4–14. doi: 10.1016/j.brainres.2005.12.115
- Jeannerod, M. (2006). *Motor Cognition: What Actions Tell the Self*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780198569657.001.0001
- Milner, A. D., and Goodale, M. A. (1995/2006). *The Visual Brain in Action, 2nd Edn.* Oxford: Oxford University Press.
- Pacherie, E. (2008). The phenomenology of action: a conceptual framework. *Cognition* 107, 179–217. doi: 10.1016/j.cognition.2007.09.003
- Vishton, P. M., Stephens, N. J., Nelson, L. A., Morra, S. E., Brunick, K. L., and Stevens, J. A. (2007). Planning to reach for an object changes how the reacher perceives it. *Psychol. Sci.* 18, 713–719. doi: 10.1111/j.1467-9280.2007.01965.x
- Wu, W. (2013). Visual spatial constancy and modularity: does intention penetrate vision? *Philos. Stud.* 165, 647–669. doi: 10.1007/s11098-012-9971-y
- Conflict of Interest Statement:** The author declares 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: 24 September 2014; accepted: 19 October 2014; published online: 05 November 2014.

Citation: Briscoe RE (2014) Do intentions for action penetrate visual experience? *Front. Psychol.* 5:1265. doi: 10.3389/fpsyg.2014.01265

This article was submitted to *Consciousness Research*, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Briscoe. 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.

¹ I am grateful to the reviewer for suggesting this possibility.



Consciousness doesn't overflow cognition

Richard Brown *

Philosophy Program, LaGuardia Community College, City University of New York, Long Island City, NY, USA

*Correspondence: onemorebrown@gmail.com

Edited by:

Aleksandra Mroczko-Wasowicz, National Yang Ming University, Taiwan

Reviewed by:

Lucia Melloni, Max Planck Institute for Brain Research, Germany

Keywords: phenomenological overflow, higher-order thought theory, consciousness, partial report, fragile visual short-term memory

Theories of consciousness can be separated into those that see it as cognitive in nature, or as an aspect of cognitive functioning, and those that see consciousness as importantly distinct from any kind of cognitive functioning. One version of the former kind of theory is the higher-order-thought theory of consciousness. This family of theories posits a fundamental role for cognitive states, higher-order thought-like intentional states, in the explanation of conscious experience. These states are higher-order in that they represent the subject herself as being in various world-directed first-order states and thus constitute a kind of cognitive access to one's own mental life. This distinctive cognitive access is postulated to account for what it is like for one to have a conscious experience.

One important challenge to this approach is Block's case for phenomenological overflow (Block, 2007, 2011, 2012). The basic argument is that, overall, the balance of evidence favors the identification of phenomenal consciousness with first-order non-cognitive states rather than our cognitive access to those states. Emerging clearly from the ensuing debate is that Block's argument is meant to establish that phenomenology overflows *working memory*. This is important because, unlike other theories, the higher-order thought theory can allow that our conscious experience overflows working memory. In addition, it can account for the subjective impression that there is overflow even if there isn't.

Take the so-called Amsterdam paradigm (Sligte et al., 2008), which builds on Sperling's (1960) partial report paradigm. In these experiments, subjects are presented with a

change-blindness-type scenario. For instance, they might be presented with a clock-like formation of rectangles. One array is presented followed by a variable interval and a second array, which may or may not contain a rectangle that had changed its orientation. Subjects are cued to the location of the potential change at various points during this process and then asked at the end if anything changed. Sligte et al. distinguish between what they call the "visual icon," which is a highly detailed but brief positive afterimage occurring shortly after stimulus presentation, and what they call "fragile short-term memory," which is less detailed but long-lasting. Subjects are able to perform the task successfully even when cued up to 6 s after the original presentation of the stimulus.

Block argues, largely on the basis on informal reports by subjects, that the best way to explain these findings is by positing a richly detailed phenomenally conscious experience of all of the shapes, rather than a sparsely detailed conscious experience corresponding to what is represented in working memory. Because the higher-order thought theory does not make the claim that encoding in working memory is required for conscious experience the theory could in principle accept this claim. The higher-order thought theory can allow that our phenomenal consciousness (that is, the contents of the relevant higher-order thoughts) overflows working memory. The relevant higher-order thoughts will be as detailed as the stream of consciousness, which, however sparse that is, will still be more detailed than what is encoded in working memory. What it cannot allow is that there is phenomenal consciousness in the absence of suitable

higher-order thoughts instantiating a kind of cognitive access to the first-order states.

On the other end of the theoretical spectrum is the claim that only what is in working memory is phenomenally conscious and subjects are mistaken about the detail of their conscious experience. If so, then the conscious experience of subjects in the Amsterdam paradigm is to some degree generic, partial, fragmented, or degraded. The reports of "reading the answers off of conscious experience" may, to some extent, be confabulated. Subjects can do the task, they have the impression that they saw all of the rectangles, and they give a commonsense explanation. If this is the case then the higher-order theory will account for this by positing correspondingly fragmented, generic, or partial contents of the relevant higher-order states.

So at this point there may or may not be phenomenal consciousness that overflows working memory, but whatever the conscious experience of subjects in these experiments turns out to be we can explain it on the higher-order thought theory. This is because the higher-order thought theory makes the general claim that people may be aware of first-order states in virtue of some of the state's properties (that they are letters, that they are blocks, that they are arranged in various ways, that this particular block is oriented in that particular orientation, etc.), but not necessarily in virtue of all of their properties. Nonetheless, the information that the first-order states encode is causally efficacious. Higher-order-thought theories maintain that the *information* that is represented by the first-order state is partially unconscious, not that the first-order state itself is unconscious.

There is some evidence for this interpretation of the data from other work on change blindness. When subjects are not consciously aware of the difference between the two stimuli, both the original shape and the changed shape show priming effects; when the difference is consciously perceived only the changed stimulus shows those priming effects (Silverman and Mack, 2006). In both cases, subjects are aware of both stimuli, but being conscious of the difference between the two makes a difference in mental functioning. In addition, there is some evidence that subjects can detect such changes unconsciously (Fernandez-Duque and Thornton, 2000; Laloyaux et al., 2006).

Block interprets these claims as committing one to robust long-lasting unconscious working memory and he argues that the evidence currently doesn't support that hypothesis. For instance, he cites work by Soto et al. (2011) that suggests that unconscious working memory doesn't have the required capacity to explain the Amsterdam results. In this study, experimenters used masking to render a stimulus close to or below threshold and then asked subjects to compare a grating to a highly visible one presented up to 5 s later. Subjects were able to do it, but at a rate that is far below what subjects in the Amsterdam paradigm are capable of. This was the case even though the Soto task was much easier than that in the Amsterdam paradigm. In response to the criticism that the stimuli in the Soto experiments were masked, Block cites Carmel et al. (2011) which suggests that unconscious representations are short lived and so would not last the up to 6 s we find in the Amsterdam paradigm. Together these results suggest that unconscious working memory is not robust enough to explain the Amsterdam results.

Block raises a legitimate worry for those theories that do appeal to working memory, but it would be a mistake to lump the higher-order theory into that camp. As we have seen above, the higher-order-thought theory does not rely on unconscious states, but rather on some aspects of the targeted first-order states not being represented in the higher-order thought. We are aware of being in the states, and so they are conscious, but not in respect of all of their

properties. By analogy compare what happens when I see a cardboard box, say, through a window but because the window is dirty I cannot make out what the box has written on it. I am aware of the box but not of all of its properties.

Another way to make the point is by stipulating a distinction between *phenomenal consciousness*—or there being something that there is like for the creature in question—and *state consciousness*—or being the target of a suitable higher-order representation (Brown, 2012, 2014). In the partial-report paradigm, the higher-order-thought theory claims that the first-order states, which are in fact the targets of the relevant higher-order representations, are state-conscious while the phenomenal consciousness of the subject is determined by the higher-order thought. In the Soto and Carmel et al. work, the relevant stimuli were all state-unconscious and so do not address the claim made by the higher-order thought theory. Subjects in the Amsterdam paradigm are maintaining a phenomenally conscious visual experience; most parties agree on that and even if one doesn't we can allow it for the sake of argument. What the higher-order theorist insists on is that this phenomenally conscious visual experience, which is determined by the content of the higher-order thought, may diverge from the informational content of the first-order states that are represented by the relevant higher-order states.

Block also appeals to work from Sligte et al. (2009), which found activity in V4 but not V1. This, he suggests, is not what we would expect if these representations were unconscious. However, it should now be clear that higher-order theory could allow that these states in V4 may be state-conscious. If these states are actually the first-order representations of the stimuli, then they are the targets of the higher-order cognitive access. The higher-order-thought theory claims that this cognitive access consists in thought-like states that result in one being aware of oneself as being in the relevant first-order states and since that cognitive access determines what it is like for you, what it is like for you will be relatively impoverished compared to "how it could have been," so to speak, if more of the mental information carried by the first-order states was represented in the

higher-order thought. For instance if one has a maximally determinate first-order representation of a grid and one's higher-order thoughts represent one as seeing only part of the grid then this is what it will be like for you. On the other hand if one is having a rich conscious experience as of the grid this will be because of the richness of the content of the relevant higher-order states. But in both cases the very same first-order states are state-conscious.

Thus, regardless of how the phenomenology of subjects turns out, the higher-order thought theory is well-situated to account for it. In fact, positing non-cognitive phenomenal consciousness itself comes with a high theoretical cost. Phenomenal consciousness consists in there being something that it is like for the subject of the experience and this suggests that there must be some kind of access to the experience, some kind of awareness of the experience as being one's own. Block has elsewhere argued that some non-cognitive form of awareness can account for this (Block, 2007), but no account of non-cognitive access to date can explain the subjective appearances.

Block does suggest a possible form of non-cognitive access. Following Sosa (2002) he offers a deflationary account on which we are aware of our mental states just in the having of them: just as we smile our own smiles just by smiling, so too we may experience our own experiences just by experiencing. When I feel a pain, not only do I experience the painful quality but I also experience it as mine. This is not the case when these states are unconscious. How can the deflationary account handle this? How is the mere having of one of these states different from the mere having of the other? While perhaps not an insurmountable problem, this is a formable obstacle to any non-cognitive account of awareness. Block also suggests the possibility of some kind of self-representational account. But there is no way to make sense of any such view except in higher-order terms. Block hasn't offered an alternative, but just appealed to there being one.

Some doubt that we can decide this issue in a theory neutral way (Kouider et al., 2012; Overgaard and Grunbaum, 2012) while others suggest

that non-cognitive consciousness is somehow unscientific (Cohen and Dennett, 2011). I agree with Block (2012) that these views are mistaken. While it seems clear that we will never know with absolute certainty whether cognition plays a role in consciousness, we need not aspire to that unreachable goal. We should ask whether, within the confines of scientifically acceptable standards of evidence, the balance of available evidence favors one theory or another. I have been arguing that the higher-order thought theory is in a position to provide a more parsimonious “mesh” between psychology and neuroscience (Block, 2007; Lau and Brown, in press) but it also makes testable predictions.

If phenomenal consciousness depends in some way on higher-order cognitive functioning then we should be able to alter the conscious experience of subjects by interfering with areas of the brain thought to be involved in higher-order cognition while simultaneously leaving first-order processing unchanged or alternatively to produce conscious experience by directly stimulating the relevant areas (Weisberg, 2011). We might also expect that we could find cases where conscious experience outstrips first-order activity and that we would be able to “read-out” or “decode” this from activity in higher-order areas. In extreme conditions we would expect that we might find conscious experience in the absence of first-order sensory activity altogether. More work needs to be done but early attempts at testing these predictions have given suggestive results (Lau and Rosenthal, 2011; Lau and Brown, in press).

The higher-order thought theory of consciousness remains a reasonable working hypothesis with a slight edge against competing accounts and a robust research program to pursue.

ACKNOWLEDGMENTS

Thanks to Jake Berger, Ned Block, and David Rosenthal, and the reviewer for this journal, for very helpful discussion of the issues and comments on earlier drafts.

REFERENCES

- Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behav. Brain Sci.* 30, 481–548. doi: 10.1017/S0140525X07002786
- Block, N. (2011). Perceptual consciousness overflows cognitive access. *Trends Cogn. Sci.* 12, 567–575. doi: 10.1016/j.tics.2011.11.001
- Block, N. (2012). Response to Kouider et al.: which view is better supported by the evidence? *Trends Cogn. Sci.* 16, 141–142. doi: 10.1016/j.tics.2012.01.008
- Brown, R. (2012). “The brain and its states,” in *Being in Time: Dynamical Models of Phenomenal Experience*, eds S. Edelman, T. Fekete, and N. Zach (Philadelphia, PA; Amsterdam: John Benjamins), 211–238. doi: 10.1075/aicr.88.10bro
- Brown, R. (2014). The HOROR theory of phenomenal consciousness. *Philos. Stud.* doi: 10.1007/s11098-014-0388-7
- Carmel, D. R., Candace, P., Elizabeth, A., and Carrasco, M. (2011). Fast unconscious fear processing. *J. Vis.* 11, 314.
- Cohen, M. A., and Dennett, D. C. (2011). Consciousness cannot be separated from function. *Trends Cogn. Sci.* 15, 358–364. doi: 10.1016/j.tics.2011.06.008
- Fernandez-Duque, D., and Thornton, I. M. (2000). Change detection without awareness: do explicit reports underestimate the representation of change in the visual system? *Vis. Cogn.* 7, 324–344. doi: 10.1080/135062800394838
- Laloyaux, C., Destrebecqz, A., and Cleeremans, A. (2006). Implicit change identification: a replication of Fernandez-Duque and Thornton (2003). *J. Exp. Psychol. Hum. Percept. Perform.* 32, 1366–1379. doi: 10.1037/0096-1523.32.6.1366
- Lau, H., and Rosenthal, D. (2011). Empirical support for higher-order theories of consciousness. *Trends Cogn. Sci.* 15, 365–373. doi: 10.1016/j.tics.2011.05.009
- Lau, H., and Brown, R. (in press). “The emperor’s new phenomenology? The empirical case for conscious experience without first-order representations,” in *Themes from Block*, eds A. Pautz and D. Stoljar (MIT).
- Overgaard, M., and Grunbaum, T. (2012). Cognitive and non-cognitive conceptions of consciousness. *Trends Cogn. Sci.* 16, 137. doi: 10.1016/j.tics.2011.12.006
- Kouider, S., Sackar, J., and de Gardelle, V. (2012). Do we still need phenomenal consciousness? Comment on block. *Trends Cogn. Sci.* 16, 140–141. doi: 10.1016/j.tics.2012.01.003
- Silverman, M. E., and Mack, A. (2006). Change blindness and priming: when it does and does not occur. *Conscious. Cogn.* 15, 409–422. doi: 10.1016/j.concog.2005.08.003
- Sligte, I. G., Scholte, H. S., and Lamme V. A. F. (2008). Are there multiple visual short-term memory stores? *PLoS ONE* 3:e1699. doi: 10.1371/journal.pone.0001699
- Sligte, I. G., Scholte, H. S., and Lamme V. A. F. (2009). V4 activity predicts the strength of visual short-term memory representations. *J. Neurosci.* 29, 7432–7438. doi: 10.1523/JNEUROSCI.0784-09.2009
- Sosa, E. (2002). “Privileged access,” in *Consciousness: New Philosophical Perspectives*, eds Q. Smith and A. Jokic (Oxford: Oxford University Press), 273–294.
- Soto, D., Mäntylä T, and Silvanto J. (2011). Working memory without consciousness. *Curr. Biol.* 21, 912–913. doi: 10.1016/j.cub.2011.09.049
- Sperling, G. (1960). The information available in brief visual presentations. *Psychol. Monogr.* 74, 1–29. doi: 10.1037/h0093759
- Weisberg, J. (2011). Abusing the notion of what-it-is-likeness. *Analysis* 71, 438–443. doi: 10.1093/analys/anr040

Conflict of Interest Statement: The author declares 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: 23 September 2014; accepted: 16 November 2014; published online: 04 December 2014.

Citation: Brown R (2014) Consciousness doesn't overflow cognition. *Front. Psychol.* 5:1399. doi: 10.3389/fpsyg.2014.01399

This article was submitted to *Consciousness Research*, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Brown. 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.

It does belong together: cross-modal correspondences influence cross-modal integration during perceptual learning

Lionel Brunel^{1*}, Paulo F. Carvalho² and Robert L. Goldstone²

¹ Laboratoire Epsilon, Department of Psychology, Université Paul-Valéry Montpellier III, Montpellier, France, ² Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA

OPEN ACCESS

Edited by:

Aleksandra Mroczko-Wasowicz,
National Yang-Ming University, Taiwan

Reviewed by:

Jim Parkinson,
Sackler Centre for Consciousness
Science, UK

Sharon Zmigrod,
Leiden University, Netherlands

*Correspondence:

Lionel Brunel,
Laboratoire Epsilon, Department
of Psychology, Université Paul-Valéry
Montpellier III, EA 4556, 4 Boulevard
Henri IV, 34000 Montpellier, France
lionel.brunel@univ-montp3.fr

Specialty section:

This article was submitted to
Consciousness Research, a section
of the journal *Frontiers in Psychology*

Received: 30 September 2014

Accepted: 14 March 2015

Published: 09 April 2015

Citation:

Brunel L, Carvalho PF and Goldstone
RL (2015) It does belong together:
cross-modal correspondences
influence cross-modal integration
during perceptual learning.
Front. Psychol. 6:358.
doi: 10.3389/fpsyg.2015.00358

Experiencing a stimulus in one sensory modality is often associated with an experience in another sensory modality. For instance, seeing a lemon might produce a sensation of sourness. This might indicate some kind of cross-modal correspondence between vision and gustation. The aim of the current study was to explore whether such cross-modal correspondences influence cross-modal integration during perceptual learning. To that end, we conducted two experiments. Using a speeded classification task, Experiment 1 established a cross-modal correspondence between visual lightness and the frequency of an auditory tone. Using a short-term priming procedure, Experiment 2 showed that manipulation of such cross-modal correspondences led to the creation of a crossmodal unit regardless of the nature of the correspondence (i.e., congruent, Experiment 2a or incongruent, Experiment 2b). However, a comparison of priming effects sizes suggested that cross-modal correspondences modulate cross-modal integration during learning, leading to new learned units that have different stability over time. We discuss the implications of our results for the relation between cross-modal correspondence and perceptual learning in the context of a Bayesian explanation of cross-modal correspondences.

Keywords: brightness–lightness, pitch, cross-modal integration, cross-modal correspondence, perceptual learning

Introduction

Perception allows us to interact with and learn from our environment. It allows us to transform internal or external inputs into representations that we can later on recognize, and it also lets us make connections between situations that we have encountered (see Goldstone et al., 2013). In other words, perception can be envisaged as an interface between a cognitive agent and its environment. However, our environment is complex and instable. Processing a situation may require integrating information from all of our senses as well as background contextual knowledge in order to reduce the complexity and the instability of the situation. In that case, what we call a “conscious experience” of a situation should involve an integration of both a particular state of the cognitive system generated by the current situation (i.e., perceptual state) and former cognitive states (i.e., memory state). Accordingly, integration should be a relevant mechanism for both perceptual and memory processes (see Brunel et al., 2009). In this article, cross-modal perceptual

phenomena (e.g., cross-modal correspondence) are employed as an effective way to further investigate this integration mechanism and its connection with perception and memory processes.

It is now well established that cross-modal situations influence cognitive processing. For instance, people are generally better at identifying (e.g., MacLeod and Summerfield, 1990), detecting, (e.g., Stein and Meredith, 1993), categorizing (e.g., Chen and Spence, 2010), and recognizing (Molholm et al., 2002) multi-sensory events compared to unisensory ones. This multisensory advantage takes place regardless of whether the sensory signals are redundant or not (see Teder-Sälejärvi et al., 2005; Laurienti et al., 2006). More interestingly, it also seems that people spontaneously associate sensory components from different modalities together in a particular, fairly consistent, way. For instance, the large majority of people agree that “Bouba” refers to a rounded shape while “Kiki” refers to an angular one (Ramachandran and Hubbard, 2001). Evidence like this shows a non-arbitrary relation between a shape and a word sound (i.e., a cross-modal correspondence, see Spence, 2011). These correspondences between sensory modalities have a direct influence on online cognitive activity.

Cross-modal correspondences modulate performance in cognitive tasks. For instance, in a speeded classification task¹, participants are faster at identifying the size of a stimulus when it is accompanied by a congruent tone (e.g., a small circle presented with a high-pitched tone; Gallace and Spence, 2006; Evans and Treisman, 2010) rather than an incongruent tone. Similarly, in a temporal order judgment² task, participants perceive congruent asynchronous stimuli (e.g., a small circle presented with a high-pitched tone) as more synchronous than incongruent stimuli (e.g., a large circle presented with a high-pitched tone; Parise and Spence, 2009). In both examples, a particular relation is defined as congruent when the features share the same directional value (e.g., large size and low-pitched sound) and incongruent when the opposite mapping is used (e.g., small size and high-pitched sound). Directional value is a psychologically salient quality because many perceptual dimensions fall on a continuum with psychologically smaller and larger ends (Smith and Sera, 1992). Larger, louder, and lower pitched values are all perceived as having greater magnitudes than their opposing smaller, quieter, and higher pitched values. Using both speeded and non-speeded measures, this magnitude-based congruency effect has been observed between apparently highly distinct features, such as brightness/lightness and pitch (Marks, 1987, see also Marks, 2004), size and pitch (Gallace and Spence, 2006; Evans and Treisman, 2010), and spatial position and pitch (Evans and Treisman, 2010).

The existence of cross-modal correspondences contributes to our understanding of perceptual processes. Historically, perception has been conceived as a modularized set of systems relatively independent of each another (e.g., Fodor, 1983). However, the

existence of a correspondence (within or between sensory modalities) indicates that perceptual components are integrated during perceptual processing. Indeed, Parise and Spence (2009) propose that correspondences affect cross-modal integration directly. Thus, congruent stimuli form a stronger integration than incongruent ones and, as a consequence, produce a more robust impression of synchrony. In other words, the perception of a cross-modal object requires not only multiple activations in sensory areas but also the synchronization and integration of these activations. In that case, features sharing the same directional value produce a stronger coupling between the different unimodal sensory signals and are therefore more robustly integrated together (see also, Evans and Treisman, 2010).

Does the fact that cross-modal integration is stronger with features sharing the same directional value mean that cross-modal integration should not be observed with other relations between features? An impressive amount of behavioral (Brunel et al., 2009, 2010, 2013; Zmigrod and Hommel, 2010, 2011, 2013; Rey et al., 2014, 2015) and brain imagery (see Calvert et al., 1997; Giard and Peronnet, 1999; King and Calvert, 2001; Teder-Sälejärvi et al., 2002, 2005) studies provide evidence of cross-modal integration between unrelated features. For instance, Brunel et al. (2009, 2010, 2013) showed that exposing participants to an association between two perceptual features (e.g., a square and a white-noise sound) results in these features being integrated within a single memory trace (or event, see Zmigrod and Hommel, 2013). Once two features have become integrated, the presence of one feature automatically suggests the presence of the other. In this view, integration is a fundamental mechanism of perceptual learning (see also, unitization; Goldstone, 2000) or contingency learning (see Schmidt et al., 2010; Schmidt and De Houwer, 2012).

If this kind of integration mechanism is involved in perceptual learning and cross-modal correspondences modulate integration, cross-modal correspondences might be expected to modulate cross-modal integration during perceptual learning. In the present work we test this hypothesis across two experiments.

The first experiment was designed in order to test an established cross-modal congruency effect between visual lightness and auditory frequency (see Marks, 1987; Klapetek et al., 2012). To do so, we used a speeded classification task in which participants had to discriminate bimodal stimuli (i.e., audiovisual) according either to the lightness of the visual shape or frequency of the auditory tone. We manipulated the relation between the stimuli's features so that half of them were congruent (i.e., light-gray + high-pitched tone or dark-gray + low-pitched tone) and the other half was incongruent (i.e., the opposite stimuli mapping). Following Marks (1987), we predicted that, irrespective of the task, we should observe an interaction between visual lightness and auditory frequency. Observing such an interaction would indicate cross-modal correspondence between those two dimensions.

Having established this cross-modal correspondence, in the second experiment we test our hypothesis that cross-modal correspondences should modulate cross-modal integration during perceptual learning. To do so, we used a paradigm derived from

¹In speeded classification tasks participants have to discriminate one component of the stimulus as fast as possible while trying to ignore any other characteristics (see Marks, 2004).

²In temporal order judgment task, participants have to make an unspeeded response on order relation in a trial sequence.

our previous work on cross-modal integration (see Brunel et al., 2009, 2010, 2013). Our paradigm employs two distinct phases. Participants first implicitly learned that a given shape (e.g., a square) was systematically presented with a sound, while another shape (e.g., a circle) was presented without any sound. Then, participants had to perform a tone-discrimination task according to pitch (i.e., low-pitched or high-pitched) in which each tone (i.e., the auditory target/target-tone) was preceded by one of the geometrical shapes previously seen during the implicit learning phase (i.e., visual prime shape). During learning, we showed (see Brunel et al., 2009, 2010, 2013) that participants integrated the visual shape and the auditory tone within a single memory trace and as a consequence the visual prime shape was able to influence the processing of the target tone. In order to avoid a conceptual or symbolic interpretation of our priming effect (i.e., “square” = “sound”), a manipulation of the stimulus onset asynchrony (SOA) during the second phase was introduced. Previous studies (see Brunel et al., 2009, 2010) have found a modulation of the priming effect depending the level of SOA. Interference was observed when the SOA between the visual prime and the tone target was shorter than the duration of the sound associated with the shape during the learning phase. In this case, there was a temporal overlap between reactivation induced by the prime and tone processing (see Brunel et al., 2009, 2010). Facilitation was observed when the SOA was equal or longer than the duration of the sound associated with the shape during the learning phase. In this latter case, no temporal overlap occurred between simulation of the learned associated sound and target-tone processing so that target-tone processing took advantage of the auditory preactivation induced by the prime (see Brunel et al., 2009, 2010). This succession of interference followed by facilitation indicates that the shape-sound form a perceptual unit that was integrated during learning (see also Brunel et al., 2009 Experiments 2a,b and 3) otherwise we might have observed only a facilitation irrespective the SOA.

Basically, our second experiment used the same general design. However, we introduced a manipulation of the cross-modal correspondence during learning. In Experiment 2a, participants had to learn bimodal congruent stimuli (i.e., either a dark-gray + low-pitched or light-gray + high-pitched) whereas, in the Experiment 2b, participants had to learn bimodal incongruent stimuli (i.e., either a light-gray + low-pitched or dark-gray + high-pitched). This manipulation of cross-modal correspondences during learning helps us directly test an influence of cross-modal correspondence on cross-modal integration during perceptual learning. The manipulation of the congruency of stimuli might be expected to lead to the creation of perceptual units either more or less stable over time. Experiments 2a,b are crucial to test this idea.

First, if learning cross-modal congruent stimuli is at least equally strong as learning seemingly unrelated cross-modal stimuli, we might expect a replication of our previous findings (see Brunel et al., 2009, 2010) in Experiment 2a. That is to say, we should observe an interference effect for SOAs shorter than the duration of the tone at learning (i.e., slower target discrimination when the prime target relation matches, rather than mismatches, the association seen during learning) and a facilitation for SOAs

equal to the duration of the tone at learning (i.e., faster target discrimination when the prime target relation matches rather than mismatches the association seen during learning). This result would indicate that participants learned new perceptual units which integrate both perceptual components. Indeed, if such a unit is not created during learning we would only observe a replication of Experiment 1 results in Experiment 2a. That is to say, we should find an interaction between visual lightness and auditory frequency irrespective the manipulation of the SOA.

Then, with Experiment 2b, we might expect two different possibilities. First, learning incongruent stimuli might disrupt the integration mechanism so that we would not observe the same pattern of results as in Experiment 2a. One could predict no priming effect (either interference or facilitation) if there was no integration between the visual and the auditory components during learning. In that case, one might expect a replication of Experiment 1's results. Alternatively, learning incongruent stimuli might interfere with the integration mechanism. That is to say, integration might still occur but could be weaker than in Experiment 2a. In that case, one would predict the replication of the pattern of results seen in Experiment 2a, but the priming effect (irrespective of the nature of this effect: interference or facilitation) should be less reliable in Experiment 2b compared to Experiment 2a.

Experiment 1

Method

Participants

Twenty undergraduate students from Indiana University volunteered to participate in exchange for course credit. Participants' consent was obtained for all participants in compliance with the IRB of Indiana University. All of the participants reported no corrected or uncorrected hearing impairment. All of the participants had normal or corrected to normal visual acuity.

Stimuli and Material

The auditory stimuli, generated using Audacity (Free Software Foundation, Boston), were pure tones with a fundamental frequency of 440 Hz (i.e., low-pitched tone) or 523 Hz (i.e., high-pitched tone). Auditory signals were amplified through Sennheiser (electronic GmbH & Co, Wedemark Wennebostel) headphones with an intensity level of ~75 Db. The visual stimuli were geometric shapes (a 7 cm square and a circle of 3.66 cm radius) that could be displayed in two different shades of gray (CIE $L^*a^*b^3$ setting value in brackets): dark gray (L: 27.96 a: 0.00, b: 0.00), or light gray (L: 85.26, a: 0.00, b: 0.00). Across the different experimental conditions, the shape could be light or dark and the background was set at mid-gray (L: 56.3, a: 0.00, b: 0.00).

All of the experiments were conducted on a Macintosh micro-computer (iMac, Apple inc., Cupertino, CA, USA). Psyscope software X B57 (Cohen et al., 1993) was used to create and manage the experiment.

³L used to refer as the perceived luminance of the eye whereas *a (green to red) and *b (blue to yellow) refer to the chroma.

Procedure

After filling out a written consent form, each participant was tested individually in a darkened room during experimental sessions lasting approximately 45 min. The procedure can be understood as a speeded classification task (see Marks, 1987). On each trial, the participant received a composite stimulus (a particular sound + light combination presented simultaneously for 500 ms), one component of the stimulus was accessory and the other was critical. Depending on the trial, participants had to judge either the lightness (i.e., dark versus light) or the auditory frequency (i.e., low-pitched vs. high-pitched) of the stimulus. At the beginning of each trial, participants received a visual warning signal (presented 1000 ms on the screen) indicating which task they had to perform on the upcoming stimulus.

Participants completed a total of 387 trials divided in three blocks. For each trial, they had to indicate their response by pressing the appropriate response key on a QWERTY keyboard. The stimulus-response mapping was counterbalanced between participants whereas the other combinations between our manipulations were randomly counterbalanced within participants.

Results and Discussion

The mean correct response latencies (RTs) and mean percentages of correct responses (CRs) were calculated across participants for each experimental condition. RTs that deviated from the mean more or less than 2 SDs were removed (this same cut-off was used throughout all of the experiments and never led to exclusion of more than 3.5% of the data).

Separate repeated measures analyses of variance were performed on latencies RT and CRs with subject as a random variable, and Modality (Visual vs. Auditory), Tone Frequency (Low-Pitched vs. High-Pitched), and Lightness (Light vs. Dark) as within-subject variables. For clarity, we report here only the analysis regarding the RTs. The results for CR are comparable to those observed for RTs. There was no evidence of a speed-accuracy trade-off – a significant congruency effect (faster RTs for bimodal congruent than incongruent) was always associated with either a significantly lower error rate for congruent pairs or no statistically significant difference.

RT Results

As expected, our analysis revealed a reliable significant interaction between the Tone's Frequency and the Shape's Lightness, $F(1,19) = 7.03$, $p < 0.05$, $\eta_p^2 = 0.27$ (see **Figure 1**).

Regardless of the sensory modality of the task, participants were faster to discriminate congruent stimuli (i.e., low-pitched + dark-gray, or high-pitched + light-gray) than incongruent stimuli (i.e., low-pitched + light-gray, or high-pitched + dark-gray). Planned comparisons revealed that participants were faster to categorize low-pitched + dark-Gray stimuli than high-pitched + dark-gray, $F(1,19) = 8.01$, $p < 0.05$. Likewise, participants tended to be faster to categorize high-pitched + light-Gray stimuli than low-pitched + light-gray, $F(1,19) = 3.55$, $p = 0.07$.

We also observed a main effect of Lightness, $F(1, 19) = 5.09$, $p < 0.05$, $\eta_p^2 = 0.21$. Participants were overall faster to categorize Light-Gray stimuli (mean = 726 ms, SE = 34) than Dark-gray stimuli (mean = 749 ms, SE = 36).

None of the other effects or interactions reached statistical significance.

In this first Experiment, we observed a magnitude-based congruency effect between visual lightness and auditory frequency (see also Marks, 1987). Irrespective of the sensory modality (either visual or auditory), participants were faster to categorize congruent stimuli compared to incongruent stimuli. This is explained by the fact that for the congruent stimuli, the features share the same directional value along the two modalities compared to incongruent stimuli.

Now that we have established a correspondence between lightness and auditory frequency, we can test our prediction that cross-modal correspondence influences cross-modal integration during perceptual learning. This is the aim of Experiments 2a,b.

Experiment 2a

Method

Participants

Thirty-two undergraduate students from Indiana University volunteered to participate in return for partial course credit. All

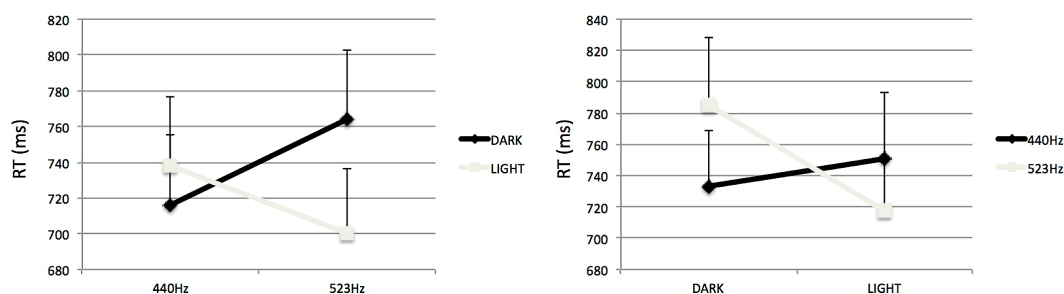


FIGURE 1 | Mean Reaction times to categorize visual stimuli in Experiment 1, as influenced by frequency of accompanying tone (left, visual discrimination task) and to categorize auditory stimuli, as influenced by lightness of accompanying light (right, auditory discrimination task). Errors bars represent ERs of the mean.

of the participants reported no corrected or uncorrected hearing impairment. All the participants had normal or corrected to normal visual acuity.

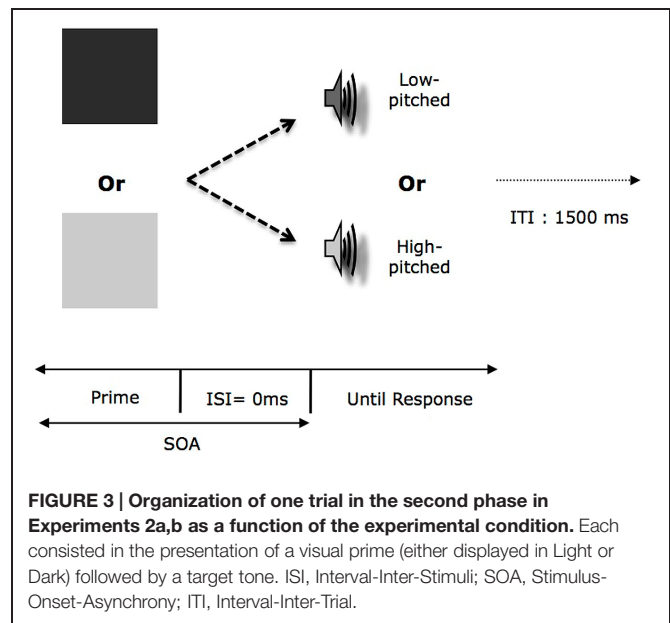
Stimuli and Material

We used the same stimuli and materials as in the first experiment. The only difference was that we used four distinct geometrical shapes (namely a square, a circle, and two octagons; see Brunel et al., 2009) equivalent in area. Since participants should categorize visual shapes according to their lightness, we introduced a manipulation of the shapes because variations on a non-relevant dimension has been demonstrated to contribute to improved perceptual learning (Goldstone et al., 2001).

Procedure

After filling out a written consent form, each participant was tested individually during a session that lasted approximately 15 min. The experiment consisted of two phases. The first phase (learning phase) was based on the hypothesis that the repetition of a sound–brightness association that was not explicitly formulated by the experimenter should lead to the integration of these two components within a single memory trace. Consequently, each trial consisted of the presentation of a shape (either displayed as dark or light gray) for 500 ms. Every shape was presented simultaneously with a tone. Participants were told that their task was to judge, as quickly and accurately as possible, whether the shape was displayed in light or dark gray. They indicated their response by pressing the appropriate key on the keyboard. All of the visual stimuli were presented in the center of the screen, and the intertrial interval was 1,500 ms. For all participants (see **Figure 2**), the shapes displayed in dark gray were presented with the low-pitched tone (440 Hz) and the shapes displayed in light gray were presented with the high-pitched tone (553 Hz). Each gray scale level was presented 32 times in a random order. Half of the participants used their left index finger for the dark-gray response and their right index finger for the light-gray response, while these responses were reversed for the other half of the participants.

The second phase consisted of a categorization task for tones along the pitch dimension (see **Figure 3**). The prime was one shape from the two set of shapes (dark or light gray) presented during the learning phase. In this task, the participants had to



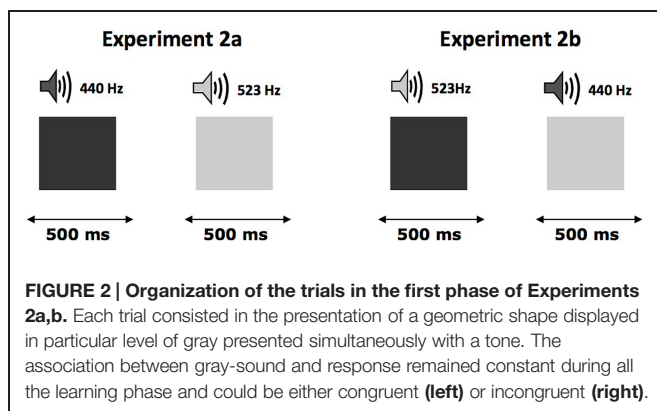
judge as quickly and accurately as possible whether the target sound was low-pitched or high-pitched and indicated their choice by pressing the appropriate key on the keyboard. It is important to stress here that all the participants were instructed to keep their eyes open during the entirety of this phase. In order to avoid a conceptual interpretation of our priming effect (i.e., “square” = “sound”), we introduced a manipulation of the SOA (either 100 or 500 ms) during the second phase. We should observe modulation of the priming effect depending on the level of SOA (i.e., an interference for 100 ms SOA followed by a facilitation at 500 ms SOA). Since participants learned specific bimodal congruent stimuli, the relation between prime (i.e., dark prime or light prime) and target (i.e., low or high-pitched tones) could be the same or opposite compared to what was experienced during the learning phase. In addition, for half of the participants the key assignment was the same between the two phases and the opposite for the other half.

Each participant saw a total of 80 trials, 40 with each target sound; half (20) of the target sounds were presented with a shade of gray that had been associated with the corresponding tone during the learning phase, and the other half were presented with a shade of gray that had been associated with the other tone. The order of the different experimental conditions was randomized within and between groups of participants.

Results and Discussion

Learning Phase

The analyses performed on the CRs and on latencies revealed no significant main effects or any interaction. These results are consistent with the idea that participants performed the gray discrimination task accurately (overall accuracy is 93.9%), and the systematic association between a sound and a shade of gray does not impact the visual nature of the task (see Gallace and Spence, 2006 for a similar interpretation). The same patterns of results were found throughout the learning phase in both experiments.



This phase led participants to integrate the visual shape and the auditory tone within a single memory trace and as a consequence the visual prime shape should be able to influence the processing of the target tone during the test phase (see also Brunel et al., 2009, 2010, 2013).

Test Phase

Separated mixed analyses of variance were performed on latencies (RT) and CRs rates with subject as a random variable, Tone Frequency (Low-Pitched vs. High-Pitched), and Prime-Type (Light vs. Dark) as within-subject variables, and SOA (100 ms vs. 500 ms) as a between-subjects variable.

The analyses performed on the CRs revealed neither a significant main effect (i.e., each $F < 1$) nor any interaction (i.e., each $F < 1$). As far as the RTs were concerned, as expected, our analyses revealed only a significant three-way interaction between SOA, the Tone's Frequency and the Prime's Type, $F(1,30) = 10.16$, $p < 0.05$, $\eta_p^2 = 0.25$. As we can see in **Table 1** the priming effect was reversed for the different SOAs.

Separate analyses of variance were performed for each SOA in order to further investigate these results. For the 100-ms SOA (see **Figure 4**) the analysis revealed a significant interaction between Tone Frequency and Prime-Type, $F(1,15) = 5.21$, $p < 0.05$, $\eta_p^2 = 0.26$. In that condition of SOA, participants were significantly slower to categorize a high-pitched tone preceded by a light-gray visual prime than a dark-gray visual prime, $F(1,15) = 11.52$, $p < 0.05$. However, for the low-pitched target the type of prime did not influence the categorization, $F < 1$.

For the 500-ms SOA (see **Figure 4**), the analysis revealed a significant interaction between Tone Frequency and Prime-Type, $F(1,15) = 5.19$, $p < 0.05$, $\eta_p^2 = 0.26$. Participants were significantly faster to categorize low-pitched tones preceded by a dark-gray visual prime than a light-gray visual prime, $F(1,15) = 15.11$, $p < 0.05$. However, for the high-pitched target the type of prime did not influence the categorization, $F < 1$.

The overall pattern of results presented here replicates what was observed in Brunel et al. (2009; Experiment 1). Indeed, we observed an interference effect for 100 ms SOA in which participants were slower at discriminating the target tone when the prime-target relation matched with the association seen during learning compared to when there was a mismatch in the prime-target relation. Conversely, for the 500 ms SOA, we observed that participants were faster at discriminating the target when the prime-target relation matched the association seen during learning. In sum, learning cross-modal congruent stimuli leads to a pattern of results that is comparable with learning cross-modal stimuli that are unrelated. This result indicates that participants have learned new perceptual units which integrate both perceptual components. Indeed, if such a unit were not created during learning we would have only observed a replication of Experiment 1 results in Experiment 2a. That is to say, we should have found an interaction between visual lightness and auditory frequency irrespective of SOA. We turn now to Experiment 2b to explore the role of incongruency in cross-modal correspondence regarding cross-modal integration.

Experiment 2b

Method

Participants

Thirty-two undergraduate students from Indiana University volunteered to participate in return for partial course credit. All the participants reported no corrected or uncorrected hearing impairment. All of the participants had normal or corrected to the normal visual acuity.

Stimuli, Material, and Procedure

We used the same stimuli, materials, and experimental design as in Experiment 2a. The only exception was that participants were exposed to incongruent stimuli (see **Figure 2**) during learning.

TABLE 1 | Mean response times (RT) and mean percentages of correct responses (CRs) in each experimental condition in Experiment 2.

Learning phase		SOA							
		100 ms				500 ms			
		440 Hz		523 Hz		440 Hz		523 Hz	
Prime	Tone	RT (ms)	CR (%)	RT (ms)	CR (%)	RT (ms)	CR (%)	RT (ms)	CR (%)
Experiment 2a									
Dark-Gray	440 Hz	539 (33)	87.6 (3.8)	508 (37)	85.5 (2.4)	504 (45)	82.7(5.0)	540 (46)	82.7 (4.6)
Light-Gray	523 Hz	538 (35)	82.3 (3.8)	558 (39)	83.9 (2.9)	582 (55)	84.1 (4.9)	549 (39)	80.7 (3.3)
Priming Effect		+1		+50		-78		9	
Experiment 2b									
Dark-Gray	523 Hz	437 (33)	91.0 (2.3)	470 (36)	84.7 (2.2)	504 (44)	92.5 (1.4)	495 (40)	90.1 (2.4)
Light-Gray	440 Hz	475 (36)	85.2 (2.1)	461 (34)	88.4 (1.7)	477 (39)	90.2 (2.4)	515 (45)	88.8 (2.2)
Priming Effect		+37		+9		-27		-20	

SEs in parenthesis. Priming effects were obtained by subtracting the matching condition from the mismatching condition. Negative values indicate facilitation effects whereas positive values indicate interference effects.

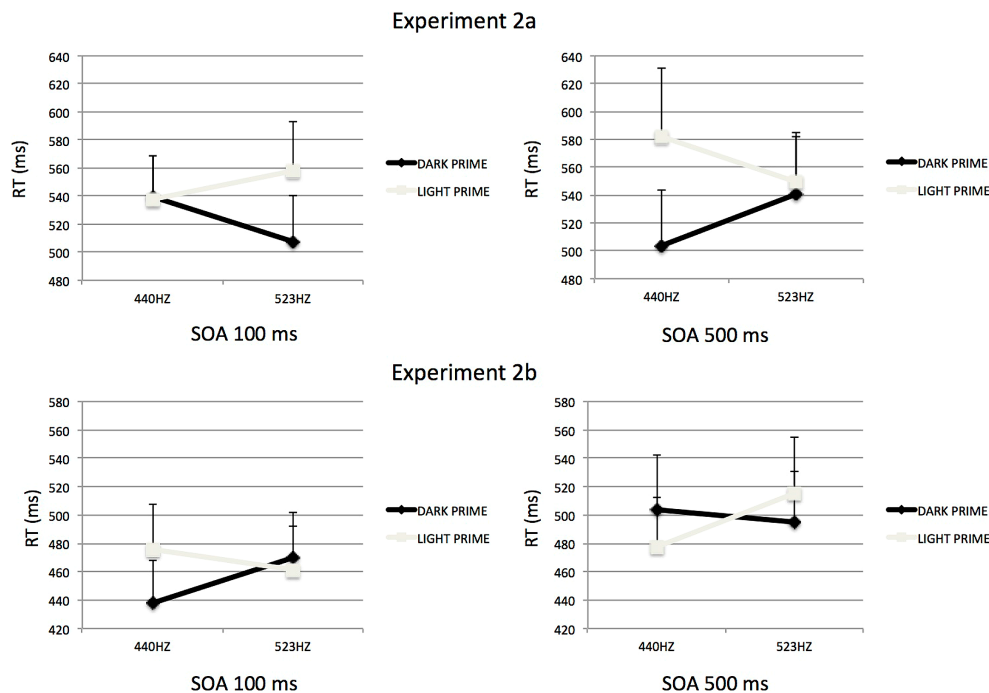


FIGURE 4 | Mean Reaction times to categorize auditory stimuli, as influenced by visual prime presented 100 ms (right) or 500 ms (left) in Experiment 2. The association at learning could be either congruent (top; Experiment 2a) or incongruent (bottom; Experiment 2b). Errors bars represent ERs of the mean.

Results and Discussion

Test Phase

Separated mixed analyses of variance were performed on latencies RT and CRs rates with subject as a random variable, Tone Frequency (Low-Pitched vs. High-Pitched), and Prime-Type (Light vs. Dark) as within-subject variables, and SOA (100 ms vs. 500 ms) as between-subjects variables. The analyses performed revealed only a significant three-way interaction between SOA, Tone Frequency and Prime-Type, respectively, $F(1,30) = 14.96$, $p < 0.05$, $\eta_p^2 = 0.33$ for RTs and $F(1, 30) = 4.83$, $p < 0.05$, $\eta_p^2 = 0.14$ for CR rates. For clarity, we further report here only the analysis regarding the RTs since the results on CR are comparable to those observed for RTs (see **Table 1**). As we can see in **Table 1** the priming effect was reversed for the different SOAs but the same for the different experiments.

Separate analyses of variance were performed for each SOA in order to interpret these results. For the 100-ms SOA (see **Figure 4**) the analysis revealed a significant interaction between Tone Frequency and Prime-Type, $F(1,15) = 7.15$, $p < 0.05$, $\eta_p^2 = 0.32$. With this short SOA, participants were significantly slower to categorize low-pitched tone preceded by a light-gray visual prime than a dark-gray visual prime, $F(1,15) = 6.74$, $p < 0.05$. However, for the high-pitched target the type of prime did not significantly influence the categorization, $F(1,15) = 1.07$, $p = 0.31$, but the trend is consistent with an interference priming effect, i.e., participants were slower to categorize high-pitched tone preceded by a dark-gray visual prime than a light-gray visual prime (see **Table 1**).

For the 500-ms SOA (see **Figure 4**), the analysis only revealed a significant interaction between Tone Frequency and Prime-Type, $F(1,15) = 7.84$, $p < 0.05$, $\eta_p^2 = 0.34$. In that condition of SOA, participants were significantly faster to categorize a low-pitched tone preceded by a light-gray visual prime than a dark-gray visual prime, $F(1,15) = 6.74$, $p < 0.05$. However, for the high-pitched target the type of prime did not significantly influence the categorization, $F(1,15) = 2.84$, $p = 0.11$, but the trend is also consistent with a facilitation priming effect, i.e., participants were faster to categorize high-pitched tone preceded by a dark-gray visual prime than a light-gray visual prime.

The overall pattern of results replicates those observed in Experiment 2a. However, the manipulation of the cross-modal correspondence at learning had a significant influence on the size of the priming effect irrespective of interference or facilitation (Mann–Whitney U test, $Z = 1.81$, $p < 0.05$), with a smaller priming effect seen for Experiment 2b (Mean = 20 ms) than for Experiment 2a (Mean = 34 ms). This difference might indicate that the decay of the priming effect over time is faster for incongruent stimuli at learning than for congruent stimuli at learning.

General Discussion

The aim of the present study was to provide evidence in support of the assumption that cross-modal correspondences modulate cross-modal integration during perceptual learning. In our

first Experiment, we established a cross-modal correspondence between visual lightness and auditory frequency (see also Marks, 1987). Indeed, regardless of the sensory modality of the task, participants were faster to categorize congruent stimuli (i.e., High-pitched + Light-Gray or Low-pitched + Dark-gray) compared to incongruent stimuli (i.e., the opposite mapping between lightness and auditory frequency). This result is consistent with previous results on cross-modal correspondence (see Gallace and Spence, 2006; Evans and Treisman, 2010; for a review see Spence, 2011). Our second Experiment explored whether learning bimodal congruent or incongruent stimuli influenced the integration mechanism. This idea is consistent with experimental evidence showing that cross-modal integration is involved during perceptual learning (see Brunel et al., 2009, 2010, 2013; Zmigrod and Hommel, 2013) and with experimental work showing that cross-modal correspondences modulate cross-modal integration (see Parise and Spence, 2009). Our experiment used an original priming paradigm that we have designed to study cross-modal perceptual learning (see Brunel et al., 2009, 2010, 2013). In this paradigm, during a learning phase, participants implicitly learn an audiovisual perceptual unit (e.g. a “sound square”). Then, the consecutive phase allows us to test for the existence of such a unit as well as its nature. To do so, we ask participants to categorize target tones preceded by a visual prime. In our previous studies, we showed a priming effect from the visual prime to the target tone limited to the visual prime that was presented with a sound during the learning phase. This result indicates that participants integrated the visual and auditory features and thus the presence of one feature as a prime automatically triggers the other. In the same vein, Meyer et al. (2007) showed that processing of a visual component (i.e., a red flash) that was previously presented with a sound (i.e., a telephone ringing) produced auditory cortex activation. Most interestingly, the manipulation of the SOA during the test phase allows us to rule out a conceptual or symbolic interpretation of the priming effect. In our previous studies, depending on the SOA value (i.e., shorter or at the same duration than the duration of the association during learning), we observed either an interference effect or a facilitation effect. The facilitation or interference of the priming effects depends on the temporal overlap between sound–target processing and the reactivation of an auditory component by the visual prime (for similar consideration, see Riou et al., 2014). It is therefore essentially this variability in the influence of the prime as a function of SOA that shows the perceptual nature of the cross-modal learnt unit. In sum, with our paradigm, we are able to test the implication of a cross-modal integration mechanism during learning.

In Experiment 2a, we showed that learning a congruent cross-modal stimulus produces a priming effect consistent with previous findings (i.e., interference followed by facilitation depending with increasing SOA, see Brunel et al., 2009). This confirms that participants exposed to an association between a visual component and an auditory component presented simultaneously created an integrated memory trace (see Versace et al., 2014) or event (see Zmigrod and Hommel, 2013). Once integrated, each component is no longer accessible individually without an effect of the other component. As a consequence, when participants see the visual component by itself, the auditory component is

automatically activated as well. Moreover, the facilitation or interference of the priming effect was dependent on the temporal overlap between sound–target processing and auditory component reactivation (i.e., SOA manipulation). We interpreted this modulation as evidence of the perceptual nature of the memory component reactivated by the visual prime and thus the integration between these two components within a memory trace or event (see also, Brunel et al., 2009; Zmigrod and Hommel, 2010, 2013). However, these results are not just a replication of previous results because of our manipulation of the pre-experimental correspondence between sensory dimensions. Indeed, to the best of our knowledge this is the first time it has been shown that participants integrate the specific relation between perceptual features. In our previous studies (Brunel et al., 2009, 2010, 2013), the relation between the prime and the target was at a dimensional level (i.e., the prime and the target shared or did not share a sound dimension). With Experiment 2a, the prime–target relation is at the feature level (i.e., prime–target relations were congruent or incongruent with the previously learned associations). Moreover, we showed that participants actively learned such a relation despite the fact that the relation between the features is already congruent. This is evident by comparing the results observed in Experiment 1 and those observed in Experiment 2a. In Experiment 1, we showed a cross-modal congruency effect between visual lightness and auditory frequency. Participants were faster at processing congruent stimuli (i.e., either dark gray + low-pitched or light gray + high-pitched) compared to incongruent ones (i.e., the opposite mapping). In Experiment 2a, this effect was modulated by the SOA. According to our previous work (see Brunel et al., 2009, 2010, 2013), this modulation necessarily indicates that participants have learned a new perceptual unit. Otherwise, we should have only observed a replication of Experiment 1.

In Experiment 2b, we showed that, even when participants learn an incongruent association, the same pattern of priming effect is still observed at test. This result indicates that learning incongruent stimuli does not disrupt the cross-modal integration. However, a comparison of the priming effect observed between Experiments 2a,b indicates a smaller priming effect with learned incongruent units than with congruent ones. It is possible that the cross-modal correspondence influences integration. Indeed, the difference in the size of the priming effect between Experiments 2a,b might be due to different decay functions over the time. Because the priming effect reported in our experiments is a consequence of the newly integrated units (i.e., 32 presentations during learning), one might assume that the units will decline over time. In other words, since bimodal incongruent stimuli were only learned during our experiment, the association would probably be expected to decay faster than a congruent correspondence that has the strength of having been reinforced frequently in the past (see Marks, 1987). So our results seem to indicate that only a weak form of integration can be created in such a short period of time.

Spence (2011) proposed cross-modal correspondences can be understood in terms of Bayesian priors. The general idea is that humans may combine stimuli in a statistically optimal manner by combining prior knowledge and sensory information and

weighting each of them by their relative reliabilities. In such an approach, cross-modal correspondences could be modeled in terms of prior knowledge (see Ernst, 2007; Parise and Spence, 2009; Spence, 2011). According to this model, the cognitive system establishes relations (or couplings) between stimuli in order to adapt to the situation and its constraints. The prior knowledge about the stimulus mapping has a consequence on the coupling of the stimuli (or the integration, see Ernst, 2007). The greater the prior knowledge in the system about the fact that two stimuli belong together, the stronger these stimuli will be coupled. In other words, the stronger the coupling, the most likely unisensory signals would be fused together, leading to the creation of a multisensory units. One major consequence would be an elevation of the discrimination threshold to detect internal conflict within a stimulus (e.g., asynchronous presentation, see Parise and Spence, 2009). In Experiment 1, we showed that prior knowledge about coupling between auditory frequency and visual lightness increases the perceptual processing of cross-modal congruent stimuli compared to incongruent ones. More interestingly, in Experiments 2a,b, we manipulated the prior knowledge distribution by creating an implicit novel association during the learning phase. This manipulation affected the cross-modal congruency effect that we observed in Experiment 1. The fact that we exposed participants to a pair of cross-modal features might have reduced the influence of coupling priors for the pair. As a consequence, the priming effect that we observed can be considered to be a measure of the modification of the influence of the coupling prior for the pair. As soon as one of element of the unit is presented the system makes an assumption about (or simulates) the presence of the other. Given that we observed a modulation of the priming effect depending on the SOA, we can argue that this assumption (or simulation) is more likely to occur at a perceptual stage rather than a decisional stage (for similar consideration, see Brunel et al., 2009, 2010, 2013; Evans and Treisman, 2010; Rey et al., 2014, 2015; Riou et al., 2014). Finally, it seems that learning congruent stimuli leads to the creation of “stronger” units (or coupling) over time because the system already has repeatedly experienced that these stimuli go together. Moreover, our results seem to indicate that the system does not need a large sampling of experiences to establish such prior knowledge distribution (or *coupling prior*). Indeed, the fact that we replicate our results in both Experiments 2a,b showed that the prior knowledge distribution depends on the experiences of the cognitive system rather than being exclusively built-in. Otherwise, we would not have observed a priming effect in Experiment 2b that conceptually replicated the one found in Experiment 2a.

References

- Brunel, L., Goldstone, R. L., Vallet, G., Riou, B., and Versace, R. (2013). When seeing a dog activates the bark: multisensory generalization and distinctiveness effects. *Exp. Psychol.* 60, 100–112. doi: 10.1027/1618-3169/a000176
- Brunel, L., Labeye, E., Lesourd, M., and Versace, R. (2009). The sensory nature of episodic memory: sensory priming effects due to memory trace activation. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 1081–1088. doi: 10.1037/a0015537
- Brunel, L., Lesourd, M., Labeye, E., and Versace, R. (2010). The sensory nature of knowledge: sensory priming effects in semantic categorization. *Q. J. Exp. Psychol.* 63, 955–964. doi: 10.1080/17470210903134369
- Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K., et al. (1997). Activation of auditory cortex during silent lipreading. *Science* 276, 593–596. doi: 10.1126/science.276.5312.593
- Chen, Y.-C., and Spence, C. (2010). When hearing the bark helps to identify the dog: semantically-congruent sounds modulate the identification of masked pictures. *Cognition* 114, 389–404. doi: 10.1016/j.cognition.2009.10.012
- Ernst, M. (2007). The sensory nature of knowledge: sensory priming effects in semantic categorization. *Q. J. Exp. Psychol.* 63, 955–964. doi: 10.1080/17470210903134369
- Evans, D. M., and Treisman, A. (2010). The sensory nature of knowledge: sensory priming effects in semantic categorization. *Q. J. Exp. Psychol.* 63, 955–964. doi: 10.1080/17470210903134369
- Rey, D., Brunel, L., and Versace, R. (2014). The sensory nature of knowledge: sensory priming effects in semantic categorization. *Q. J. Exp. Psychol.* 63, 955–964. doi: 10.1080/17470210903134369
- Riou, B., Brunel, L., and Versace, R. (2014). The sensory nature of knowledge: sensory priming effects in semantic categorization. *Q. J. Exp. Psychol.* 63, 955–964. doi: 10.1080/17470210903134369
- Spence, C. (2011). The sensory nature of knowledge: sensory priming effects in semantic categorization. *Q. J. Exp. Psychol.* 63, 955–964. doi: 10.1080/17470210903134369
- Spence, C., and Calvert, G. A. (1997). Activation of auditory cortex during silent lipreading. *Science* 276, 593–596. doi: 10.1126/science.276.5312.593
- Spence, C., and Chen, Y.-C. (2010). When hearing the bark helps to identify the dog: semantically-congruent sounds modulate the identification of masked pictures. *Cognition* 114, 389–404. doi: 10.1016/j.cognition.2009.10.012

Conclusion

Our results support the idea that cross-modal correspondences, through the modification of coupling priors, modulate cross-modal integration during perceptual learning. Thus, perceptual consciousness could be considered as emerging from the integration of the current situation and the knowledge about prior situations. In that case, we can envisage that integration is crucial to conscious processing and might be a form of signature to those processing (see also, Dehaene et al., 2014)

However, there are still remaining open questions about how cross-modal integration might be linked to a very specific form of perceptual consciousness (e.g., synesthesia). Like for cross-modal correspondences, synesthetic experiences could be considered as structurally, semantically or statistically mediated (see Spence, 2011). However, recent findings seem to indicate that synesthetic experience could be understood as a consequence of some hyper-integration (or hyper-binding, see Mroczko-Wąsowicz and Werning, 2012) between an unusually large number of sensory or semantic attribute domains. This would be consistent with the idea that integration could be involved during the emergence of conscious states.

Ethics statement

This research was conducted in accordance to the declaration of Helsinki, and had ethical approval from the Indiana University IRB office. All participants provided written informed consent and received partial course credit in return for their participation.

Acknowledgments

This research was supported by ACCEPT (“Assistance Tools and Cognitive Contribution: Embodied Potential of Technology”), French research ministerial mission (MiRe-DREES and CNSA), FYSEN Foundation (94, rue de Rivoli, 75001 PARIS, FRANCE), and National Science Foundation REESE grant 0910218. PC was also supported by Graduate Training Fellowship SFRH/BD/78083/2011 from the Portuguese Foundation for Science and Technology (FCT), co-sponsored by the European Social Fund

- Cohen, J. D., MacWhinney, B., Flatt, M., and Provost, J. (1993). PsyScope: a new graphic interactive environment for designing psychology experiments. *Behav. Res. Meth. Ins. Comp.* 25, 257–271. doi: 10.3758/BF03204507
- Dehaene, S., Charles, L., King, J.-M., and Marti, S. (2014). Toward a computational theory of conscious processing. *Curr. Opin. Neurobiol.* 25, 76–84. doi: 10.1016/j.conb.2013.12.005
- Ernst, M. (2007). Learning to integrate arbitrary signals from vision and touch. *J. Vis.* 7, 1–14. doi: 10.1167/7.5.7
- Evans, K. K., and Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *J. Vis.* 10, 1–12. doi: 10.1167/10.1.6
- Fodor, J. A. (1983). *Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press.
- Gallace, A., and Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Percept. Psychophys.* 68, 1191–1203. doi: 10.3758/BF03193720
- Giard, M. H., and Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behaviour and electrophysiological study. *J. Cogn. Neurosci.* 11, 473–490. doi: 10.1162/089892999563544
- Goldstone, R. L. (2000). Unitization during category learning. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 86–112. doi: 10.1037/0096-1523.26.1.86
- Goldstone, R. L., Landy, D., and Brunel, L. C. (2013). Improving the perception to make distant connections closer. *Front. Psychol.* 2:385. doi: 10.3389/fpsyg.2011.00385
- Goldstone, R. L., Lippa, Y., and Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition* 78, 27–43. doi: 10.1016/S0010-0277(00)00099-8
- Klapetek, A., Ngo, M. K., and Spence, C. (2012). Does crossmodal correspondence modulate the facilitatory effect of auditory cue on visual search? *Atten. Percept. Psychophys.* 74, 1154–1167. doi: 10.3758/s13414-012-0317-9
- King, A., and Calvert, G. (2001). Multisensory integration: perceptual grouping by eye and ear. *Curr. Biol.* 11, 322–325. doi: 10.1016/S0960-9822(01)00175-0
- Laurienti, P., Burdette, J., Maldjian, J., and Wallace, M. T. (2006). Enhanced multisensory integration in older adults. *Neurobiol. Aging* 27, 1155–1163. doi: 10.1016/j.neurobiolaging.2005.05.024
- MacLeod, A., and Summerfield, Q. (1990). A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: rationale, evaluation, and recommendations for use. *Br. J. Audiol.* 24, 29–43. doi: 10.3109/03005369009077840
- Marks, L. E. (1987). On cross-modal similarity: auditory-visual interactions in speeded discrimination. *J. Exp. Psychol. Hum. Percept. Perform.* 13, 384–394. doi: 10.1037/0096-1523.13.3.384
- Marks, L. E. (2004). “Cross-modal interactions in speeded classification,” in *Handbook of Multisensory Processes*, eds G. A. Calvert, C. Spence, and B. E. Stein (Cambridge: MIT Press), 85–105.
- Meyer, M., Baumann, S., Marchina, S., and Jancke, L. (2007). Hemodynamic responses in human multisensory and auditory association cortex to purely visual stimulation. *BMC Neurosci.* 8:14. doi: 10.1186/1471-2202-8-14
- Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., and Foxe, J. J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Brain Res. Cogn. Brain Res.* 14, 115–128. doi: 10.1016/S0926-6410(02)00066-6
- Mroczo-Wąsowicz, A., and Werning, M. (2012). Synesthesia, sensory-motor contingency, and semantic emulation: how swimming style-color synesthesia challenges the traditional view of synesthesia. *Front. Psychol.* 3:279. doi: 10.3389/fpsyg.2012.00279
- Parise, C. V., and Spence, C. (2009). “When birds of a feather flock together”: synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PLoS ONE* 4:e5664. doi: 10.1371/journal.pone.0005664
- Ramachandran, V. S., and Hubbard, E. M. (2001). Synaesthesia—A window into perception, thought and language. *J. Conscious. Stud.* 8, 3–34.
- Rey, A. E., Riou, B., Muller, D., Dabic, S., and Versace, R. (2015). “The mask who wasn’t there”: visual masking effect with the perceptual absence of the mask. *J. Exp. Psychol. Learn. Mem. Cogn.* 41, 567–573. doi: 10.1037/xlm0000051
- Rey, A. E., Riou, B., and Versace, R. (2014). Demonstration of an ebbinghaus illusion at a memory level: manipulation of the memory size and not the perceptual size. *Exp. Psychol.* 61, 378–384. doi: 10.1027/1618-3169/a000258
- Riou, B., Rey, A., Vallet, G. T., Cuny, C., and Versace, R. (2014). Perceptual processing affects the reactivation of a sensory dimension during a categorization task. *Q. J. Exp. Psychol.* 1–8. doi: 10.1080/17470218.2014.978876 [Epub ahead of print].
- Schmidt, J. R., and De Houwer, J. (2012). Contingency learning with evaluative stimuli. *Exp. Psychol.* 59, 175–182. doi: 10.1027/1618-3169/a000141
- Schmidt, J. R., De Houwer, J., and Besner, D. (2010). Contingency learning and unlearning in the blink of an eye: a resource dependent process. *Conscious. Cogn.* 19, 235–250. doi: 10.1016/j.concog.2009.12.016
- Smith, L. B., and Sera, M. (1992). A developmental analysis of the polar structure of dimensions. *Cogn. Psychol.* 24, 99–142. doi: 10.1016/0010-0285(92)90004-L
- Spence, C. (2011). Crossmodal correspondences: a tutorial review. *Atten. Percept. Psychophys.* 73, 971–995. doi: 10.3758/s13414-010-0073-7
- Stein, B. E., and Meredith, M. A. (1993). *The Merging of Senses*. Cambridge, MA: MIT Press.
- Teder-Sälejärvi, W. A., McDonald, J. L., Di Russo, F., and Hillyard, S. A. (2002). An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Cogn. Brain Res.* 14, 106–114. doi: 10.1016/S0926-6410(02)00065-4
- Teder-Sälejärvi, W. A., Russo, F., McDonald, J. L., and Hillyard, S. A. (2005). Effects of spatial congruity on audio-visual multimodal integration. *J. Cogn. Neurosci.* 17, 1396–1409. doi: 10.1162/0898929054985383
- Versace, R., Vallet, G. T., Riou, B., Lesourd, M., Labeye, E., and Brunel, L. (2014). Act-in: an integrated view of memory mechanisms. *J. Cogn. Psychol.* 26, 280–306. doi: 10.1080/20445911.2014.892113
- Zmigrod, S., and Hommel, B. (2010). Temporal dynamics of unimodal and multimodal feature binding. *Atten. Percept. Psychophys.* 72, 142–152. doi: 10.3758/APP.72.1.142
- Zmigrod, S., and Hommel, B. (2011). The relationship between feature binding and consciousness: evidence from asynchronous multi-modal stimuli. *Conscious. Cogn.* 20, 586–593. doi: 10.1016/j.concog.2011.01.011
- Zmigrod, S., and Hommel, B. (2013). Feature integration across multimodal perception and action: a review. *Multisens. Res.* 26, 143–157. doi: 10.1163/22134808-00002390

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.

Copyright © 2015 Brunel, Carvalho and Goldstone. 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.

Cross-modal metaphorical mapping of spoken emotion words onto vertical space

Pedro R. Montoro^{1*}, María José Contreras¹, María Rosa Elosúa¹ and Fernando Marmolejo-Ramos²

¹ Departamento de Psicología Básica I, Facultad de Psicología, Universidad Nacional de Educación a Distancia, Madrid, Spain, ² Gösta Ekman Laboratory, Department of Psychology, Stockholm University, Stockholm, Sweden

OPEN ACCESS

Edited by:

Aleksandra Mroczko-Wasowicz,
National Yang Ming University, Taiwan

Reviewed by:

Ljubica Damjanovic,
University of Chester, UK
Suzanne Oosterwijk,
University of Amsterdam, Netherlands
Liad Mudrik,
California Institute of Technology, USA

*Correspondence:

Pedro R. Montoro,
Departamento de Psicología Básica I,
Facultad de Psicología, Universidad
Nacional de Educación a Distancia,
C/Juan del Rosal 10, 28040 Madrid,
Spain
prmontoro@psi.uned.es

Specialty section:

This article was submitted to
Consciousness Research,
a section of the journal
Frontiers in Psychology

Received: 20 January 2015

Accepted: 29 July 2015

Published: 11 August 2015

Citation:

Montoro PR, Contreras MJ,
Elosúa MR and Marmolejo-Ramos F
(2015) Cross-modal metaphorical
mapping of spoken emotion words
onto vertical space.
Front. Psychol. 6:1205.
doi: 10.3389/fpsyg.2015.01205

From the field of embodied cognition, previous studies have reported evidence of metaphorical mapping of emotion concepts onto a vertical spatial axis. Most of the work on this topic has used visual words as the typical experimental stimuli. However, to our knowledge, no previous study has examined the association between affect and vertical space using a cross-modal procedure. The current research is a first step toward the study of the metaphorical mapping of emotions onto vertical space by means of an auditory to visual cross-modal paradigm. In the present study, we examined whether auditory words with an emotional valence can interact with the vertical visual space according to a 'positive-up/negative-down' embodied metaphor. The general method consisted in the presentation of a spoken word denoting a positive/negative emotion prior to the spatial localization of a visual target in an upper or lower position. In Experiment 1, the spoken words were passively heard by the participants and no reliable interaction between emotion concepts and bodily simulated space was found. In contrast, Experiment 2 required more active listening of the auditory stimuli. A metaphorical mapping of affect and space was evident but limited to the participants engaged in an emotion-focused task. Our results suggest that the association of affective valence and vertical space is not activated automatically during speech processing since an explicit semantic and/or emotional evaluation of the emotionally valenced stimuli was necessary to obtain an embodied effect. The results are discussed within the framework of the embodiment hypothesis.

Keywords: emotions, vertical space, cross-modal procedure, embodiment, metaphorical mapping

Introduction

Emotion concepts have been researched extensively, particularly in relation to abstract and concrete concepts (see Altarriba and Bauer, 2004), and have become a topic of particular interest in the embodied cognition framework (e.g., Niedenthal et al., 2005, 2009; see also Meteyard et al., 2012). Specifically, it has been argued that abstract and emotion concepts have sensorimotor properties much like concrete concepts. For example, it has been shown that there is a metaphorical association between emotionally valenced concepts and the vertical plane (e.g., Meier and Robinson, 2004; Meier et al., 2011; Sasaki et al., 2012, 2015; Santiago et al., 2012; Marmolejo-Ramos and Dunn, 2013; Marmolejo-Ramos et al., 2013, 2014; Xie et al., 2014, 2015; Damjanovic and Santiago, 2015). Most of the work

on the association between emotion words and the vertical axis has used visual words as typical experimental stimuli. Indeed, most tasks rely on within-modal tasks (e.g., effects of images on word processing), chiefly visual tasks, and to a certain extent rarely examine cross-modal effects (e.g., effects of sounds on word processing); let alone study emotions in cross-modal processing (see Gerdes et al., 2014). This study aims to investigate the association between emotional auditory stimuli and vertical visual space. Specifically, an auditory to visual cross-modal paradigm task is used to explore the limits of the metaphorical mapping between concepts and bodily simulated space. That is, we studied whether auditory words with an affective valence can influence the spatial localization of visual stimuli in line with the 'positive-up/negative-down' vertical spatial metaphor. Additionally, we included different intervals between the auditory word and visual cue in order to explore the automaticity (or lack thereof) of this audiovisual emotional processing.

The introduction in this article is divided as follows. Firstly, some examples of research in the embodiment of emotions are presented to outline the overarching topic of the studies reported in this article. Secondly, the specific case of conceptual metaphors and emotions is considered. In particular, research on the valence-space metaphor is discussed since it constitutes the exact scope of this article. Finally, research relating to the timing of the valence-space metaphor is examined. This is a new aspect, which is currently being investigated.

Embodiment of Emotions

Traditionally, most research on emotions has employed visual stimuli. However, recent work has used non-visual emotional stimuli in relation to other modalities. In a study by Tajadura-Jiménez et al. (2010), it was shown that unpleasant approaching sounds elicit a more intense emotional response than pleasant or neutral preceding sounds. The emotional response of the participants consisted of a greater self-reported emotional experience and a greater facial muscle response during unpleasant approaching sounds than during the preceding conditions. Furthermore, listening to white noise, a type of sound rated as unpleasant, while people provided odor ratings for different smells, led to lower pleasantness and sweetness and higher dryness odor ratings (Velasco et al., 2014). Other studies using emotionally valenced tactile stimuli have found that even touch gestures communicated remotely (i.e., via a tactile device) can convey different emotional intentions (Rantala et al., 2013). Specifically, Rantala et al. (2013) found that squeeze actions are associated with unpleasant and aroused emotional intentions, whereas finger touch was better at conveying pleasant and relaxed emotional intentions. These studies suggest that emotional stimuli can indeed influence the body's somatosensory and sensorimotor systems and emotions can indeed be conveyed through these systems.

Research has shown that emotions are subserved by somatosensory and sensorimotor systems that work together in response to internal or external stimulus events (e.g., Scherer, 2005). That is, emotions are made up of interoceptions, exteroceptions, and memories that are instantiated whenever an

emotion is re-experienced, recalled or evoked (see Niedenthal et al., 2005, 2009). Indeed, recent neuroimaging research further indicates that experiencing emotions entails the activation of distributed multimodal brain networks involved in various psychological processes (Oosterwijk et al., 2012). Further, research on the embodiment of emotions has shown that even emotion concepts and words can elicit the activation of such multimodal networks (see Niedenthal et al., 2005, 2009; see also Marmolejo-Ramos et al., 2015). For example, Wilson-Mendenhall et al. (2013) found that the sensorimotor cortex, the amygdala and the hippocampus were activated during vivid imagination of situations referring to physical danger and social evaluations¹ (where sensorimotor activations may have resulted from the simulation *per se* rather than the emotional activation). As the authors indicated, the hippocampus is involved in other psychological processes such as binding multimodal mnemonic information and simulating future and imagined situations. Such findings support the idea that multimodal simulations are deployed when emotions are processed. Recent research on the conceptual modality-switching cost effect (see Pecher et al., 2003) lends further support to the multimodality associated to emotions. Specifically, it has been found that there is a switching cost when shifting from somatosensory to emotional modality but not the other way around (Dagaev and Terushkina, 2014). The authors further argued that activation of emotion concepts entails the activation of somatosensory modalities such as touch, pressure, vibration, temperature, pain, and joint and muscle sensitivity. Overall, these studies indicate that the processing of emotion concepts entails the activation of concepts relating to somatosensory and sensorimotor systems. However, research investigating how emotions are transferred across somatosensory and sensorimotor modalities in real modality-switching tasks is just emerging (see Velasco et al., 2014). Specifically, it would be important to investigate what emotions are more or less easily transferred across somatosensory and sensorimotor systems and what their time course is during switching tasks.

Conceptual Metaphors and Emotions

Conceptual metaphors occur when concepts are mapped from a source domain onto a target domain. Specifically, target domains refer to abstract concepts that are to be mapped onto source domains that are concrete and bodily based (see Gallese and Lakoff, 2005). Based on these premises, Wilson and Gibbs (2007) found that performing body actions, or even imagining them, facilitated the comprehension of metaphorical sentences, when compared to not performing any sort of action. In the specific case of emotions and space, emotions (target domain) are conceptualized as spatial locations (source domain). That is, just as upper locations tend to be evaluated as more positive than lower spatial locations, positive words tend to be allocated in upper spatial areas, while negative words tend to be allocated in lower areas (Marmolejo-Ramos et al., 2013). According to Lakoff (2014), metaphorical mappings rely on the human body itself and its neurological substratum as the source domain in

¹We thank one of the reviewers for suggesting this possible explanation.

order to represent physical properties (e.g., motion and space). That is, the sensorimotor and somatosensory systems dictate the experience with the physical environment used to ground abstract concepts (e.g., these systems enable the processing of physical properties like high-low as analogs of valenced abstract concepts). Thus, the metaphorical mapping of emotions onto space requires somatosensory and sensorimotor simulations in order to comprehend the linkage from the target to the source domain.

In order to test the metaphorical mapping of emotions onto space, Ansorge and Bohnner (2013; Ansorge et al., 2013) used an implicit association task in which participants categorized words like 'up' as elevated or less elevated and affective words like 'happy' as positive or negative. Interestingly, when spatial and affective stimuli were metaphorically congruent and required the same response (e.g., up-happy), faster responses and fewer errors were found than when spatial and affective stimuli were incongruent and required a different response (see Experiment 1). That is, faster responses were observed when target words were presented in spatial congruent locations (e.g., 'happy' in the upper part of a computer screen) than when they were presented in incongruent locations (e.g., 'sad' in the upper part of a computer screen) and this association was seemingly implicit. Other studies confirmed that there is even a mapping of emotion sentences onto space; however, such an association holds only when the task demands an explicit affective evaluation of the target (Marmolejo-Ramos et al., 2014; Experiment 2). In the case of emotion words, it has been shown that such an association occurs only when valence is to be explicitly evaluated or when they refer to emotional states that have discernible body postures (Dudschig et al., 2015). Note that these types of results are informative as to within-modal emotion processing.

However, not much is known as to multimodal and cross-modal emotion processing. Indeed, the employment of cross-modal paradigms for the study of conceptual-physical interactions could contribute relevant data to decide between alternative models of embodiment. Recently, Meteyard et al. (2012) reviewed four different theories of embodiment arranged in a continuum from "strong embodiment" (complete dependence on the relationship to sensory-motor systems) to completely "unembodied" (complete independence between both). The evidence revised by Meteyard et al. (2012) supports balanced/moderate versions of the embodiment hypothesis, which propose that sensory and motor information is activated when a semantic representation is accessed. In the present study, we test the hypothesis that deep semantic processing is needed to display the effect of embodiment. To prove this, two experiments were planned; in Experiment 1 only shallow processing was required, while in Experiment 2, the effect of emotional versus non-emotional processing was contrasted.

A review paper by Gerdes et al. (2014), noted that there is behavioral, physiological, and electrophysiological evidence showing the effects that emotional visual stimuli have on auditory processing. However, only a couple of studies have investigated how emotional sounds influence visual processing (see also

Table 1 in Gerdes et al., 2014). In one of these studies, it was found that when emotionally valenced stimuli were visually presented, recognition of visually presented neutral stimuli was impaired. However, when emotionally valenced stimuli were auditorily presented, recognition of visually presented neutral stimuli was enhanced (Zeelenberg and Bocanegra, 2010). Furthermore, another study found behavioral effects of emotional sounds on visual processing only when visual items were presented on the right visual hemifield² (Harrison and Davies, 2013; see Brosch et al., 2008 for a study in which emotionally valenced pseudowords were used). Thus, these studies suggest that auditory emotional stimuli affect visual processing. A pending issue, though, is the automaticity accompanying such an effect and whether the effect carries over onto metaphorical mapping (see above).

Automaticity of the Metaphorical Mapping of Emotions onto Space

Some researchers have found that the mapping of visually presented emotion words onto vertical space seem to occur automatically even when the experimental task requires a shallow processing of such mapping; however, such a finding is not clear-cut. As Brookshire et al. (2010) argued, finding vertical space-valence congruity depends on contextual modulation such that the effect disappears with repetition (Experiment 1) and reappears with attention orientation (Experiment 2). Studies on the metaphorical mapping of emotion words on the horizontal (left-right) plane have also found that explicit attention to the valence of the words activates space-valence associations (de la Vega et al., 2012). Thus, these authors argued that an association between horizontal space and valence is not automatic and occurs only when explicit valence assessment is required.

Few studies tapping the effect of auditorily valenced stimuli on visual processing have dealt with the automaticity of this process. A study in which emotional pseudowords were listened to prior to the localisation of a rightward- or leftward-presented dot on the screen indicated that visual spatial cuing by auditory emotional stimuli seems to happen at the very early stages of processing (i.e., between 130 and 190 ms); particularly in the striate visual cortex (Brosch et al., 2009). In this study, visual targets were presented 550 to 750 ms (in increments of 50 ms) after auditory cue onset yet this data was not entered into the statistical analyses. While such SOAs could have been used to further examine the behavioral time-course of the cross-modal audiovisual effect, they were included in the study in order to approximate temporal changes (e.g., variations in stress and pitch) that affect prosody.

It is not known, however, whether such automaticity operates during metaphorical mapping on the vertical space. As mentioned above, some tasks using within-modality visually presented words seem to find a rather automatic mapping from emotions onto space (e.g., Ansorge and Bohnner, 2013; see Brookshire et al., 2010 and de la Vega et al., 2012 for

²It is worth clarifying at this point that in Table 1 in Gerdes et al. (2014), the study by Harrison and Davies (2013, p. 5) is perhaps mistakenly cited as an example of no influence of emotional sounds on a visual task but in the text this mistake is not present.

examples of studies challenging a clear-cut automatic mapping), but, when longer linguistic units are used, the effect exists only when an explicit emotional evaluation is required (Marmolejo-Ramos et al., 2014). In this line, the systematic manipulation of the time interval between auditory and visual stimuli may provide relevant information about the temporal course of the metaphorical interaction of emotions onto bodily space. Due to the use of different sensory inputs, a crucial benefit of a cross-modal paradigm lies in the possibility of a careful examination of the time intervals between stimuli from total overlapping to long time delays. In a similar manner, the study of De Vega et al. (2013) made use of a dual-task approach to better capture the time course of the embodied interaction between action-related language comprehension and action performance. Interestingly, their results showed reverse effects of interference (with SOAs around 100–200 ms) and facilitation (with a SOA of 350 ms), depending exclusively on the timing between action-related words and motor responses.

The Present Experiments

The present investigation aimed to study the auditory-visual cross-modal mapping of spoken words onto vertical bodily simulated space. The general method consisted in the prior presentation of a spoken word denoting a positive or negative emotion followed by the display of a visual target whose upper or lower location had to be detected by the participants as soon as possible. It is hypothesized that an interaction between emotion words and the vertical spatial axis may be found in the context of a cross-modal procedure according to a ‘positive-up/negative-down’ embodied metaphor. In particular, this could be owing to a faster detection of upper targets after presenting positive auditory words and lower targets after negative words compared with the other alternative combinations between affective valence and vertical position (i.e., positive-down, negative-up).

The study comprised of two experiments. Experiment 1 was a first attempt to study a possible metaphorical association between emotion and vertical space by means of an auditory to visual cross-modal task. Worth noting, the spoken affective words were passively heard by the participants as they were not required to do any task with these auditory words. Perhaps, this passive procedure was the main reason for the absence of affective and embodied effects found in Experiment 1. For this reason, we decided to introduce a task requiring more active listening of the spoken words than in Experiment 2. Here, two groups of participants carried out different tasks with the auditory words in order to compare an explicit emotion-focused task with a non-emotional activity. In both experiments, the time delay between the auditory and visual stimuli was manipulated in order to explore the temporal course of metaphorical mapping between affect and space.

Experiment 1

The current experiment examined whether auditory infinitive verbs with an affective valence could modulate the response

to a localization task in line with the positive-up, negative-down, vertical spatial metaphor. After playing the auditory files containing the affective words, the participants had to speedily detect the position of a visual target, displayed in either a high or low position on the screen. This task did not require that the auditory stimuli were evaluated in order to test whether mere passive listening could be enough to produce an embodied effect based on a metaphorical conceptual-spatial association.

Method

Participants

Seventeen undergraduate students (12 women and 5 men, $M = 31.6$, $SD = 7.9$, $age_{range} = 19$ –48,) from the *Universidad Nacional de Educación a Distancia* (UNED, Spain) participated in the experiment and received course credits for their participation. The experimental protocol was approved by the Bioethics Committee of the UNED. All of them were native Spanish speakers and reported to have normal or corrected-to-normal vision. Two of the participants were left-handed, and the others were right-handed.

Apparatus and Stimuli

The visual stimuli were displayed on 19-inch LCD-LED color monitors with a screen resolution of 1024×768 pixels, controlled by microcomputers running E-Prime 2.0 software (Psychology Software Tools, 1996–2002). The auditory words were presented through stereo headphones. The visual targets could be displayed in one of two $11.3 \text{ cm} \times 3.0 \text{ cm}$ white (255 RGB) boxes ($10.8^\circ \times 2.9^\circ$ of visual angle), presented 8.0 cm (7.6° of visual angle) above and below the center of the screen (center-to-center). The visual targets were printed in black (0 RGB) and presented against a light gray background (192 RGB; “silver” according to the E-Prime color palette). The masks were made up of a 29×8 matrix checkerboard of black and gray squares (0 and 192 RGB, respectively).

Forty-eight Spanish infinitive verbs denoting emotional states were used. Half of them referred to positive emotions [e.g., *divertir* (to entertain)] and the other half to negative emotions [e.g., *sufrir* (to suffer); see Data Sheet 1]. The verbs were obtained by converting 48 emotional adjectives from Santiago et al.’s (2012) study into their infinitive verbal tense. Twelve additional verbs were selected for the practice block: six positive and six negative. The infinite verbs were spoken by an expert Spanish-speaking female radio announcer in a neutral voice tone and were digitally recorded in a professional radio studio belonging to the UNED’s audiovisual services.

Mean auditory word duration was 640.2 ms ($SD = 137.8$ ms; range = 359–932 ms). There was no significant difference between the mean duration of positive ($M = 622.75$; $SD = 135.95$) and negative words [$M = 657.67$; $SD = 134.43$; $t(46) = -0.88$; $p > 0.10$]. Additional analyses were conducted to compare the number of letters and the frequency of use (according to LEXESP; Sebastián-Gallés et al., 2000) of the positive and negative words. There were no differences in frequency of use [$t(46) = 0.01$, $p > 0.10$] nor in number of letters [$t(46) = -0.82$, $p > 0.10$] between both samples of words.

Procedure and Design

Participants were tested individually in a dimly lit, quiet room. The viewing distance was approximately 60 cm. They were instructed to make their responses as quickly as possible while making the minimum number of errors. Each trial started with the presentation of a 1 cm × 1 cm (0.96° of visual angle) cross-shape fixation mark at the center of the screen and two rectangular boxes above and below the fixation. Participants were instructed to remain fixated on the cross until the completion of the trial. After a variable time period oscillating between 500 and 1,000 ms, randomly selected by the program, an auditory word was presented through the headphones. Participants were instructed to passively listen to the auditory word and wait for the presentation of the visual target. At the end of the auditory word, one of two possible inter-stimulus intervals (ISI; 200 or 350 ms) was previously included to the presentation of the visual target, which consisted of a hash sign (#) printed in black (0 RGB). The visual target was displayed for 200 ms in one of the two boxes and, then, two pattern masks filled in the boxes for 200 ms. The target position was determined at random in each trial but ensured an equal proportion of upper and lower trials in the experiment. Participants were instructed to detect (as fast as possible) the position of the target in the vertical axis by indicating whether the hash sign was displayed in the up or down box. The key response procedure was similar to those used by De Vega et al. (2013). The keys “5,” “2,” and “8” from the right-hand side of the keyboard were assigned as the “resting” key, the “up” key and the “down” key, respectively. The iconic arrows printed on the keys “8” (up arrow) and “2” (down arrow) reinforced the spatial interpretation of the response keys in order to simulate a bodily space. The participants placed the index finger of their dominant hand on the “resting” key until they detected the position of the target by pressing “up” or “down” key. After a maximum time of 2,000 ms to respond, the trial was aborted and a message of “no response, try to respond faster” was shown. There was a practice block and six experimental blocks.

Each experimental block consisted of 96 trials, for a total of 576 experimental trials, whereas the practice block had 48 trials. Feedback was provided only in the practice trials. The experiment lasted about 40–45 min (see **Figure 1A**). After a short break, an unexpected free-recall test of the spoken words was conducted. A sheet of paper was provided and participants were asked to write down as many words from the experiment as possible for 5 min.

The experimental design included three within-subjects factors: emotional valence of the word (positive vs. negative), visual target position (up vs. down), and ISI (200 vs. 350 ms).

Results

Participants responded correctly in 99.5% of all trials (9,741 of 9,792). For the response time (RT) analyses, only correct responses and RTs longer than 200 ms (9,729 of 9,741) were taken into account. The median RT was estimated for each participant in each condition, and these averages were submitted to a parametric ANOVA. The median was chosen since it is an estimator of central tendency robust to outliers (Whelan, 2008). A $2 \times 2 \times 2$ repeated measures ANOVA of the median RTs revealed a main effect of the factor ISI: response were faster (398 ms) after a longer interval between auditory word and visual target compared to shorter interval (407 ms), $F(1,16) = 18.54$, $MSE = 15.45$, $p = 0.001$, $\eta_p^2 = 0.54$. Additionally, a marginally significant effect of visual target position was observed, $F(1,16) = 3.33$, $MSE = 2327.7$, $p = 0.087$, $\eta_p^2 = 0.17$, suggesting a trend to respond faster to upper positions (395 ms) than to lower locations (410 ms). No other main effects or interactions were significant (see **Figure 2**).

An identical analysis was conducted on accuracy rates. No effects or interactions were significant in this case.

Free Recall Test

Missing data for one particular participant was addressed by excluding this particular participant from the analyses. The

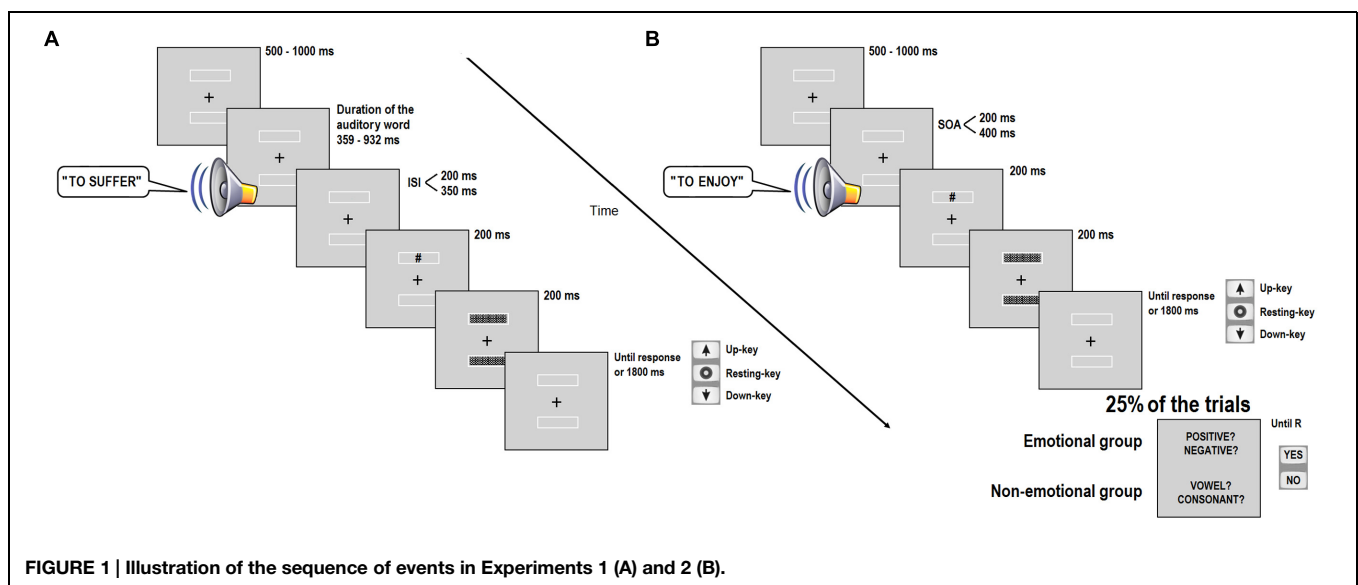
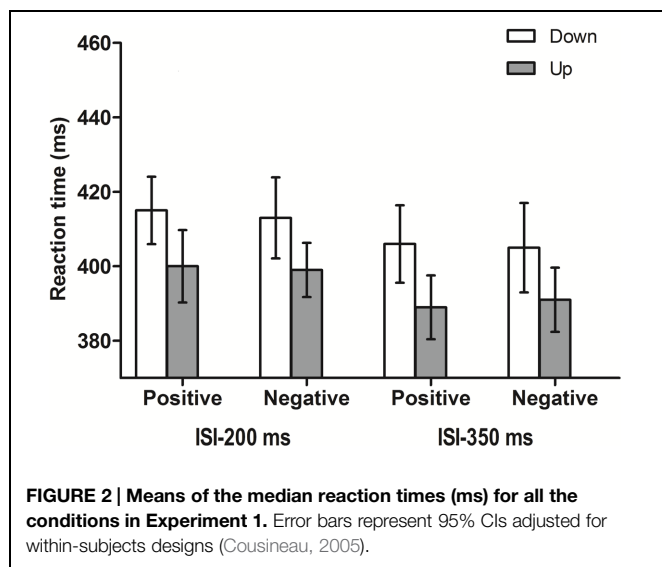


FIGURE 1 | Illustration of the sequence of events in Experiments 1 (A) and 2 (B).



following analyses included data from the remaining sixteen participants. The global mean number of words recalled was 8.31 ($SD = 5.19$; range = 1–20), whereas for positive words 4.00 ($SD = 2.78$) and for negative ones 4.31 ($SD = 3.05$). The global mean number of words correctly recalled was 6.81 ($SD = 4.9$); for positive words 3.25 ($SD = 2.7$) and for negative ones 3.56 ($SD = 2.5$). On average, participants correctly remembered 14.2% ($SD = 10.2$) of all the words presented; 13.5% ($SD = 11.2$) for positive words and 14.8% (10.4) for negative ones. Finally, a conditional proportion correct score was computed for each participant by dividing the number of correct responses by the overall amount of words recalled. The mean conditional correct score was .76 ($SD = 0.27$) with values oscillating between 0.33 and 1.00.

Discussion

The results of this experiment indicated there was no effect in the emotional content of the auditory words on the response to the visual target. The unique reliable effect of ISI seems related to a temporal orienting process in a similar manner to the *foreperiod effect* (see Niemi and Näätänen, 1981; Coull, 2009, for reviews). This effect consists in faster RTs for long vs. short intervals between warning signals (e.g., a brief tone) and the imperative stimulus, provided that the different foreperiod durations are equally distributed and randomly presented (Capizzi et al., 2015). In contrast, the task has not been sensitive either to the emotional meaning of the spoken words or the metaphorical link between concepts and space. A possible reason of these null effects might lie in the passive listening induced by the procedure. Participants did not have to apply any cognitive operation involving the auditory word and, apparently, the exclusive role of the auditory words was acting as a preparatory signal. In this line, the relatively poor performance observed in the free recall test suggests that the participants completely ignored the auditory words. According to the results of a previous visual-to-visual experiment of our group (Marmolejo-Ramos et al., 2014), the interaction of emotion with vertical space might require an explicit use in the affective

content of the word to obtain a reliable behavioral effect. Other studies displaying single visual tasks (Niedenthal et al., 2009; de la Vega et al., 2012) have reported embodied effects of specific emotions but only in the context of emotion-focused processing tasks. Thus, active listening of the auditory word could lead to a deeper processing of its semantic content and, possibly, promote an embodied interaction with the vertical space.

A simpler explanation for the null embodied effects should not be ruled out. That is, are the auditory words selected representative samples of positive and negative affective stimuli? We adapted the adjectives denoting emotional states that Santiago et al. (2012) included in their experiments to infinitive verbs (see Santiago et al., 2012, p. 1059, for a revision of the method used for the selection of the words). However, we did not confirm that these stimuli were emotionally stimulating for our participants and, therefore it might be possible that the verbs did not represent sufficiently polarized affective values. An explicit evaluation of the emotional valence of each was conducted in Experiment 2 to rule out this possible cause.

Another procedural limitation of Experiment 1 has to do with the interval between the spoken words and the visual target. This interval was introduced as an ISI, that is, the time was counted from the end of the digital file until the onset of the visual target. The marked variability of the files' duration (range = 359–932 ms; $M = 640.2$ ms; $SD = 137.8$ ms) might have introduced a confound variable that could make it difficult to stabilize the procedural conditions.

Experiment 2

The results found in Experiment 1 suggested the possibility of studying a cross-modal embodied effect by means of a different task demanding an active listening of the auditory stimuli, as suggested by previous embodied studies. The current experiment aimed to examine the specific conditions under which the conceptual-physical interaction could emerge. An active listening of the spoken words was introduced by means of two different between-subjects tasks: one requiring an explicit task about the positive or negative affective meaning of auditory stimuli (emotional condition), and another task demanding a mere distinction of the first letter of the word as a vowel or a consonant (non-emotional condition).

If an explicit use of the affective content is needed to produce a semantic-spatial interaction, then an embodied interactive effect restricted to the results from the emotional group should be obtained. In contrast, if a mere active listening of the auditory words is required, then, a conceptual-spatial interaction will also be observed in the results from the non-emotional group. Additionally, a more strict control of the time interval between spoken word and visual target by means of a SOA-procedure (i.e., stimulus onset asynchrony) was used in order to avoid the possibility that the variable duration of the auditory files (from 359 to 932 ms) introduced a disturbing effect. Then, two different time intervals were introduced to the current experiment: 200 and 400 ms. Notice that SOA was measured as

time from the onset of the auditory word to the onset of the visual target, causing a partial overlap between both stimuli in most of the trials. However, this procedure does not signify that the visual target appeared before the word was fully processed but only before the word was completely reproduced given that the visual processing requires some time. Thus, it could be assumed that the ending of a process clashes with the beginning of another, making the occurrence of a hypothetical interaction between both cognitive operations easier. Indeed, several classical paradigms inducing semantic interference typically display the target and the distractor/s at the same time, making use of a $SOA = 0$ ms. Examples of these paradigms are flanker task, parafoveal priming or dichotic listening and all of them are well-known experimental procedures to obtain a consistent effect of semantic interference (see Lachter et al., 2004, for a review).

Besides these improvements, a recognition test and an emotional valence evaluation were included at the end of the experimental session. The inclusion of a recognition test aimed to obtain a more sensible indirect measure of the processing level devoted to the auditory words during the experiment. On the other hand, the emotional valence evaluation was included in the experiment in order to reliably measure the affective salience that the auditory word had in our sample of participants.

Method

Participants

Thirty undergraduate students (twenty-one women and nine men, $M_{age} = 24.3$, $SD_{age} = 6.2$, $age_{range} = 19-45$), from the UNED (Spain) participated in the experiment and received course credits for their participation. The experimental protocol was approved by the UNED's Bioethics Committee. All of the participants were native Spanish speakers and reported to have normal or corrected-to-normal vision. Two of the participants were left-handed, and the others were right-handed. The participants were randomly assigned to two groups of 15 individuals each; an emotional group ($M_{age} = 24.3$, $SD_{age} = 3.5$), and non-emotional group ($M_{age} = 24.3$, $SD_{age} = 8.2$).

Apparatus and Stimuli

The stimuli and apparatus were identical to those of Experiment 1, with the exception of a new set of 48 infinitive verbs selected as “new” distracter items for the recognition test (24 positive and 24 negative). This new set of infinitive verbs (both positive and negative) were synonyms of, and were matched in length to, those used in Experiment 1.

Procedure and Design

The procedure was very similar to Experiment 1 but included several relevant changes. In the current experiment, the temporal interval between the auditory word and the visual target was manipulated by a SOA instead of an ISI, in order to control this variable among the trials, independently of the different duration of the digital audio files. The main task of the participants was the same as Experiment 1, that is, to detect as soon as possible the location of the visual target on the vertical axis. The same keys as Experiment 1 were used (see Figure 1B).

A crucial manipulation was related to the different instructions provided for both experimental subgroups. In the emotional group, participants were instructed to carefully listen to the auditory word and judge the emotional valence of the verb as either negative or positive, with the aim of correctly responding to the retrospective question that could be displayed at the end of the trial. In the non-emotional group, participants had to identify whether the first letter of the word was a vowel or a consonant, also with the aim of answering the retrospective question. Retrospective questions were randomly distributed in 25% of the trials so participants could not predict their appearance. Here, a word was displayed in the middle of the screen (e.g., ‘POSITIVE’ for the emotional group or ‘VOWEL’ for the non-emotional group) and the observers had to respond ‘yes’ or ‘no’ by pressing one of two available keys (‘1’ and ‘2’ keys in the top row of numbers with stickers indicating “SÍ”/yes and “NO”) without time response demand and with a different hand than had been used in the main task. There was a practice block and six experimental blocks. Each experimental block consisted of 96 trials, for a total of 576 experimental trials, whereas the practice block had 24 trials. This part of the experiment lasted about 45–50 min.

After the main task was finished, a free-recall task was conducted. The participants had to remember as many auditory words as possible during a 5 min period by writing them on a sheet of paper. Then, participants carried out a computerized recognition task. Ninety-six infinitive, positive or negative, verbs (48 ‘old,’ and 48 ‘new’) were randomly displayed on the screen (one word per trial) and participants judged whether the word was heard during the main task by clicking the mouse over the button containing the chosen response (‘YES’ or ‘NO’) without any time constraints. After the presentation of each word, participants had to make a self-paced confidence judgment of their recognition memory. They indicated their confidence in having listened to the presented word by pressing one of eleven response keys from “0” to “10.” A “10” response indicated that they were completely sure of their response, whereas a “0” response indicated that they were completely unsure of the response. Finally, a valence emotional rating task of the 48 auditory words was administered by computer. Participants rated the emotional valence of the words on a 9-point rating scale from -4 (extremely negative) to $+4$ (extremely positive), considering zero as a neutral value. On the screen, together with the word, nine squares with digits inside from -4 to $+4$ were displayed on the screen. Participants made their ratings by clicking the mouse over the square containing the chosen number without RT demands.

The experiment resulted in a mixed design with one between-subjects factor (emotional vs. non-emotional groups) and three within-subject factors; emotional valence of the word (positive vs. negative), visual target position (up vs. down), and SOA (200 ms vs. 400 ms).

Results and Discussion

Retrospective Question Task

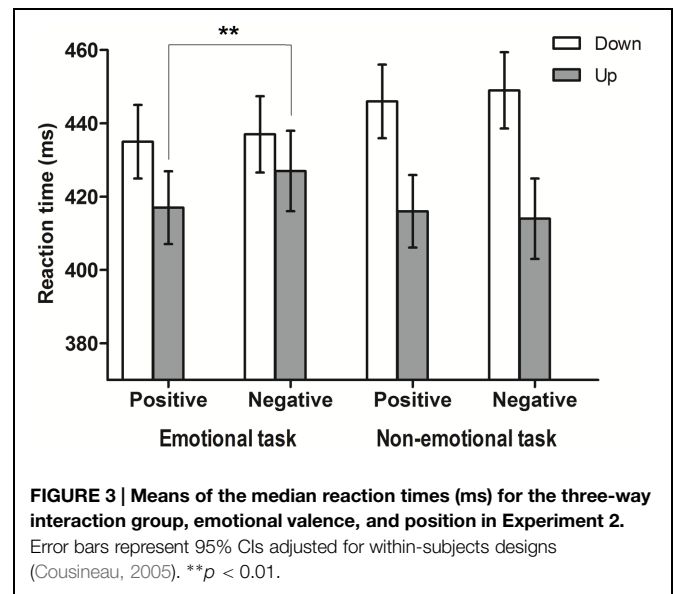
Performance on the retrospective question trials was high ($M = 97\%$; $SD = 3.4\%$; range = 87–100%; emotional group:

$M = 96.87\%$; $SD = 3.6\%$; non-emotional group: $M = 97.27\%$; $SD = 3.3\%$). A one-factor between-subjects ANOVA intended to rule out differences in hit rates between the emotional and non-emotional group showed that there were no differences; $F < 1$.

Reaction Times

Regarding the main task, participants responded correctly in 97.6% of all trials (16,870 of 17,280). For the RT analyses, only correct responses and RTs longer than 200 ms (16,828 of 16,870) were considered. As in Experiment 1, the median RT was computed for each participant in each condition, and these averages were submitted to an ANOVA. A $2 \times (2 \times 2 \times 2)$ mixed ANOVA of the RTs showed main effects of all three within-subjects factors: responses were faster with positive auditory words (429 ms) than with negative words [432 ms; $F(1,28) = 5.71$, $MSE = 102.06$, $p = 0.024$, $\eta_p^2 = 0.17$], after a SOA of 400 ms (411 ms) compared with a SOA of 200 ms [449 ms; $F(1,28) = 95.8$, $MSE = 897.9$, $p < 0.001$, $\eta_p^2 = 0.77$], and with visual targets displayed in the upper position (419 ms) compared to the lower location [442 ms; $F(1,28) = 11.54$, $MSE = 2809.7$, $p = 0.002$, $\eta_p^2 = 0.29$]. Interestingly, there was no main effect of the between-subjects factor group ($F < 1$) showing similar global RTs in both groups. The interaction between group and emotional valence of the word was significant [$F(1,28) = 5.26$, $MSE = 537$, $p = 0.03$, $\eta_p^2 = 0.21$], showing that the speeding-up effect of the positive words respect to negative words was reliable in the emotional group ($\Delta 6$ ms) but not in the non-emotional group ($\Delta 0$ ms). There was also a significant interaction between visual target position and SOA [$F(1,28) = 13.2$, $MSE = 220.7$, $p = 0.001$, $\eta_p^2 = 0.30$] suggesting a multiplier effect of the joining together of the longer SOA and the upper position that leads to an even faster response in this condition (396 ms, $\Delta 30$ ms compared to longer SOA and lower position) than those in the shorter SOA and upper position (441 ms, $\Delta 16$ ms compared to shorter SOA and lower position).

Interestingly, the main effect of the emotional valence of the word, as well as its interaction with the factor group, support a semantic effect of the affective meaning on the response only reliable for the emotional group (which is in contrast with the null effect of this factor in Experiment 1). The effect of SOA replicates the result of the factor ISI of Experiment 1 and may also be based on a temporal orienting effect. Remarkably, the critical effect for the purposes of the present work is a significant three-way interaction between the factors group, emotional valence of the words and visual target position [$F(1,28) = 4.4$, $MSE = 114.8$, $p = 0.045$, $\eta_p^2 = 0.16$]. *Post hoc* comparisons using Bonferroni correction showed that the emotional group detected the target faster in the upper position after a positive word (417 ms) compared to a negative one (427 ms; $p = 0.004$). In contrast, no differences in the non-emotional group between those conditions were observed (416 ms vs. 413 ms; $p = 0.454$). No significant effect of the emotional valence of the word on the responses to lower positions was observed, neither in the emotional group (positive word: 435; negative word: 437 ms; $p = 0.363$) nor the non-emotional group (positive word: 446; negative word: 448 ms; $p = 0.454$; see Figure 3).



Lastly, a three-way interaction including emotional valence \times target position \times SOA was significant too [$F(1,28) = 9.44$, $MSE = 60.4$, $p = 0.005$, $\eta_p^2 = 0.25$], indicating that the combination of a long SOA, positive valence and upper position generated an even faster response (393 ms) compared with the response to a short SOA, negative valence and upper position (400 ms; $p = 0.041$).

Accuracy Rates

Identical analyses were conducted on accuracy rates. The $2 \times (2 \times 2 \times 2)$ mixed ANOVA of the hit rates only revealed a significant interaction effect between visual target position and SOA [$F(1,28) = 4.56$, $MSE < 0.001$, $p < 0.05$, $\eta_p^2 = 0.14$]. However, pair-wise comparisons applying Bonferroni correction did not indicate any significant effect (all $p > 0.05$).

Free Recall Test

The mean number of words evoked for the emotional group was 11.94 ($SD = 4.11$; range = 7–19), whereas, for the non-emotional group 11.4 ($SD = 5.17$; range = 4–22). The mean number of words correctly evoked for the emotional group was 10.13 ($SD = 3.2$) and, for the non-emotional group 7.8 ($SD = 4.66$). On average, participants of the emotional group correctly remembered 21.1% ($SD = 6.7$) of all the words presented; whereas, the non-emotional group correctly remembered 16.25% ($SD = 9.7$). A comparison between the group from Experiment 1 and the two groups from the Experiment 2 was carried out by means of a mixed 3×2 ANOVA with the factors group and emotional valence of the word, as well as the proportion of items correctly evoked as dependent measure. Neither significant main effects nor interactions between factors reached statistical significance (all $p > 0.05$). A conditional proportion correct score was obtained for each participant. The emotional group obtained an average of 0.86 ($SD = 0.10$), whereas the non-emotional group showed an average of 0.67 ($SD = 0.23$). A one-factor between-subjects ANOVA with the factor group and

the conditional correct score as measure suggested differences between the groups [$F(2,46) = 3.12, p = 0.054$]. However, *post hoc* comparisons did not detect any significant pair-wise comparisons (all $p > 0.05$).

Recognition Test

A direct measure of word recognition (d') was calculated for each participant. The measures were obtained by treating “old” words as signal and “new” words as noise. The individual d' values ranged between 1.18 and 2.54, and the overall mean was 1.65 (SD = 0.19). A mixed 2×2 ANOVA with the factors group (emotional vs. non-emotional) and emotional valence of the word and d' values as the dependent variable was conducted. Only the between-subject factor group showed a significant effect [$F(1,24) = 19.72, p < 0.001$], in that there was a better recognition rate in the emotional group ($d' = 2.02$) than in the non-emotional sample ($d' = 1.43$). Neither a difference between positive and negative words nor an interaction between the two factors was found (both $F < 1$).

To analyse the data from the recognition confidence judgments, a mixed $2 \times (2 \times 2)$ ANOVA with the factors group (emotional vs. non-emotional), emotional valence (positive vs. negative) and the type of word (“new” vs. “old”) was conducted on the mean rating of the confidence scores. A significant difference between groups was observed [$F(1,25) = 6.5$, MSE = 3.98, $p < 0.05$, $\eta_p^2 = 0.21$], in the sense of a higher confidence in the emotional group ($M = 7.79$; SD = 1.01) with respect to the non-emotional group ($M = 6.81$; SD = 0.99). Additionally, the “old” stimuli had significantly more confident judgments ($M = 7.77$; SD = 1.07) than the “new” verbs [$M = 6.80$; SD = 1.25; $F(1,25) = 41.3$, MSE = 0.616, $p < 0.001$, $\eta_p^2 = 0.62$]. Neither a difference between positive and negative words nor any interaction between the factors was found (all $F < 1.2$).

Emotional Valence Evaluation

Mean valence rating for each word was averaged (see Data Sheet 1). The mean rating for the positive words was 2.78 (SD = 0.61) and for the negative ones -2.61 (SD = 0.67), showing a clear polarized difference between them. Two single-sample t -tests showed that these mean ratings were significantly different from the neutral score of zero both for positive words [$t(23) = 22.5, p < 0.001$] and negative items [$t(23) = -19.12, p < 0.001$]. Additionally, a single-sample t -test comparing the absolute values (or modulus) of positive and negative words was insignificant, $t(23) = 0.88, p = 0.39$, suggesting that both categories of words are polarized to a similar degree. When comparing the ratings provided by both subject groups, more extreme responses were offered by the emotional group than the non-emotional, both for positive words [$M = 3.00$ and $M = 2.57$, respectively; $t(23) = 6.12, p < 0.001$] and for negative words [$M = -2.76$ and $M = -2.49$, respectively; $t(23) = -4.5, p < 0.001$]. Interestingly, there was no difference in a global comparison of mean valence between both groups including all the words [$t(47) = 1.19, p > 0.10$], which suggests that the polarized responses from the emotional group were mutually compensated.

Discussion

Strikingly, the pattern of RTs supports an interaction between affective activation and the visual vertical axis that is reliable only when the participants were engaged in an emotion-focused task. These results can be interpreted as supporting an association between emotional concepts and the physical vertical axis but only when the instructions demand an explicit decision on the emotional valence of the spoken words or, at least, a processing of the meaning of the word. The absence of this conceptual-spatial mapping in the results from the non-emotional group suggests that an active scrutiny of the words is not enough to generate an embodied interaction. Two possible requirements could be considered in addition to the active listening of the word to obtain an interactive effect between affect and space. On the one hand, a direct handling of the emotional meaning of the words, on the other hand, a deeper level of processing of the information (in terms of Craik and Lockhart, 1972). The results from the recognition task showed a typical level-of-processing effect since a deep processing (i.e., semantic processing) leads to a more robust memory trace as well as higher confidence judgments, while shallow processing (i.e., phonemic or orthographic analysis) results in a more fragile memory and lower confidence in recognition. However, the current design is not qualified to disentangle between both alternatives.

Remarkably, the lack of a main effect of the emotional valence (for the non-emotional group) suggests that the participants may have indeed failed to interpret the meaning of the words. This result contrasts with findings from previous studies that did not ask observers to explicitly judge emotional valence but still found behavioral effects linked to emotionally charged stimuli (e.g., Eastwood et al., 2003; Harris and Pashler, 2004; Estes and Adelman, 2008)³. However, it should be taken into account that most of the previous findings have been obtained by visual stimulation (affecting visual processing too; e.g., Harris and Pashler, 2004). In our case, the auditory nature of the affective stimuli especially the cross-modal interaction with another sensory modality could have diminished the usual effects given by other procedures.

Similar to Experiment 1, while the SOA exerted a main effect over the responses based on a temporal preparation, this factor did not modulate the interaction between emotion and spatial axis. Longer intervals between stimuli should be implemented in future experiments to carefully explore the temporal requirements of this conceptual-physical interaction. Regarding this point, an important element of our procedure must be considered, i.e., the retrospective question relating to the spoken word at the end of the sequence of events (in 25% of the trials, strictly speaking). The introduction of such a question forced participants to maintain in working memory (WM) an active representation of the relevant information extracted from the auditory stimuli. This procedure is similar to prior studies showing that the contents of WM can exert an influence over the deployment of attention in visual search tasks arising with asynchronies ranging from 200 to 4000 ms (see Soto et al., 2008, for a revision). Importantly, the WM effects on search were

³We thank one of the reviewers for suggesting this point.

absent when observers were merely exposed to the memory cue without a later report (Soto et al., 2005; Olivers et al., 2006), similar to our Experiment 1. Accordingly, it might be considered that the embodied interaction observed here could be due to the active maintenance in WM of emotionally valenced information irrespective of the specific interval between stimuli (although see Xie et al., 2015). Crucially, this possibility should be taken into account for future research on this topic.

General Discussion

In the context of within-modal visual tasks, previous studies have reported evidence for an association between emotionally valenced concepts and the vertical as well as horizontal space (e.g., Meier and Robinson, 2004; de la Vega et al., 2012; Santiago et al., 2012; Marmolejo-Ramos et al., 2014; Damjanovic and Santiago, 2015). Other lines of research have studied the influence of emotions activated by other sensory modalities different from visual system (e.g., Tajadura-Jiménez et al., 2010; Rantala et al., 2013) and, even, the interaction between different perceptual modalities during the processing of affective stimuli (e.g., Gerdes et al., 2014; Velasco et al., 2014). Nevertheless, to our knowledge, no previous work has tackled the study of cross-modal interactions between emotionally valenced concepts and bodily space from an embodied standpoint. The current research is the first step toward the study of the metaphorical mapping of emotions onto vertical space by means of an auditory to visual cross-modal paradigm.

Experiment 1 failed to observe such a cross-modal embodiment suggesting that passive listening to the conceptual stimuli was not enough to generate a bias in the detection of the visual target. The participants were not assigned any task related to the spoken words and this absence of cognitive analysis might have led to a null bias of the affective load on the response. Previous studies have provided evidence for the necessary explicit use of the semantic information to observe the embodiment of specific emotions (Niedenthal et al., 2009; de la Vega et al., 2012; Marmolejo-Ramos et al., 2014). On the basis of these findings, we introduced active listening of the spoken words in Experiment 2, by including two different tasks that were applied to different participant samples. Here, the results did show a mapping between emotions induced by auditory words and the vertical space involved in the location detection task of a visual stimulus. This result consisted of a faster detection of the target in the upper position after positive words compared with those after negative words. In contrast, the responses to target displayed at the lower position were not sensible to the different emotional content of the spoken words. This result has similar precedents in previous related research. Recently, Marmolejo-Ramos et al. (2014; Experiment 2) have reported a reliable priming effect of the emotional valence of sentences representing emotional contexts on the processing of visual probes at the upper position, which was not observed for lower positions (see also Xie et al., 2015). In the same direction, the metaphorical congruency effects between affect and vertical space found by Meier and Robinson (2004) and Santiago et al.

(2012; see Experiments 1 and 3) showed a higher effect size (in the sense of a higher difference between mean RTs of congruent and incongruent trials) at an upper than lower location; although it is true that the embodied congruency effect was also significant at lower positions, in contrast to our findings. Interestingly, all the three cited studies observed a significant main effect of position, showing global faster responses to targets displayed at upper than lower locations, thus being congruent with our work.

Crucially, the cross-modal embodied interaction found in the present study was restricted to the participants that carried out a semantic emotion-focused analysis of the auditory information. Taken together, the results of our experiments suggest that the association of affective valence and vertical space is not activated automatically during speech processing. However, the exact nature of the task needed to obtain the embodied effect cannot be distinctly established with our experimental design. The emotional group performed a valence-decision task while the other group had to identify the first letter of the spoken word for which emotional content was irrelevant. Notice that the application of an emotional-based criterion was not the exclusive difference due to both between-subjects conditions. That is, an evident divergence regarding the level of processing between a semantic versus a phonemic analysis was presented without a choice to separate them considering the present results. Undoubtedly, this crucial issue should be examined in future research.

The current study is a cross-modal task in that auditory stimuli preceded the presentation of visual items. However, the relation between the auditory and visual stimuli was metaphorical in that, as previous research shows (e.g., Marmolejo-Ramos et al., 2013), emotion concepts can be represented in bodily space. Thus, the task used herein is in fact a cross-modal metaphorical-mapping task. Although the results indicate such mapping seems to occur only when an explicit evaluation of the stimuli is required, it does not exclude that task-dependent factors could lead to different results. It might be the case, for example, that an implicit association did not occur simply because the horizontal (i.e., over the left and right ear) presentation of the auditory stimuli did not facilitate mapping onto the visual vertical plane. Thus, a task in which the location source of the auditory emotional stimuli matches visual vertical spatial locations (see e.g., Spence et al., 2001) could be instrumental in further studying cross-modal metaphorical mapping. By the same token, it would be informative to know whether the mapping holds the other way around. That is, would the cross-modal mapping hold when visual emotional stimuli precede the location of auditory sources onto space? Note that in this study only two sensory modalities are being considered. Hence, the cross-modal processing occurring among these and other modalities (i.e., tactile, olfactory, and gustative) need to be investigated in the context of emotions and metaphorical mapping.

From a theoretical perspective, merely tentative, the current results fit better with the restrained embodiment theories described by Meteyard et al. (2012), such as “secondary embodiment” (sensory and motor system are independent but

directly associated) or “weak embodiment” (partial dependence in the relationship to sensory-motor systems). In this line, our results are compatible with an activation of both sensorial and motor systems when a semantic representation is explicitly accessed and opposed to previous findings supporting an automatic, incidental valenced-induced activation of spatial features (e.g., Meier and Robinson, 2004; Gozli et al., 2013; but see Brookshire et al., 2010). Note that the access to the semantic representation may be modulated by the definition of semantic processing and what counts as deeper or more “explicit processes” in comparison to shallow processing required by the task. The distinction between weak and strong theories has also been related to current theories about the embodiment of emotion concepts and explicitly describes how the current findings would confirm balanced/moderate versions of the embodiment hypothesis. In this context, a crucial issue is related to the conceptualization of automaticity. In contrast with the traditional view of automaticity as a dichotomous *all-or-nothing* variable, Brookshire et al. (2010) investigated to what extent is the activation of embodied representations automatic. Remarkably, our procedure could be useful to explore the limits of automaticity in the occurrence of space-valence congruity effects. The inclusion of retrospective measures of memory of the valenced stimuli provide us with an indirect measure of the degree or level of processing devoted to the valenced words, which might be correlated to the effect size of the embodied effects obtained. In Experiment 2, a better recognition of the auditory words in the *emotional group* is compatible with a deeper processing of the valenced stimuli that, at the same time, is correlated with a significant space-valence interaction. However, it is possible that the small sample size in the current study, and hence lack of power, could have hidden this specific kind of embodiment phenomena.

It might be the case that cross-modal metaphorical mapping needs mild to low embodiment and neuropsychological and neuroimaging research could be instrumental in determining the timing and brain localization of this type of cross-modality (be it real or conceptual). In regards to the localization, it could be entertained that metaphorical mapping onto bodily space could be processed in the left hemisphere hippocampus as this area is known for dealing with information, mainly linguistic, that feeds into the generation of semantic spaces (see O’Keefe and Nadel, 1979). Indeed, some entorhinal cortex activation could be expected since this area deals with the representation of position, direction, and velocity (Sargolini et al., 2006). In other words, if entorhinal and hippocampal structures aid in the representation of space (see Moser et al., 2008), it is thus tenable that these structures play some role in the representation of metaphorical mappings onto bodily space. We believe that most of the metaphorical processing could be handled by these areas; however, as these areas project to the neocortex, and vice versa via perirhinal and parahippocampal cortex, some mild activation of sensorimotor and somatosensory cortical areas could be observed. This speculation leads us to believe that the processing of cross-modal metaphorical mapping might need mild to low levels of

embodiment. Nonetheless, this conjecture is yet to be empirically investigated.

The current work examined the association between affect and vertical space by using a cross-modal procedure. The auditory stimuli selected for our study were spoken words denoting an emotion. Interestingly, for future research, it might be relevant to include emotion sounds (e.g., grunts, sighs, screams) or even to manipulate the prosody of the spoken words in an affective fashion. Such a novelty would provide us with a more direct test of the cross-modal interactions between affect and spatial location⁴. An important advantage of our procedure is that it allows manipulation, in a completely independent manner, of the timing of the visual and auditory stimulation in order to explore the temporal requirements of a metaphorical mapping between emotion and bodily space. The visual and auditory stimuli can be displayed simultaneously or with different SOAs or ISIs. Another potential innovation would be the introduction of a dichotic listening procedure in order to manipulate the extent of cognitive resources devoted to the auditory items. Undoubtedly, such a procedural improvement will serve as an important step in the study of the role of attention, level of processing, and the limits of automaticity in the occurrence of interactive effects between affect and bodily space.

Conclusion

Our study is a first step toward the study of a cross-modal metaphorical mapping of emotions onto vertical space. The results obtained show that (i) a cross-modal association of affective valence and vertical space is possible but that (ii) this embodied association is not activated automatically because (iii) an explicit evaluation of the emotionally valenced words is needed to observe an interaction between emotion concepts and bodily simulated space.

Acknowledgments

This work was supported by grants EDU2013-46437-R from the Ministerio de Economía y Competitividad (MINECO) of Spain and 2012V/PUNED/0009 from the UNED. We would like to thank Maite Pérez de Albéniz, from the audiovisual services of the UNED, for her kind collaboration in the recording of the auditory material. The authors thank Rosie Gronthos, Joanna Lindström, and María Fernández Cahill for proofreading this manuscript. FM-R dedicates this paper to the memory of Javier Emiro Sánchez Ramos.

Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2015.01205>

⁴We thank one of the reviewers for suggesting this interesting innovation for future research.

References

- Altarriba, J., and Bauer, L. M. (2004). The distinctiveness of emotion concepts: a comparison between emotion, abstract, and concrete words. *Am. J. Psychol.* 117, 389–410. doi: 10.2307/4149007
- Ansorge, U., and Bohner, G. (2013). Investigating the association between valence and elevation with an implicit association task that requires upward and downward responding. *Univ. Psychol.* 12, 1453–1471. doi: 10.11144/Javeriana.UPSY12-5.iave
- Ansorge, U., Khalid, S., and König, P. (2013). Space-valence priming with subliminal and supraliminal words. *Front. Psychol.* 4:81. doi: 10.3389/fpsyg.2013.00081
- Brookshire, G., Ivry, R., and Casasanto, D. (2010). “Modulation of motor-meaning congruity effects for valenced words,” in *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, eds S. Ohlsson and R. Catrambone (Austin, TX: Cognitive Science Society), 1940–1945.
- Brosch, T., Grandjean, D., Sander, D., and Scherer, K. R. (2008). Behold the voice of wrath: cross-modal modulation of visual attention by anger prosody. *Cognition* 106, 1497–1503. doi: 10.1016/j.cognition.2007.05.011
- Brosch, T., Grandjean, D., Sander, D., and Scherer, K. R. (2009). Cross-modal emotional attention: emotional voices modulate early stages of visual processing. *J. Cogn. Neurosci.* 21, 1670–1679. doi: 10.1162/jocn.2009.21110
- Capizzi, M., Correa, A., Wojtowicz, A., and Rafal, R. D. (2015). Foreperiod priming in temporal preparation: testing current models of sequential effects. *Cogn.* 134, 39–49. doi: 10.1016/j.cognition.2014.09.002
- Coull, J. T. (2009). Neural substrates of mounting temporal expectation. *PLoS Biol.* 7:e1000166. doi: 10.1371/journal.pbio.1000166
- Cousineau, D. (2005). Confidence intervals in within-subjects designs: a simpler solution to Loftus and Masson's method. *Tutor. Quant. Methods Psychol.* 1, 42–45.
- Craik, F. I. M., and Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *J. Verbal Learning Verbal Behav.* 11, 671–684. doi: 10.1016/S0022-5371(72)80001-X
- Dagaev, N. I., and Terushkina, Y. I. (2014). Conceptual knowledge of emotions includes somatosensory component: evidence from modality-switch cost effect. *J. Cogn. Psychol.* 26, 322–332. doi: 10.1080/20445911.2014.892111
- Damjanovic, L., and Santiago, J. (2015). Contrasting vertical and horizontal representations of affect in emotional visual search. *Psychon. Bull. Rev.* doi: 10.3758/s13423-015-0884-6 [Epub ahead of print].
- de la Vega, I., Filippis, M., Lachmair, M., Dudschig, C., and Kaup, B. (2012). Emotional valence and physical space: limits of interaction. *J. Exp. Psychol.* 38, 375–385. doi: 10.1037/a0024979
- De Vega, M., Moreno, V., and Castillo, D. (2013). The comprehension of action-related sentences may cause interference rather than facilitation on matching actions. *Psychol. Res.* 77, 20–30. doi: 10.1007/s00426-011-0356-1
- Dudschig, C., de la Vega, I., and Kaup, B. (2015). What's up? Emotion-specific activation of vertical space during language processing. *Acta Psychol.* 156, 143–155. doi: 10.1016/j.actpsy.2014.09.015
- Eastwood, J. D., Smilek, D., and Merikle, P. M. (2003). Negative facial expression captures attention and disrupts performance. *Percept. Psychophys.* 65, 352–358. doi: 10.3758/BF03194566
- Estes, Z., and Adelman, J. S. (2008). Automatic vigilance for negative words is categorical and general. *Emotion* 8, 453–457. doi: 10.1037/a0012887
- 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
- Gerdes, A. B. M., Wieser, M. J., and Alpers, G. W. (2014). Emotional pictures and sounds: a review of multimodal interactions of emotion cues in multiple domains. *Front. Psychol.* 5:1351. doi: 10.3389/fpsyg.2014.01351
- Gozli, D. G., Chow, A., Chasteen, A. L., and Pratt, J. (2013). Valence and vertical space: saccade trajectory deviations reveal metaphorical spatial activation. *Visual Cogn.* 21, 628–646. doi: 10.1080/13506285.2013.815680
- Harris, C. R., and Pashler, H. (2004). Attention and the processing of emotional words and names not so special after all. *Psychol. Sci.* 15, 171–178. doi: 10.1111/j.0956-7976.2004.01503005.x
- Harrison, N. R., and Davies, S. J. (2013). Modulation of spatial attention to visual targets by emotional environmental sounds. *Psychol. Neurosci.* 6, 247–251. doi: 10.3922/j.psns.2013.3.02
- Lachter, J., Forster, K. I., and Ruthruff, E. (2004). Forty-five years after Broadbent (1958): still no identification without attention. *Psychol. Sci.* 111, 880–913. doi: 10.1037/0033-295x.111.4.880
- Lakoff, G. (2014). Mapping the brain's metaphor circuitry: metaphorical thought in everyday reason. *Front. Hum. Neurosci.* 8:958. doi: 10.3389/fnhum.2014.00958
- Marmolejo-Ramos, F., and Dunn, J. (2013). On the activation of sensorimotor systems during the processing of emotionally-laden stimuli. *Universitas Psychol.* 12, 1511–1542. doi: 10.11144/Javeriana.UPSY12-5.assp
- Marmolejo-Ramos, F., Elosúa, M. R., Yamada, Y., Hamm, N. F., and Noguchi, K. (2013). Appraisal of space words and allocation of emotion words in bodily space. *PLoS ONE* 8:e81688. doi: 10.1371/journal.pone.0081688
- Marmolejo-Ramos, F., Hellemans, K., Comeau, A., Heenan, A., Faulkner, A., Abizaid, A., et al. (2015). ERP signatures to perceived and imagined emotional and real-life photos. *Neurosci. Bull.* 31, 317–330. doi: 10.1007/s12264-014-1520-6
- Marmolejo-Ramos, F., Montoro, P., Elosúa, M. R., Contreras, M. J., and Jiménez, W. A. (2014). The activation of representative emotional verbal contexts interacts with vertical spatial axis. *Cogn. Process.* 15, 253–267. doi: 10.1007/s10339-014-0620-6
- Meier, B. P., Moller, A. C., Chen, J. J., and Riemer-Peltz, M. (2011). Spatial metaphor and real estate: north-south location biases housing preference. *Soc. Psychol. Personal. Sci.* 2, 547–553. doi: 10.1177/1948550611401042
- Meier, B. P., and Robinson, M. D. (2004). Why the sunny side is up: associations between affect and vertical position. *Psychol. Sci.* 15, 243–247. doi: 10.1111/j.0956-7976.2004.00659.x
- Meteyard, L., Rodríguez, S., 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
- Moser, E. I., Kropff, E., and Moser, M.-B. (2008). Place cells, grid cells, and the brain's spatial representation system. *Annu. Rev. Neurosci.* 31, 69–89. doi: 10.1146/annurev.neuro.31.061307.090723
- 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
- Niedenthal, P. M., Winkielman, P., Mondillon, L., and Vermeulen, N. (2009). Embodied emotion concepts. *J. Pers. Soc. Psychol.* 96, 1120–1136. doi: 10.1037/a0015574
- Niemi, P., and Näätänen, R. (1981). Foreperiod and simple reaction time. *Psychol. Bull.* 89, 133–162. doi: 10.1037/0033-2909.89.1.133
- O'Keefe, J., and Nadel, L. (1979). Précis of O'Keefe & Nadel's The hippocampus as a cognitive map. *Behav. Brain Sci.* 2, 487–533. doi: 10.1017/S0140525X00063949
- Olivers, C. N. L., Meijer, F., and Theeuwes, J. (2006). Feature-based memory-driven attentional capture: visual working memory content affects visual attention. *J. Exp. Psychol.* 32, 1243–1265. doi: 10.1037/0096-1523.32.5.1243
- Oosterwijk, S., Lindquist, K. A., Anderson, E., Dautoff, R., Moriguchi, Y., and Barrett, L. F. (2012). States of mind: emotions, body feelings, and thoughts share distributed neural networks. *Neuroimage* 62, 2110–2128. doi: 10.1016/j.neuroimage.2012.05.079
- Pecher, D., Zeelenberg, R., and Barsalou, L. W. (2003). Verifying different-modality properties for concepts produces switching costs. *Psychol. Sci.* 14, 119–124. doi: 10.1111/1467-9280.t01-1-01429
- Rantala, J., Salminen, K., Raisamo, R., and Surakka, V. (2013). Touch gestures in communicating emotional intention via vibrotactile stimulation. *Internat. J. Hum. Comput. Stud.* 71, 679–690. doi: 10.1016/j.ijhcs.2013.02.004
- Santiago, J., Ouellet, M., Román, A., and Valenzuela, J. (2012). Attentional factors in conceptual congruency. *Cogn. Sci.* 36, 1051–1077. doi: 10.1111/j.1551-6709.2012.01240.x
- Sargolini, F., Fyhn, M., Hafting, T., McNaughton, B. L., Witter, M. P., Moser, M.-B., et al. (2006). Conjunctive representation of position, direction, and velocity in entorhinal cortex. *Science* 312, 758–762. doi: 10.1126/science.1125572
- Sasaki, K., Seno, T., Yamada, Y., and Miura, K. (2012). Emotional sounds influence vertical vection. *Perception* 41, 875–877. doi: 10.1068/p7215
- Sasaki, K., Yamada, Y., and Miura, K. (2015). Post-determined emotion: motor action retrospectively modulates emotional valence of visual images. *Proc. R. Soc. B Biol. Sci.* 282:20140690. doi: 10.1098/rspb.2014.0690
- Scherer, K. R. (2005). What are emotions? And how can they be measured? *Soc. Sci. Inform.* 44, 695–729. doi: 10.1177/0539018405058216

- Sebastián-Gallés, N., Martí, M., Carreiras, M., and Cuetos, F. (2000). *LEXESP: Léxico Informatizado del Español*. Barcelona: Universitat de Barcelona.
- Soto, D., Heinke, D., Humphreys, G. W., and Blanco, M. J. (2005). Early, involuntary top-down guidance of attention from working memory. *J. Exp. Psychol.* 31, 248–261. doi: 10.1037/0096-1523.31.2.248
- Soto, D., Hodsoll, J., Rotshtein, P., and Humphreys, G. W. (2008). Automatic guidance of attention from working memory. *Trends Cogn. Sci.* 12, 342–348. doi: 10.1016/j.tics.2008.05.007
- Spence, C., Nicholls, M. E. R., and Driver, J. (2001). The cost of expecting events in the wrong sensory modality. *Percept. Psychophys.* 63, 330–336. doi: 10.3758/BF03194473
- 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
- Velasco, C., Balboa, D., Marmolejo-Ramos, F., and Spence, C. (2014). Cross-modal effect of music and odor pleasantness on olfactory quality perception. *Front. Psychol.* 5:1352. doi: 10.3389/fpsyg.2014.01352
- Whelan, R. (2008). Effective analysis of reaction time data. *Psychol. Record* 58, 475–482.
- Wilson, N. L., and Gibbs, R. W. (2007). Real and imagined body movement primes metaphor comprehension. *Cogn. Sci.* 31, 721–731. doi: 10.1080/15326900701399962
- Wilson-Mendenhall, C. D., Barrett, L. F., and Barsalou, L. W. (2013). Situating emotional experience. *Front. Psychol.* 7:764. doi: 10.3389/fnhum.2013.00764
- Xie, J., Huang, Y., Wang, R., and Liu, W. (2015). Affective valence facilitates spatial detection on vertical axis: shorter time strengthens effect. *Front. Psychol.* 6:277. doi: 10.3389/fpsyg.2015.00277
- Xie, J., Wang, R., and Chang, S. (2014). The mechanisms of valence-space metaphors: ERP evidence for affective word processing. *PLoS ONE* 9:e99479. doi: 10.1371/journal.pone.0099479
- Zeelenberg, R., and Bocanegra, B. R. (2010). Auditory emotional cues enhance visual perception. *Cognition* 115, 202–206. doi: 10.1016/j.cognition.2009.12.004

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.

Copyright © 2015 Montoro, Contreras, Elosúa and Marmolejo-Ramos. 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.

Cross-modal associations between materic painting and classical Spanish music

Liliana Albertazzi^{1,2*}, Luisa Canal³ and Rocco Micciolo³

¹ Center for Mind/Brain Sciences (CIMEC), University of Trento, Trento, Italy, ² Department of Humanities, University of Trento, Trento, Italy, ³ Department of Psychology and Cognitive Sciences, University of Trento, Trento, Italy

OPEN ACCESS

Edited by:

Aleksandra Mroczko-Wasowicz,
National Yang Ming University, Taiwan

Reviewed by:

David R. Simmons,
University of Glasgow, UK
Mats B. Küssner,
Royal College of Music, UK

*Correspondence:

Liliana Albertazzi,
Center for Mind/Brain Sciences
(CIMEC), University of Trento, Corso
Bettini, 31, 38068 Rovereto, Italy
liliana.albertazzi@unitn.it

Specialty section:

This article was submitted to
Consciousness Research,
a section of the journal
Frontiers in Psychology

Received: 29 September 2014

Accepted: 25 March 2015

Published: 21 April 2015

Citation:

Albertazzi L, Canal L and Micciolo R
(2015) Cross-modal associations
between materic painting and classical
Spanish music. *Front. Psychol.* 6:424.
doi: 10.3389/fpsyg.2015.00424

The study analyses the existence of cross-modal associations in the general population between a series of paintings and a series of clips of classical (guitar) music. Because of the complexity of the stimuli, the study differs from previous analyses conducted on the association between visual and auditory stimuli, which predominantly analyzed single tones and colors by means of psychophysical methods and forced choice responses. More recently, the relation between music and shape has been analyzed in terms of music visualization, or relatively to the role played by emotion in the association, and free response paradigms have also been accepted. In our study, in order to investigate what attributes may be responsible for the phenomenon of the association between visual and auditory stimuli, the clip/painting association was tested in two experiments: the first used the semantic differential on a unidimensional rating scale of adjectives; the second employed a specific methodology based on subjective perceptual judgments in first person account. Because of the complexity of the stimuli, it was decided to have the maximum possible uniformity of style, composition and musical color. The results show that multisensory features expressed by adjectives such as “quick,” “agitated,” and “strong,” and their antonyms “slow,” “calm,” and “weak” characterized both the visual and auditory stimuli, and that they may have had a role in the associations. The results also suggest that the main perceptual features responsible for the clip/painting associations were hue, lightness, timbre, and musical tempo. Contrary to what was expected, the musical mode usually related to feelings of happiness (major mode), or to feelings of sadness (minor mode), and spatial orientation (vertical and horizontal) did not play a significant role in the association. The consistency of the associations was shown when evaluated on the whole sample, and after considering the different backgrounds and expertise of the subjects. No substantial difference was found between expert and non-expert subjects. The methods used in the experiment (semantic differential and subjective judgements in first person account) corroborated the interpretation of the results as associations due to patterns of qualitative similarity present in stimuli of different sensory modalities and experienced as such by the subjects. The main result of the study consists in showing the existence of cross-modal associations between highly complex stimuli; furthermore, the second experiment employed a specific methodology based on subjective perceptual judgments.

Keywords: connotative dimensions, cross-modal associations, music, painting, subjective judgments

Introduction

In recent years the field of perception studies has seen an increasing amount of research showing the tendency for a sensory feature, or attribute, in one modality to be matched with a sensory feature in another modality (Simner et al., 2005, 2011; Sagiv and Ward, 2006; Ward et al., 2006a,b; Cohen Kadosh et al., 2009. For a review see Spence, 2011). The phenomenon had already been pointed out by Köhler, who showed the tendency of the general population systematically to associate visual and auditory attributes (the so-called “takete-maluma” phenomenon) (Köhler, 1929; Gallace et al., 2011; Nielsen and Rendall, 2011). Initially prompted by interest in the field of synesthesia (Wicker, 1968; Melara and O’Brien, 1987; Cytowic, 1995; Baron-Cohen and Harrison, 1997; Ward and Simner, 2003; Simner et al., 2006), studies then considered similar phenomena occurring in the general population (Martino and Marks, 2001; Maurer and Mondloch, 2005; Sagiv and Ward, 2006; Spector and Maurer, 2008, 2011; Parise and Spence, 2009; Deroy et al., 2013; Deroy and Spence, 2013). In regard to the nature of such associations, Spence has distinguished among structural correspondences (due to neural correlates, hence potentially universal), statistical correspondences (due to learning, hence potentially influenced by different environments), and semantic correspondences (due to language influence, hence potentially different among cultures) (Spence, 2011). Recently, a growing number of researchers have sought to explain synesthetic and cross-modal associations in semantic more than sensory terms by re-evaluating the possible role of cognitive factors in the associations. In particular, the question has been raised in regard to inducers that take the form of concepts—such as the days of the week or the months—usually associated with colors. In short, since inducers have a conceptual nature, it has been asked whether a full account of synesthesia should not go beyond the standard sensory-sensory approach (Dixon et al., 2006; Simner and Ward, 2006; Ward et al., 2007; Eagleman, 2012; Jürgens and Nikolić, 2012, 2014; Mroczko-Wąsowicz and Werning, 2012; Simner, 2012; Ward, 2013; Mroczko-Wąsowicz and Nikolić, 2014).

Bottom up and Top Down Explanations

The opposition between the sensory interpretation (bottom-up, i.e., sense-driven) and the conceptual interpretation (top-down, i.e., concept driven) of synesthetic associations has arisen within the classical framework of cognitive science, which counterposes the two levels of information processing. Consequently, the former interpretation has sought to explain the associations in terms of direct synaptic connections between neurons (representing the inducer and the concurrent); the latter is based on high-level processes due to language, culture, abstract symbolization, learning, etc. (Ward et al., 2007). The second interpretation, however, would seem better suited to explaining cases of color sequence synesthesia (Simner et al., 2006; Tomson et al., 2013) and spatial sequence synesthesia (Sagiv et al., 2006; Eagleman, 2010), where the names of time units and ordinal categories are involved. This second interpretation, which considers cases of synesthesia occurring independently of external inducers, has taken the name of ideasthesia (Meier, 2013; Jürgens and Nikolić, 2014 for

a thorough discussion of the topic see Mroczko-Wąsowicz and Nikolić, 2014).

There is also another interpretation. It rests on a not necessarily linguistic or symbolic conception of semantics. This third approach, of Gestalt derivation (Albertazzi, 2013), explains associations in terms of *patterns of qualitative similarity* present in different sensory modalities and perceived as such: for example, hot and cold, sad and happy, and pleasant and unpleasant, are connotative properties of both sounds and colors. This therefore concerns, not semantic information projected top-down into other domains, but qualities intrinsic to perceived phenomena. This position obviously does not preclude investigation of *correlations* at neuronal level or of the presence of cognitive dimensions due to learning, language, symbolization, etc. This interpretation has been adopted in studies on the associations between color and shape in the general population (Dadam et al., 2012; Albertazzi et al., 2013, 2014, 2015).

Whatever viewpoint is adopted in interpreting the phenomenon, there is growing interest in cross-modal associations both within synesthesia (Simner, 2012; Ward, 2013) and in the general population (Deroy and Spence, 2013).

In the field of cross-modal associations occurring in the general population, a perceptual attribute which proves to play an important role is color. In fact, it has been shown that color is associated with olfaction (Gilbert et al., 1996; Kemp and Gilbert, 1997; Demattè et al., 2006; Hanson-Vaux et al., 2013; Levitan et al., 2014), touch (Ludwig and Simner, 2013), and acoustics (Ward et al., 2006a; Moos et al., 2013, 2014).

Cross-Modal Associations between Visual and Auditory Stimuli

As regards studies on associations between visual and auditory stimuli, initially considered were predominantly single tones by means of psychophysical methods and forced choice responses (Walker, 1987). More recently, the relation between visual and more complex auditory stimuli (i.e., music clips) has also been tested (Tan and Kelly, 2004; Küssner, 2013b), and free response paradigms have also been accepted (Reybrouck et al., 2009). In particular, the task in Tan and Kelly (2004) was to create marks or drawings that visually represented five short orchestral compositions, and to write essays explaining their graphic representations; while in Küssner (2013b) the task was to visualize sound and music by creating representations with an electronic graphics tablet (in two different experimental conditions, i.e., drawing during and after the sounds). In their experiment, Tan and Kelly tested musically trained and untrained subjects, and found a difference between them: i.e., the musically trained participants provided abstract representations, such as lines of symbols, while the untrained participants produced pictorial representations, such as images or pictures telling a story. A second difference consisted in the fact that trained participants focused more on musical characteristics (such as theme, mode, changes in pitch, etc.), while untrained participants focused more on emotions.

It has also been found that music in the major mode more closely matches lighter colors than does music in the minor mode (Bresin, 2005), while faster music in the major mode

more closely matches more saturated and lighter colors than does slower music in the minor mode (Palmer et al., 2013). In particular, Bresin's study explicitly addressed the role of expressivity in music, which was verified by testing the association made by expert subjects between colors and performances of classical music. The results of this study showed that participants used different color profiles to classify the same piece of music, but these differences depended mainly on the performance and on the instrument. The study by Palmer et al. (2013) instead tested participants of different cultures (United States and Mexico), and found that in both cultures, faster music in the major mode produced color choices that were more saturated, lighter, and yellower, whereas slower, minor music produced the opposite pattern. Similarly, other studies have been conducted in order to explain the association between visual and auditory stimuli on the basis of their shared emotional content, such as the association between happy music and happy colors (Whiteford et al., 2013; Langlois et al., 2014).

As is well known, cross-modal and cross-dimensional associations have always played a role in aesthetics. Recently, the association between color and shape has been experimentally tested by studies relating to Kandinsky's hypothesis of a systematic association between geometrical shapes and colors (Droste, 1990; Lupton and Miller, 1991; Jacobsen, 2002; Kharkhurin, 2012; Albertazzi et al., 2013, 2015; Makin and Wuerger, 2013; Chen et al., 2015). Besides the above-mentioned shape/color association in Kandinsky, the analogy between scales of colors and scales of notes is a major component of the harmony theories developed by Klee (1956), Kandinsky (1926/1994), and Itten (Itten et al., 1970; Gage, 1999) with particular regard to Schönberg (Schönberg and Kandinsky, 1980; Bidaine, 2004). In our study, and following previous studies of ours dealing with aesthetics (Albertazzi et al., 2013, 2015), we tested the association in the general population between some artistic works in painting and some pieces of classical music, in order to evaluate whether systematic cross-modal associations occur among stimuli of high complexity. Specifically, we tested whether images with varying perceptual characteristics and contents led to consistent associations with the music clips, and what attributes might be responsible for the phenomenon. The choice of the images and of the clips was based on the hypothesis that their artistic modes of expressions (the coloratura of the flamenco and the materic style of the paintings), and a series of connotative properties holding for colors and tones (like weak and strong, calm and agitated) play a role in the associations. Precisely, the specific coloratura (Tonkolorit) of flamenco music is characterized by very brief sharp notes and a minor scale. As to materic (or material) painting, this is painting realized with a great quantity of pictorial material, and characterized by a thick and tendentially 3D pictorial surface. The study that may show an affinity with ours was conducted in the 1930s by Cowles (1935), who also made use of complex stimuli (8 pieces of classical music, although composed by different musicians, and 8 paintings by various well-known artists), and with expert and non-expert participants. There are differences with our study, however, both in the number and in the kind of stimuli: in Cowles (1935) the pictures were mainly

landscapes or scenes with simple content, without uniformity of style; the auditory stimuli were taken from works by different composers, differing in character, although there were no more than slight variations in volume, tempo, or tone quality. Finally, the aim of one of the two experiments conducted in Cowles' study, differently from ours, was to verify whether similar affective moods were found between the musical selections and the pictures.

The Study

The purpose of our research was to test whether the general population exhibits cross-modal associations between complex stimuli of two different modalities (vision and sound). Specifically, the aim of our research was to test whether significant associations existed between a series of paintings and a series of clips of classical (guitar) music, and whether these associations were consistent when evaluated on the same subjects. The research also sought to evaluate whether the findings were confirmed on different subjects with different backgrounds and expertise. Our expectation was that, if found, these associations would be consistent from one subject to another, suggesting a predisposition to perceive specific cross-modal natural associations between complex visual and auditory stimuli.

The selection of the paintings and the music clips was discussed with the painter (Matteo Boato: <http://www.matteoboato.net/>), who is also a musician and who provided a description of the individual art works and the characteristics of the music clips selected. The choice of Boato's works was made (apart from personal preference) on the basis of their characteristics of high chromaticity and saturation. Our hypothesis was that corresponding to these visual characteristics are similar patterns in the acoustic modality as to vibrato, coloratura, and quick tempo: for example, we expected that a quick tempo would correspond to a very chromatic and saturated red or yellow. Specifically, the hypothesis was that the association, if found, would be due to multisensorial and connotative features present in both the visual stimuli and the auditory stimuli, such as warmth/coldness, brightness/darkness, sadness/happiness, softness/hardness, etc. The prediction was therefore that the subjects would make systematic associations between the paintings and the music, and that the associations would be due to the presence of similar features in the paintings and the music, as also evidenced by the semantic differential. Because of the complexity of the stimuli, we tried to keep the maximum amount of uniformity possible. The purpose of using works by the same painter as stimuli was to maintain the same style (materic painting) and composition (expressionist) notwithstanding the diversity of content and colors (achromatic and mainly chromatic paintings depicting landscapes and figures were tested). The purpose of using clips from the same repertoire was to maintain the coloratura of classical Spanish flamenco music. The clips were instead chosen for their specific musical features, such as having a strong, hard, agitated sound and a quick tempo (*presto*) (for example, *Asturias* by Albéniz). The recorded music clips were performed by Boato himself. Finally, we did not test individual preferences because it was not an objective of our experiment.

Methods

Participants

Sixty-three participants volunteered for two experiments: 38 women and 25 men (mean age: 22.6 years; standard deviation: 3.5; median: 22 years). All participants were recruited by e-mail from students in the Department of Cognition and Education Sciences, University of Trento, Italy. The address list of the students was provided by the student office. We firstly sent a mail asking the students to adhere to the experiment, mentioning that we were looking for people with a background in music, people with a background in art, and non-expert people. We didn't ask for professional people, however. When we contacted the students who adhered to the experiment, we decided to accept people who had a public or private artistic education in music for at least 4 years, people who had a private or public education in art, and non-expert people. The questionnaire reported this information. The subjects were also asked about a possible conscious synesthesia (Palmer and Schloss, 2010; Albertazzi et al., 2013, 2015; Palmer et al., 2013). The only exclusion criterion was self-reported defective color vision or acoustic impairment. After the experiment, the subjects were asked whether they had previously known the paintings and the pieces of music that they evaluated. For all the subjects the stimuli were totally new.

The first experiment was performed using the semantic differential on a unipolar rating scale of adjectives. It was decided to use a unipolar scale instead of the classic bipolar one of the Osgood semantic differential (Osgood, 1956) because the bipolar scale is not always one-dimensional. Sixty-one subjects participated in the experiment (two subjects who did not complete the experiment were excluded from the analysis). The second experiment evaluated the association between visual and auditory stimuli and was completed by all the 63 subjects. We tested non-experts (31), music experts (20), and art experts (12), the purpose being to investigate a possible influence of expertise on the associations. Individuals with training in private or public schools were considered expert participants in the present study. All the subjects signed an informed consent form. The experiments reported here complied with the ethical guidelines of the University of Trento.

Procedure

The experiment was performed in a laboratory with constant and controlled lighting conditions (230–250 lux) in the room, correlated color temperature 3400K, halogen lamp). The visual stimuli appeared on a Quato Display 242ex (Intelli Prof 242 excellence) 24" screen (51.8 × 32.4 cm visible area); the auditory stimuli were administered through Sennheiser HD580 Precision headphones. Automatic 48 bit USB-hardware calibration with 3 × 16 bit 3D Look-Up Table and luminance inside the monitor, dedicated luminance stability circuit, UDACT display analysis built-in; the measurement device was a 4-channel Silver Haze Pro colorimeter. The resolution used was 1920 × 1200 pixels (the native and the maximum possible for the monitor Display Quato 242).

Participants were seated at a desk. The distance from the center of the screen to the eye was about 65 cm. Chin supports were not used, but during each session the postures of the participants

were checked and corrected if their chests approached the screen or their backs were hunched.

General Materials

The materials consisted of a series of 15 paintings (by the same painter), a series of 15 music clips and a list of 22 adjectives. The titles of the paintings were: (1) "Padova, 2007," (2) "Verona, 2009," (3) "Mantova, 2009," (4) "Trento, 2008," (5) "Full Moon," (6) "The Circle," (7) "Trento, 2006," (8) "Burano, 2009," (9) "Sky of Fields," (10) "Sea II," (11) "Land—Hora et Labora," (12) "In Dream II," (13) "In Dream 2006," (14) "Leopard," (15) "Matilada and Beatrice" (see Supplementary Material, reproduction permitted by Boato). For presentation of the digital images a high-resolution digital transcription was performed by an expert in the graphic reproduction of works of art.

The clips (performed by the same player) were taken from the following musical works: (1) Heitor Villa Lobos, *Prelude n. 4*; (2) Heitor Villa Lobos, *Mazurka, Suite populaire brésilienne*; (3) Francisco Tárrega, *Recuerdos de la Alhambra*; (4) Isaac Albéniz, *Asturias—Part I*; (5) Fernando Sor, *Variations on a theme by Mozart—II var*; (6) Gaspar Sanz, *Canarios, Suite Española*; (7) Fernando Sor, *Variations on a theme by Mozart—Theme*; (8) Fernando Sor, *Variations on a theme by Mozart—I var*; (9) Manuel Ponce, *Giga*; (10) Gaspar Sanz, *Espanoletas, Suite Española*; (11) Heitor Villa Lobos, *Prelude n. 5—Part I*; (12) Heitor Villa Lobos, *Prelude n. 5—Part II*; (13) Isaac Albéniz, *Leyenda, Asturias*; (14) Heitor Villa Lobos, *Study n. 6*; (15) Francisco Tárrega, *Arabian caprice*.

The assessments of the adjectives were arranged on a continuous scale between 0 and 1024. We selected for the experiment mainly adjectives that could be applied to both music and paintings. The experiment was preceded by a pilot test with the same characteristics as the experiment itself but a much longer list of adjectives. The original list of adjectives included 49 items evaluated by 35 subjects. After a correlational study, the list of adjectives was shortened to include 22 items. The final list of adjectives (presented in Italian) was the following: slow, quick, agitated, calm, happy, sad, warm, cold, heavy, light, continuous, rhythmic, strong, weak, dark, bright, hard, soft, impression of horizontality, impression of verticality, adagio, presto (the two last items were left in the adverbial form as they are in Italian). As to the chromatic dimensions, neither hue nor saturation were considered (all the paintings were uniformly drawn with very saturated hues), but rather the dimensions of warmth (warm/cold) and brightness (light/dark) (relying on the contrast between the fragments of colors and the painted background used by the painter). The choice of dimensions was due to their perceptual salience and to the fact that they are the most meaningful dimensions in cross-modal associations where color is involved. The asymmetric choice of having the subjects listen to a music clip and asking them to associate three paintings with them, and not vice versa, was dictated by the complexity of the task, which was of considerable duration (about an hour and a half, with a pause). We also hypothesized that asking the subjects to look at the paintings and associate three music clips from the classical guitar repertoire with them would have been an excessively burdensome task. In fact, it would have required listening to 15

clips sequentially (although in random order) for each painting. Instead, as shown in **Figure 1**, the 15 paintings were seen all together.

Experiment 1

The experiment was performed using the semantic differential on a unidimensional rating scale of adjectives. First the individual images (in random order) were presented on the screen and then each music clip was executed (also in this case, the order of presentation was randomized). Participants were told that they would first see a set of images (each was displayed on the screen for 10 s) and then hear a series of music clips (each lasting 60 s). For each stimulus the subject had to evaluate, on a continuous scale, his/her degree of agreement with a series of adjectives. Participants were given the following written instructions for the task:

You will be presented with images on the screen or music clips through your headphones accompanied by a series of adjectives in succession. You should evaluate these adjectives with reference to the image or music presented. Evaluation of the adjective will be made on a continuous scale. You should prefer accuracy to promptness of response.

The purpose of the experiment was to check whether complex images and music clips with varying perceptual characteristics led to consistent choices of adjectives. Images were shown one by one (in random order) on the left half of the screen, while on the right half of the screen participants saw one after the other the adjectives presented randomly (**Figure 2**). The same occurred with the music clips, which could be heard by clicking on a button positioned on the left side of the screen.

Experiment 2

The purpose of the second experiment was to check whether images with varying perceptual characteristics and contents led to consistent associations with music clips taken from the repertoire of classical (guitar) music. Each subject saw a series of images of paintings in preview on the screen. The subject clicked on a specific image, which thus appeared in full screen mode, and likewise with the other images, in no particular order. The subject viewed the images while simultaneously listening to a music clip. The subject had to choose the image(s) that s/he most naturally associated with that music. S/he could list up to three images associated with the clip, arranging them in order of appropriateness from 1 to 3 in three different boxes at the bottom on the screen (**Figure 1**). The subject could go back to re-view images already seen, and s/he could also listen repeatedly to the music clip. Once the association had been decided, the images selected were transported down into one of three boxes, depending on the degree of association, in order from 1 to 3. Once the choice had been confirmed, it could not be changed, and the task continued with re-presentations of all the images and further music clips until the latter were exhausted.

Participants were given the following written instructions for the task:

You will see a series of images of paintings in preview on the screen. Click on one of them, which will appear in full screen mode, and then do likewise with the other images. At the same time, you will hear a music clip. Select which image(s) you most naturally associate with the music. You can go back to re-view images already seen, and also to hear the music clip again. You can list up to three images associated with the music, placing them in order

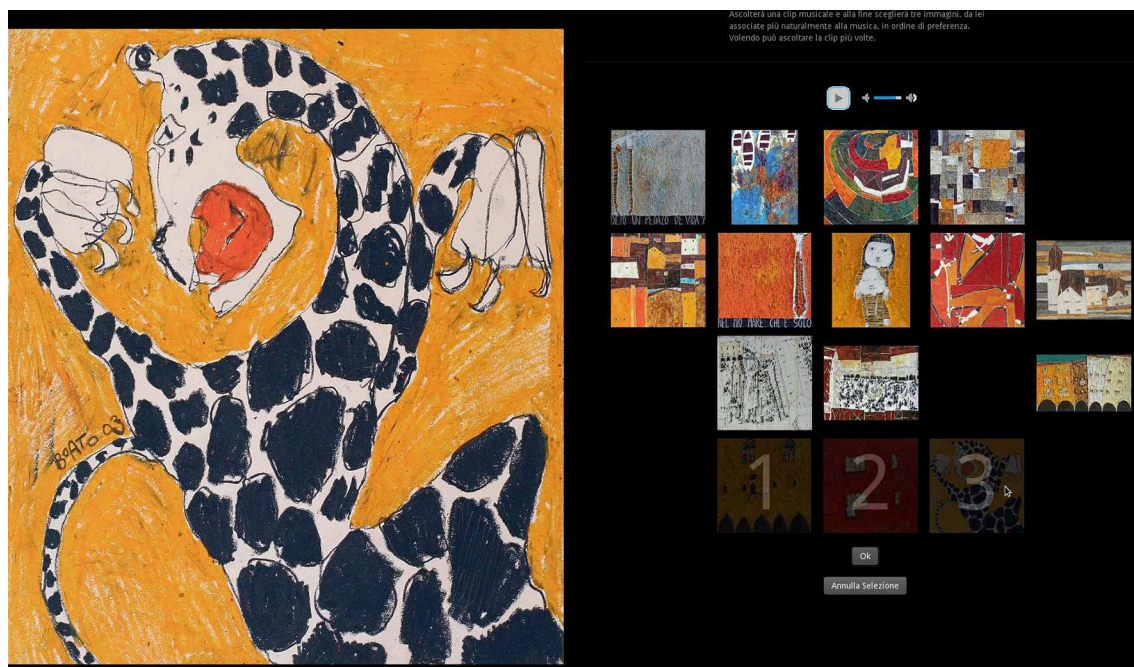


FIGURE 1 | Example of a painting selected in association with a given music clip (The arrow points where to click to hear the music clip again).

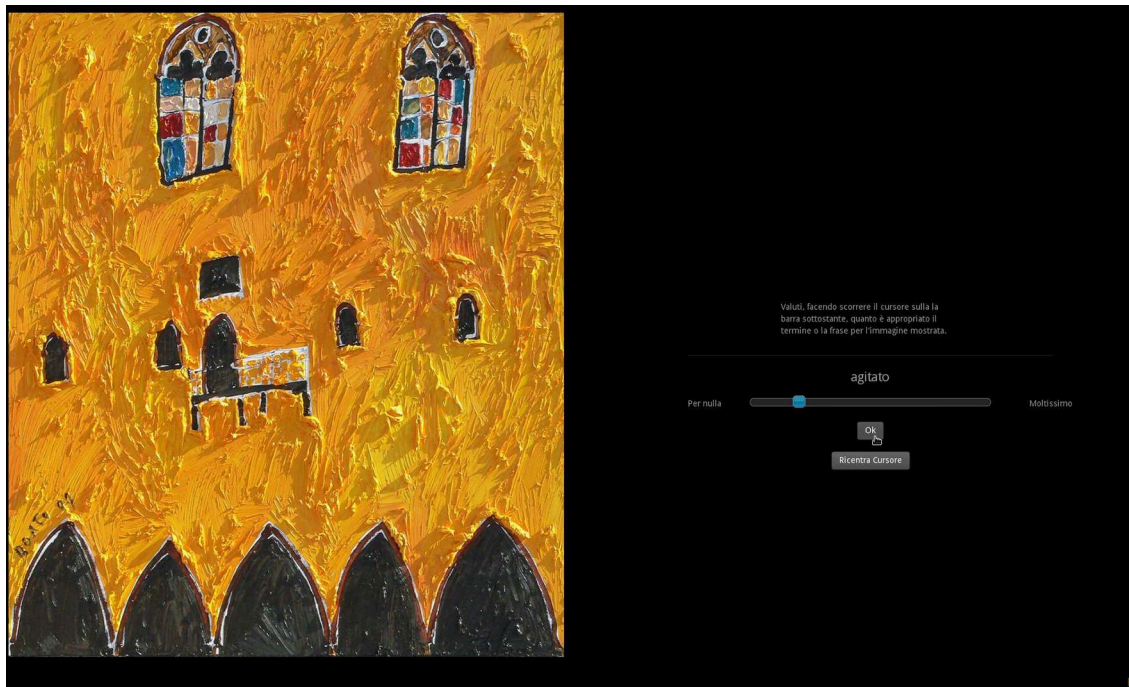


FIGURE 2 | Example of a painting to be evaluated according to a series of adjectives.

of appropriateness from 1 to 3. Once you have confirmed your choice, it cannot be changed, and the task will continue with further music clips until there are none left. You should prefer accuracy to promptness of response.

Statistical Methods

Associations between quantitative variables were evaluated by means of the non-parametric “rho” correlation coefficient. The chi-square test for a contingency table was employed to evaluate the associations between the paintings and the music clips. A residual analysis was performed to identify which painting/clip combinations were significant (Canal and Micciolo, 2013). Analyses were performed with R 3.0.0 software (R Core Team, 2013).

Results

Experiment 1

Table 1 reports the mean rating values for each word-painting pair given by the 61 participants. Means range between 186 and 842. This latter value was obtained when considering painting number 3 (“Mantova, 2009”) and the adjective “bright”; therefore this painting was considered the most luminous. The minimum value was obtained when considering painting number 14 (“Leopard”) and the adjective “weak”; therefore this painting was considered the least weak of the 15 paintings.

Table 2 reports the mean rating values for each word-clip pair given by the 61 participants. Means range between 133 and 864. This latter value was obtained when considering clip no. 14 (Villa Lobos, *Study n. 6*) and the adjective “agitated”; therefore this clip was considered the most agitated. The minimum value was obtained when considering clip no. 8 (Fernando Sor, *Variations*

on a theme by Mozart—I var) and the adjective “dark”; therefore this clip was considered the least dark of the 15 clips.

To evaluate the degree of association between the semantic rating (i.e., considering the mean ratings of the 22 words) of one selected painting and one selected clip, non-parametric rho correlation coefficients were calculated. The results are shown in **Table 3** (the rows contain the 15 music clips, the columns the 15 paintings).

These correlations ranged between -0.69 and 0.90. This latter value was obtained when considering the mean ratings of the 22 words given to painting no. 11 (“Land—Hora et Labora”) (see **Table 1**) and to clip no. 6 (Gaspar Sanz, *Canarios*) (see **Table 2**). The highest negative correlation was found between painting no. 14 (“Leopard”) and clip no. 3 (Francisco Tárrega, *Recuerdos de la Alhambra*).

Experiment 2

Table 4 shows the results of Experiment 2.

For each clip listened to (shown in the rows of the table), the percentage of painting choices is reported (considering only the first choice of a painting). It seems evident from visual inspection of the table that some paintings were more frequently associated with a given clip: for example, 30.2% of participants associated painting no. 5 (“Full Moon”) with clip no. 3 (Francisco Tárrega, *Recuerdos de la Alhambra*). On the other hand, some paintings were less frequently associated with a given clip; for example, none of the participants associated painting no. 14 (“Leopard”) with clip no. 1 (Villa Lobos, *Prelude n. 4*). The chi-square test revealed that the association between the variables “painting”

TABLE 1 | Mean ratings for each word and for each painting given by the 61 participants.

Word	Painting°														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Adagio	376	427	511	457	557	377	502	401	469	430	393	493	427	238	520
Presto	582	548	393	435	270	476	370	445	415	492	509	436	401	585	325
Agitated	619	617	362	533	236	613	347	487	457	628	608	460	447	777	325
Calm	362	426	599	433	701	401	629	471	552	469	319	500	512	188	609
Bright	527	696	842	767	507	691	720	649	677	704	738	582	732	724	748
Dark	444	348	254	284	506	268	275	368	272	254	271	273	341	332	373
Cold	628	393	258	239	590	196	358	339	463	439	244	593	235	291	291
Warm	254	509	762	709	373	794	609	575	494	547	724	373	749	624	653
Continuous	519	583	546	470	578	458	595	643	530	479	710	530	577	453	534
Rhythmic	563	591	401	447	449	584	555	621	686	522	721	448	435	527	395
Happy	363	499	570	477	252	535	529	413	447	553	664	424	446	395	502
Sad	556	361	320	324	652	282	391	469	376	339	268	419	449	295	434
Hard	480	424	445	565	480	494	477	550	502	384	427	407	444	590	367
Soft	303	401	441	374	330	373	387	311	327	429	423	466	409	289	487
Heavy	397	425	448	576	549	437	482	547	417	342	470	342	517	467	482
Light	447	411	439	350	407	425	462	309	480	532	356	529	411	330	437
Horizontality	329	617	455	401	696	467	514	423	497	361	429	248	270	272	310
Verticality	718	549	638	571	587	679	703	761	661	694	525	800	810	652	718
Quick	577	542	335	440	243	553	367	427	396	544	630	438	405	689	266
Slow	350	398	511	412	674	370	481	465	505	465	341	496	499	218	576
Strong	482	507	566	686	369	688	506	556	519	563	629	444	631	733	469
Weak	323	316	307	251	494	258	363	321	377	290	268	374	266	186	391

°See text for the correspondence between the ID number and the title of the painting.

and “clip” cannot be considered random but instead systematic (chi-square = 517; d.f. = 196; $p < 0.001$). Given that the lowest expected frequency was less than 5, a Monte Carlo simulation was performed which confirmed the significance of the association ($p < 0.001$).

Since the test did not indicate which clip was associated (positively or negatively) with which painting, a residual analysis was performed. A standardized form of the residual was employed. This behaves like a normal deviate to determine whether the residual is large enough to indicate a departure from a random choice. In this case, there is only about a 5% chance that any particular standardized residual exceeds 1.96 in absolute value. When we inspected 225 cells, about 11 residuals (i.e., 5% of 225) could have been so large solely because of random variation. On the other hand, as can be seen in **Table 5**, there were 40 residuals greater than 1.96 in absolute value.

Overall, there were 22 residuals greater than 1.96, and 18 residuals lower than -1.96. A positive residual means that the selected clip “attracted” the corresponding painting; a negative residual means that the selected clip “repelled” the corresponding painting.

There were five clips which showed a very strong attraction (a residual greater than 4). Clip no. 3 (Tárrega, *Recuerdos de la Alhambra*) was strongly associated with image no. 5 (“Full Moon”); clip no. 4 (Albéniz, *Asturias*) was strongly associated with image no. 14 (“Leopard”); clip no. 14 (Villa Lobos, *Study n. 6*) was strongly associated with image no. 14 (“Leopard”); clip

no. 1 (Villa Lobos, *Prelude n. 4*) was strongly associated with image no. 10 (“Sea II”); clip no. 7 (Sor, *Variations on a theme by Mozart*) was strongly associated with image no. 15 (“Matilada and Beatrice”).

On the other hand, the negative associations were weaker; the lowest residual was -2.60. Clip no. 12 (Villa Lobos, *Prelude n. 5—Part II*) was negatively associated with image no. 2 (“Verona, 2009”); clip no. 14 (Villa Lobos, *Study n. 6*) was negatively associated with image no. 15 (“Matilada and Beatrice”); clips no. 1 (Villa Lobos, *Prelude n. 4*), no. 3 (Tárrega, *Recuerdos de la Alhambra*), and no. 12 (Villa Lobos, *Prelude n. 5—Part II*) were all negatively associated with image no. 14 (“Leopard”).

A Comparison between the Results of Experiment 1 and Experiment 2

To evaluate if and to what extent the “direct” associations found in Experiment 2 were in agreement with the correlations in terms of semantic differential (Experiment 1), we counted how many times the sign of the “significant” residuals (i.e., residuals greater than 1.96 in absolute value) shown in **Table 5** for the 40 clip/painting combinations was the same as the corresponding correlation shown in **Table 3**. For 21 combinations, both the residuals and the correlations were positive, showing that, when a particular painting was attracted by a given clip, the 22 words had similar ratings. On the other hand, for 12 combinations both the residuals and the correlations were negative, showing that, when a particular painting was repelled by a given clip, the 22 words

TABLE 2 | Mean ratings for each word and for each clip given by the 61 participants.

Word	Music Clip ^o														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Adagio	444	312	595	182	333	206	350	221	269	603	409	597	274	163	392
Presto	489	620	392	655	580	682	583	696	664	410	559	349	592	696	549
Agitated	598	661	346	805	645	685	500	634	725	370	500	397	646	864	578
Calm	453	366	600	137	327	209	415	238	254	597	423	570	308	162	406
Bright	423	646	376	376	631	818	790	793	592	505	659	358	545	459	499
Dark	538	336	512	535	272	137	156	133	366	392	225	533	357	506	438
Cold	443	383	468	480	335	215	226	279	361	344	332	412	385	456	419
Warm	492	558	458	460	582	713	661	657	531	546	579	529	516	467	488
Continuous	448	588	736	623	624	633	691	667	621	614	558	588	489	607	521
Rhythmic	543	651	491	778	669	790	646	785	742	518	634	508	638	767	581
Happy	320	570	269	411	613	780	781	820	553	476	657	306	516	456	429
Sad	636	408	682	485	410	147	167	188	397	532	261	678	437	430	551
Hard	414	382	355	617	366	322	213	273	459	296	302	398	475	632	410
Soft	390	459	526	233	501	443	584	477	393	595	505	541	343	248	431
Heavy	446	333	352	474	280	250	183	180	359	313	243	441	336	522	317
Light	416	525	512	354	543	585	742	641	496	617	563	497	444	312	504
Horizontality	494	535	669	430	524	460	655	483	584	574	574	636	461	385	518
Verticality	585	587	382	651	485	609	414	578	479	434	492	380	536	649	562
Quick	604	745	377	817	714	806	687	819	758	401	656	363	706	830	636
Slow	462	263	598	157	241	166	254	159	214	556	295	647	286	137	348
Strong	537	478	374	727	485	624	406	472	582	361	418	422	554	730	481
Weak	305	336	472	160	303	195	342	236	262	497	334	447	274	217	366

^oSee text for the correspondence between the ID number and the title of the clip.

had opposite ratings. In the remaining seven combinations, the sign of the residual and the sign of the correlation disagreed. If the clip/painting associations shown in **Table 5** randomly agreed with the correlations shown in **Table 3**, a total of 20 combinations would have the same sign and 20 combinations would have different signs. An exact binomial test yielded a significant result ($p < 0.001$), in contrast with the hypothesis that the associations found were essentially random.

Furthermore, the correlation between the values reported in **Table 3** and all the standardized residuals shown in **Table 5** was significantly different from zero ($\rho = 0.338$; $p < 0.001$). Therefore, at least in part, the painting/clip association could be explained by similar perceptual characteristics.

Quite similar results were found when the analyses described above were performed considering all the three paintings selected for a given clip. The final correlation coefficient was 0.365 (0.338 was found when only the first painting chosen was considered).

The subjects who participated in the experiment were classified into three groups: music experts, painting experts, and non-experts. When all the analyses were repeated selecting only the subjects of the same group, similar results were found. When music experts were selected, the correlations were 0.293 (considering only the first painting chosen) and 0.341 (all the three paintings chosen). When painting experts were selected, the correlations were, respectively, 0.210 and 0.319. Painting/clip association in terms of similar perceptual characteristics was confirmed also within the three groups.

Discussion

The study tested whether the general population exhibits cross-modal associations between complex stimuli of two different modalities, and specifically between a series of paintings and a series of clips of classical (guitar) music. The test was conducted with subjects who were both expert and non-expert in visual and musical arts.

The study tested the association in two experiments. One was conducted using the semantic differential on a unidimensional rating scale of adjectives; the other was based on subjective judgments on the association between visual and auditory stimuli. The hypothesis was that the association, if found, would be linked to the presence of characteristics of the paintings and the music clips, perceived as such by the subjects, and evidenced also when evaluated by means of the semantic differential. Due to the experimental nature of the study, the link between the two experiments cannot be consistently found for each clip/image couple; in some cases such a link may not be consistent. Overall, the results show the existence of an association between paintings and music clips among experts in music, experts in painting, and subjects with no artistic training, within each group and overall.

These results were consistent when considering both the first painting chosen and all the three paintings selected for a given clip. Specifically, there were five clip/image couples for which a very strong attraction was found: specifically, clip no. 3

TABLE 3 | Correlation coefficients (Spearman's rho) for each combination clip/painting evaluated employing the semantic ratings reported in Tables 1, 2.

		Painting°														
Music Clip°		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	1	0.47	0.33	−0.15	0.09	0.03	0.29	−0.04	0.39	0.06	0.18	0.18	−0.11	0.20	0.41	−0.20
	2	0.53	0.78	0.10	0.35	−0.52	0.70	0.18	0.39	0.23	0.71	0.74	0.19	0.15	0.70	−0.14
	3	−0.29	−0.24	−0.09	−0.58	0.63	−0.55	0.07	−0.23	0.04	−0.47	−0.42	0.13	−0.18	−0.69	0.12
	4	0.74	0.45	−0.27	0.29	−0.45	0.49	−0.18	0.43	0.02	0.33	0.45	−0.09	0.06	0.77	−0.46
	5	0.36	0.75	0.14	0.37	−0.58	0.69	0.22	0.31	0.22	0.67	0.78	0.15	0.13	0.62	−0.10
	6	0.32	0.81	0.36	0.60	−0.59	0.84	0.38	0.45	0.36	0.80	0.90	0.17	0.29	0.72	0.06
	7	−0.04	0.65	0.45	0.31	−0.37	0.56	0.45	0.13	0.33	0.66	0.69	0.27	0.16	0.28	0.23
	8	0.30	0.71	0.27	0.38	−0.55	0.68	0.33	0.26	0.28	0.73	0.77	0.23	0.11	0.54	0.00
	9	0.49	0.78	0.03	0.38	−0.56	0.70	0.11	0.37	0.19	0.58	0.76	−0.01	0.07	0.73	−0.28
	10	−0.44	0.05	0.31	−0.23	0.31	−0.17	0.38	−0.16	0.21	−0.02	−0.03	0.29	0.07	−0.52	0.45
	11	0.05	0.71	0.38	0.33	−0.43	0.64	0.43	0.17	0.32	0.67	0.73	0.18	0.09	0.36	0.11
	12	−0.49	−0.33	−0.04	−0.41	0.61	−0.51	0.05	−0.16	0.00	−0.57	−0.42	−0.11	−0.03	−0.65	0.17
	13	0.60	0.73	0.09	0.49	−0.58	0.80	0.13	0.50	0.25	0.72	0.75	0.06	0.22	0.86	−0.21
	14	0.68	0.54	−0.13	0.47	−0.54	0.63	−0.08	0.52	0.10	0.44	0.59	−0.08	0.17	0.88	−0.35
	15	0.61	0.60	−0.13	0.11	−0.33	0.50	0.03	0.32	0.12	0.47	0.48	0.12	0.06	0.57	−0.25

°See text for the correspondence between the ID number and the title of the clip/painting.

TABLE 4 | For each clip listened to, the number of times (in percentage) each painting was chosen is reported (Each row of the table sums up to 100).

		Painting°														
Music Clip°		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	1	4.8	6.3	3.2	6.3	12.7	1.6	0.0	4.8	4.8	19.0	6.3	12.7	9.5	0.0	7.9
	2	11.1	14.3	6.3	4.8	1.6	11.1	12.7	6.3	4.8	3.2	9.5	3.2	1.6	3.2	6.3
	3	7.9	3.2	1.6	3.2	30.2	1.6	7.9	7.9	6.3	3.2	3.2	11.1	6.3	0.0	6.3
	4	11.1	4.8	1.6	3.2	1.6	11.1	0.0	0.0	6.3	4.8	15.9	4.8	1.6	30.2	3.2
	5	9.5	4.8	1.6	4.8	1.6	7.9	4.8	11.1	12.7	3.2	7.9	6.3	7.9	6.3	9.5
	6	4.8	17.5	7.9	7.9	0.0	6.3	1.6	4.8	7.9	7.9	15.9	1.6	4.8	7.9	3.2
	7	6.3	15.9	11.1	4.8	3.2	6.3	4.8	9.5	0.0	4.8	3.2	4.8	1.6	1.6	22.2
	8	9.5	22.2	9.5	3.2	0.0	6.3	4.8	4.8	0.0	9.5	7.9	7.9	1.6	4.8	7.9
	9	6.3	14.3	4.8	9.5	3.2	7.9	3.2	4.8	6.3	1.6	14.3	6.3	3.2	11.1	3.2
	10	4.8	11.1	7.9	3.2	12.7	0.0	9.5	4.8	4.8	11.1	3.2	7.9	4.8	1.6	12.7
	11	6.3	12.7	14.3	3.2	4.8	9.5	3.2	9.5	1.6	4.8	9.5	6.3	9.5	1.6	3.2
	12	4.8	0.0	4.8	7.9	15.9	0.0	6.3	6.3	12.7	6.3	1.6	11.1	4.8	0.0	17.5
	13	4.8	1.6	11.1	6.3	3.2	7.9	1.6	9.5	6.3	3.2	4.8	11.1	7.9	3.2	17.5
	14	7.9	1.6	0.0	4.8	0.0	17.5	0.0	1.6	11.1	3.2	15.9	4.8	3.2	28.6	0.0
	15	9.5	6.3	9.5	1.6	9.5	4.8	6.3	3.2	6.3	7.9	3.2	7.9	11.1	11.1	1.6

°See text for the correspondence between the ID number and the title of the clip/painting.

(Francisco Tárrega, *Recuerdos de la Alhambra*) was strongly associated with image no. 5 (“Full moon”); clip no. 4 (Isaac Albéniz, *Leyenda*) was strongly associated with image no. 14 (“Leopard”); clip no. 14 (Villa Lobos, *Study n. 6*) was strongly associated with image no. 14 (“Leopard”); clip no. 1 (Villa Lobos, *Prelude n. 4*) was strongly associated with image no. 10 (“Sea II”); clip no. 7 (Fernando Sor, *Variations on a theme by Mozart—Theme*) was strongly associated with image no. 15 (“Matilda and Beatrice”). As an example, **Figure 3** shows the scatterplot of the ratings given to the 22 adjectives for clip 3 (on the vertical axis) and image 5 (on the horizontal axis).

Most of the adjectives show a linear pattern, with low values for “happy,” “agitated,” “quick,” “presto,” and “strong” and high values for “sad,” “horizontal,” “slow,” “calm,” and “continuous.”

As a second example, **Figure 4** shows the scatterplot of the ratings given to the 22 adjectives for clip 14 (on the vertical axis) and image 14 (on the horizontal axis).

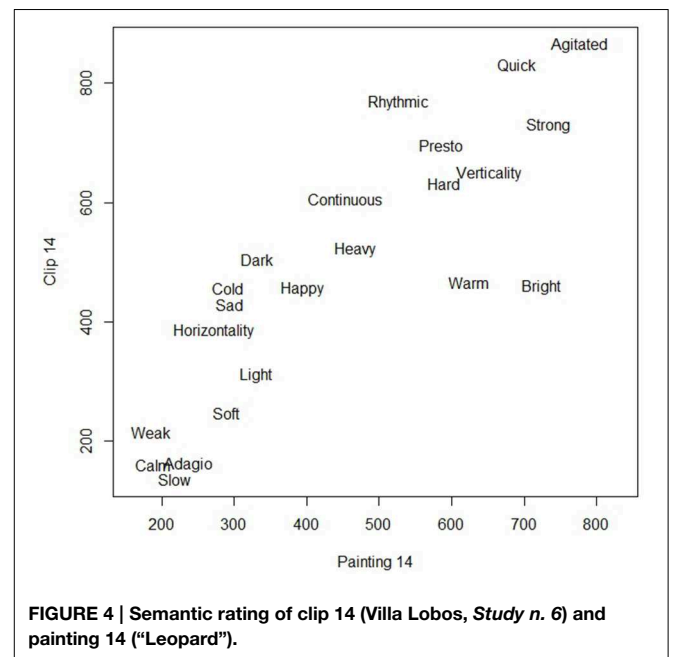
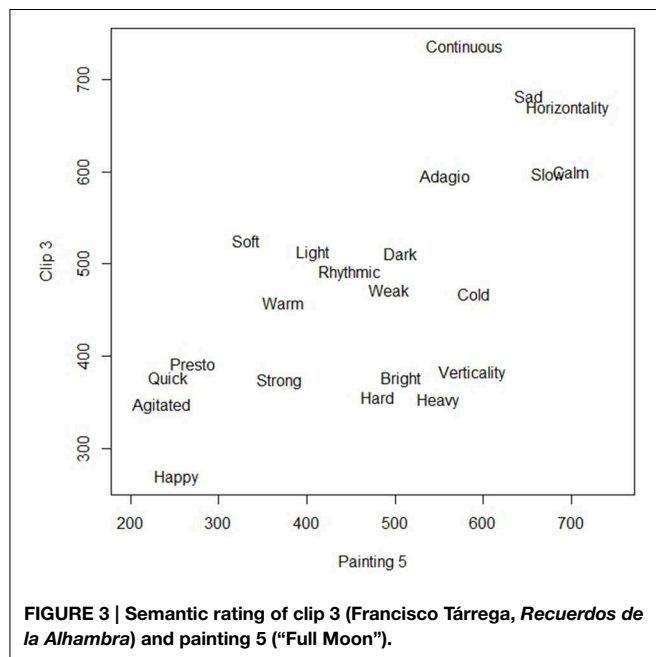
Also in this case the adjectives show a linear pattern, with low values for “calm,” “slow,” “adagio,” and “weak,” and high values for “agitated,” “quick,” “strong,” and “presto.”

The study considered stimuli of great complexity. Consequently, we chose to have the maximum possible uniformity of

TABLE 5 | Standardized residuals of the contingency table between the music clips and the first choice of a painting (Residuals greater than 1.96 in absolute value are shown in bold).

		Painting°														
Music Clip°		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	1	−0.80	−0.79	−1.07	0.52	1.99	−1.67	−1.77	−0.41	−0.47	4.35	−0.54	1.75	1.55	−2.32	−0.06
	2	1.20	1.48	0.00	−0.08	−1.67	1.46	3.29	0.15	−0.47	−1.04	0.41	−1.28	−1.36	−1.33	−0.54
	3	0.20	−1.69	−1.60	−0.68	7.74	−1.67	1.39	0.70	0.07	−1.04	−1.49	1.24	0.39	−2.32	−0.54
	4	1.20	−1.24	−1.60	−0.68	−1.67	1.46	−1.77	−2.06	0.07	−0.50	2.32	−0.77	−1.36	7.14	−1.49
	5	0.70	−1.24	−1.60	−0.08	−1.67	0.42	0.13	1.80	2.25	−1.04	−0.06	−0.27	0.97	−0.33	0.41
	6	−0.80	2.39	0.53	1.12	−2.20	−0.10	−1.14	−0.41	0.62	0.57	2.32	−1.78	−0.19	0.17	−1.49
	7	−0.30	1.93	1.60	−0.08	−1.15	−0.10	0.13	1.25	−2.10	−0.50	−1.49	−0.77	−1.36	−1.83	4.23
	8	0.70	3.75	1.07	−0.68	−2.20	−0.10	0.13	−0.41	−2.10	1.11	−0.06	0.24	−1.36	−0.83	−0.06
	9	−0.30	1.48	−0.53	1.72	−1.15	0.42	−0.51	−0.41	0.07	−1.58	1.84	−0.27	−0.78	1.16	−1.49
	10	−0.80	0.57	0.53	−0.68	1.99	−2.20	2.02	−0.41	−0.47	1.65	−1.49	0.24	−0.19	−1.83	1.37
	11	−0.30	1.03	2.67	−0.68	−0.63	0.94	−0.51	1.25	−1.56	−0.50	0.41	−0.27	1.55	−1.83	−1.49
	12	−0.80	−2.60	−0.53	1.12	3.03	−2.20	0.76	0.15	2.25	0.04	−1.97	1.24	−0.19	−2.32	2.80
	13	−0.80	−2.15	1.60	0.52	−1.15	0.42	−1.14	1.25	0.07	−1.04	−1.02	1.24	0.97	−1.33	2.80
	14	0.20	−2.15	−2.14	−0.08	−2.20	3.56	−1.77	−1.51	1.70	−1.04	2.32	−0.77	−0.78	6.64	−2.45
	15	0.70	−0.79	1.07	−1.28	0.94	−0.63	0.76	−0.96	0.07	0.57	−1.49	0.24	2.14	1.16	−1.97

°See text for the correspondence between the ID number and the title of the clip/painting.



style, composition and musical coloratura. The results show that the associations were made on specific characteristics that the subjects perceived as similar between the paintings and the music clips.

In particular, the associations between paintings and music clips proved to be consistent (even within the triads of images selected in the associations). Specifically, a strong positive association was found between clip no. 4 (Isaac Albéniz, *Asturias*) and image no. 14 ("Leopard"); between clip no. 14 (Villa Lobos, *Study n. 6*) and image no. 14 ("Leopard"); between clip no.

1 (Villa Lobos, *Prelude n. 4*) and image no. 10 ("Sea II"). To be noted is that the tempo of the first two music clips was either presto or prestissimo, and the images associated with them showed high values for the adjectives "quick," "agitated," and "strong." The strongest negative association was instead between clip no. 3 (moderate tempo) and image no. 14 ("Leopard"). Clip no. 3 was associated with characteristics such as "continuous," "slow," "calm," while the image "Leopard" was associated with opposite characteristics such as "agitated," "bright," "quick," and "presto." Considering the results presented in **Figures 3, 4,**

the attributes that seemed to play the most significant role in the associations obtained were “calm,” “agitated,” “slow,” “quick,” “strong,” “presto,” “adagio.” Consequently, relevant features in the association with paintings seem to be the timbre and the musical tempo, as shown by the positive associations between clip no. 3 (Francisco Tárrega, *Recuerdos de la Alhambra*) and painting no. 5 (“Full Moon”), and between clip no. 14 (Villa Lobos, *Study n. 6*) and painting no. 14 (“Leopard”); and by the negative association between clip no. 3 and painting no. 14.

Contrary to what was expected (Bresin, 2005), the results instead show that the musical mode usually related to feelings of happiness (major mode), or to feelings of sadness (minor mode), and the spatial orientation (vertical and horizontal) as expressed by the attributes tested with the semantic differential, did not play a significant role in the association. Finally, no substantial difference was apparent between expert and non-expert subjects. Because all the images had highly saturated colors, we did not test for a potential association between these color dimensions and major and minor modes, or slow and fast music (see Bresin, 2005; Palmer et al., 2013). The purpose of our study was not to analyse the production of visual representation of a sequence of sounds in simple drawings as in Küssner (2013b; see also Küssner, 2013a), but the association of highly complex paintings and musical pieces of classical music. The two studies are then only partially comparable because our goal was not the visualization of music. What we asked the subjects to do was associate highly complex Gestalten in the visual and acoustic fields (not single parameters such as pitch and loudness) while listening to the clips. In other words, the task was much more complex and closer to the natural global perception of stimuli in the environment (in this case, of an artistic kind). Also different from Küssner (2013b) was the expertise of the participants; in fact we tested experts in music, non-experts in music, and art experts, but obviously we did not test experts in dance because our goal was not to test the motor action aspects of the associations (see also Maes et al., 2014). As to the study conducted by Cowles (1935), there were differences in the number and the kind of stimuli, in the aims and in the methodology: in Cowles’ test the pictures, as mentioned, were mainly landscapes or scenes with simple content, while there was greater uniformity in our stimuli as to the paintings (which were by the same artist, and in the same style, materic and expressionist) and the music (all our clips were taken from Spanish classical guitar music). The contents of the paintings, instead, were different. In our experiment, besides the cross-modal association between auditory and visual stimuli, we also made use of the semantic differential method. But similarly to Cowles, our results showed no difference between experts and non-experts. The methods used in the experiment (i.e., semantic differential and subjective judgements) corroborated the interpretation of the results as associations due to patterns of qualitative similarity present in stimuli of different sensory modalities and experienced as such by the subjects (Albertazzi et al., 2013, 2014). Also in this respect the methodology that we used differed from the standard ones: we did not rely on psychophysical methods, reaction times (as in Marks, 2004; Spence, 2011), and forced choice responses (Walker, 1987); and we obviously did not make use of computational technologies. Our aim

was to remain as close as possible to the natural perception of auditory and visual items. As said, the tested adjectives very frequently exhibited a linear pattern in the association between the paintings and the music clips: for example, having low values for “happy,” “agitated,” “quick,” “presto,” and strong and high values for “sad,” “horizontal,” “slow,” “calm,” and “continuous.” On the basis of these findings, and the fact that we didn’t find any difference between expert and non-expert subjects, the tested semantic connotations of the stimuli might be considered as affordances playing the role of *general semantic information clues*, which makes perfect sense in a framework of an *ecology of meaning*. It has been recently shown, for example, that subjects in the general population group natural shapes on the basis of certain visual qualitative characteristics: specifically, non-spiculed, non-holed, and flat shapes are experienced and classified as harmonic and static, while rounded shapes are classified as harmonic and dynamic, and elongated shapes as somewhat disharmonious and somewhat static (Albertazzi et al., 2014). Because of the complex nature of the stimuli, and on the basis of our results, one can conclude that there are aesthetic, sometimes ideasthetic dimensions in perceptual awareness. These dimensions act as Gestalten or templates playing the role of an immediate understanding of the complex objects we usually encounter in the environment. Furthermore, these Gestalten exhibit common patterns in the different modalities, as we have found in our study.

Finally, in our study we did not specifically test the emotional response, as in Cowles (1935), Di Dio and Gallese (2009), Juslin and Sloboda (2001), Krumhansl and Lerdahl (2011), Langlois et al. (2014), Madison (2011), Palmer et al. (2013) and Zaidel (2010), because it was not our primary interest. However, some of the adjectives tested with the semantic differential test, such as “calm” and “agitated,” “happy” and “sad” proved to have an important role.

In light of the overall results, one cannot exclude the presence of potential top-down influences (however unconscious), although our study did not aim to investigate these aspects. In this regard, what we did in our experiments was to invite the subjects to be as careful as possible to avoid the influence of past experience.

On the basis of our results, it is likely that the choice of a different number of adjectives restricted to a small number of characteristics, and limiting the range of associations and the length of the experiment, might yield further consistent information about the cross-modal associations obtained. Presenting adjectives in pairs, like calm/agitated, weak/strong, might also contribute to shortening the duration of the test. However, such a choice would have overestimated the correlation which in our study is also possibly overestimated, because the adjectives were not entirely independent. It is also likely that choosing a more uniform theme for the paintings (only landscapes, for example) would make the test shorter. A further development of the design might consist in testing the associations between the paintings and a series of music clips from a different musical repertoire, reducing the uniformity of patterns. Nevertheless, it seems worthwhile to continue testing cross-modal associations in complex stimuli, because these are usually experienced in perceiving. Finally, it

would be advisable to repeat the experiment with subjects from other cultures, such as oriental ones, in order to test for the presence of possible pictorial and musical biases in the associations found.

In conclusion, our study shows (i) the existence of cross-modal associations between complex visual and auditory stimuli, (ii) the existence of associations between visual and auditory stimuli when evaluated employing the semantic differential, and (iii) that these associations were at least partially consistent with each other. These findings corroborate the interpretation that the associations are partially due to patterns of qualitative similarity present in stimuli of different sensory modalities.

References

- Albertazzi, L. (2013). "Experimental phenomenology. An introduction," in *The Wiley Blackwell Handbook of Experimental Phenomenology. Visual Perception of Shape, Space and Appearance*, ed L. Albertazzi (London: Wiley-Blackwell), 1–36. doi: 10.1002/9781118329016
- Albertazzi, L., Canal, L., Dadam, J., and Micciolo, R. (2014). The semantics of biological forms. *Perception* 43, 1365–1376. doi: 10.1068/p7794
- Albertazzi, L., Canal, L., Malfatti, M., and Micciolo, R. (2015). The hue of angles. Was Kandinsky right? *Art Percept.* 3, 81–92. doi: 10.1163/22134913-00002025
- Albertazzi, L., Da Pos, O., Canal, L., Micciolo, R., Malfatti, M., and Vescovi, M. (2013). The hue of shapes. *J. Exp. Psychol. Hum. Percept. Perform.* 39, 37–47. doi: 10.1037/a0028816
- Baron-Cohen, S., and Harrison, J. E. (1997). *Synesthesia: Classic and Contemporary Readings*. Cambridge, MA: Blackwell.
- Bidaine, P. (Ed.). (2004). *Sons & Lumières. Une Histoire du son Dans l'art du XXe Siècle*. Paris: Éditions du Centre Pompidou.
- Bresin, R. (2005). "What is the color of that music performance?," in *Proceedings of the International Computer Music Conference-ICMC 2005* (International Computer Music Association, San Francisco, CA.)
- Canal, L., and Micciolo, R. (2013). "Measuring the immeasurable: quantitative analyses of perceptual experiments," in *Handbook of Experimental Phenomenology. Visual Perception of Shape, Space and Appearances*, ed L. Albertazzi (London: Wiley-Blackwell), 477–498. doi: 10.1002/9781118329016.ch20
- Chen, N., Tanaka, K., and Watanabe, K. (2015). Color-shape associations revealed with implicit association tests. *PLoS ONE* 10:e0116954. doi: 10.1371/journal.pone.0116954
- Cohen Kadosh, R. C., Henik, A., and Walsh, V. (2009). Synesthesia: learned or lost? *Dev. Sci.* 12, 484–491. doi: 10.1111/j.1467-7687.2008.00798.x
- Cowles, J. T. (1935). An experimental study of the pairing of certain auditory and visual stimuli. *J. Exp. Psychol.* 18, 461–469. doi: 10.1037/h0062202
- Cytowic, R. E. (1995). *Synesthesia: A Union of the Senses, 2nd Edn*. Cambridge MA: MIT Press.
- Dadam, J., Albertazzi, L., Da Pos, O., Canal, L., and Micciolo, R. (2012). Morphological patterns and their colours. *Percept. Mot. Skills* 114, 363–377. doi: 10.2466/03.22.23.PMS.114.2.363-377
- Demattè L., Sanabria, D., and Spence, C. (2006). Cross-modal associations between odors and colours. *Chem. Senses* 31, 531–538. doi: 10.1093/chemse/bjj057
- Deroy, O., Crisinel, A., and Spence, C. (2013). Crossmodal correspondences between odours and contingent features: odours, musical notes and abstract shapes. *Psychon. Bull. Rev.* 20, 1–19. doi: 10.3758/s13423-013-0397-0
- Deroy, O., and Spence, C. (2013). Why we are not all synesthetes (not even weakly so). *Psychon. Bull. Rev.* 20, 643–664. doi: 10.3758/s13423-013-0387-2
- Di Dio, C., and Gallese, V. (2009). Neuroaesthetics: a review. *Curr. Opin. Neurobiol.* 19, 682–687. doi: 10.1016/j.conb.2009.09.001
- Dixon, M. J., Smilek, D., Duffy, P. L., Zanna, M. P., and Merikle, P. M. (2006). The role of meaning in grapheme-colour synesthesia. *Cortex* 42, 243–252. doi: 10.1016/S0010-9452(08)70349-6
- Droste, M. (1990). *Bauhaus*. Köln: Taschen.
- Eagleman, D. M. (2010). Synaesthesia. *Br. Med. J.* 340, b4616. doi: 10.1136/bmj.b4616
- Eagleman, D. M. (2012). Synesthesia in its protean guises. *Br. J. Psychol.* 103, 16–19. doi: 10.1111/j.2044-8295.2011.02020.x
- Gage, J. (1999). *Color and Meaning: Art, Science and Symbolism*. Berkeley, CA: University of California Press.
- Gallace, A., Boschini, E., and Spence, C. (2011). On the taste of "bouba" and "kiki": an exploration of word-food associations in neurologically normal participants. *Cogn. Neurosci.* 2, 34–46. doi: 10.1080/17588928.2010.516820
- Gilbert, A. N., Martin, R., and Kemp, S. E. (1996). Cross-modal correspondence between vision and olfaction: the color of smells. *Am. J. Psychol.* 109, 335–351.
- Hanson-Vaux, G., Crisinel, A.-S., and Spence, C. (2013). Smelling shapes: cross-modal correspondences between odors and shapes. *Chem. Senses* 38, 161–166. doi: 10.1093/chemse/bjs087
- Itten, J., Birren, F., and van Haagen, E. (1970). *The Elements of Color: A Treatise on the Color System of Johannes Itten Based on his Book The Art of Color*. New York, NY: Van Nostrand Reinhold.
- Jacobsen, T. H. (2002). Kandinsky's questionnaire revisited: fundamental correspondence of basic colors and forms? *Percept. Mot. Skills* 95, 903–913. doi: 10.2466/pms.2002.95.3.903
- Jürgens, U. M., and Nikolić, D. (2012). Ideasthesia: conceptual processes assign similar colours to similar shapes. *Translat. Neurosci.* 3, 22–27. doi: 10.2478/s13380-012-0010-4
- Jürgens, U. M., and Nikolić, D. (2014). "Synesthesia as an Ideasthesia - cognitive implications," in *Synesthesia and Children-Learning and Creativity*, eds J. R. Sinha and C. Söffing (Kassel: Kassel University Press), (in press).
- Juslin, P. N., and Sloboda, J. A. (2001). *Music and Emotion: Theory and Research*. Oxford: Oxford University Press.
- Kandinsky, W. (1926/1994). "Point and line to plane (P. Vergo, Trans.)," in *Kandinsky, Complete Writings on Art*, eds K. C. Lindsay and P. Vergo (New York, NY: Da Capo Press), 524–700.
- Kemp, S. E., and Gilbert, A. N. (1997). Odor intensity and color lightness are correlated sensory dimensions. *Am. J. Psychol.* 110, 35–46. doi: 10.2307/1423699
- Kharkhurin, A. V. (2012). Is triangle really yellow? An empirical investigation of Kandinsky's correspondence theory. *Empir. Stud. Arts* 30, 167–182. doi: 10.2190/EM.30.2.d
- Klee, P. (1956). *Das bildnerische Denken. Schriften zur Form und Gestaltungslehre*. Basel-Stuttgart: Jürg Spiller editor.
- Köhler, W. (1929). *Gestalt Psychology*. New York, NY: Liveright.
- Krumhansl, C. H., and Lerdahl, F. (2011). "Musical tension," in *Art and the Senses*, eds F. Bacci and D. Melcher (Oxford: Oxford University Press), 311–328.
- Küssner, M. B. (2013a). "Shaping music visually," in *Kreativität – Struktur und Emotion*, eds A. C. Lehmann, A. Jeßulat, and C. Wünsch (Würzburg: Königshausen & Neumann), 203–209.
- Küssner, M. B. (2013b). Music and shape. *Lit. Linguist. Comput.* 28, 472–479. doi: 10.1093/lc/fqs071

Acknowledgments

We wish to thank Matteo Boato for giving us permission to use images of his paintings, for selecting the music clips and performing them, and for participating in the preliminary discussions. We also thank Pietro Chasseur for producing the high-resolution images of Boato's paintings used in the experiment.

Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2015.00424/abstract>

- Langlois, T., Peterson, J., and Palmer, S. (2014). *Visual Texture, Music and Emotion. Abstract presented at the 13th Annual Meeting of the Vision Sciences Society* (St. Pete Beach, FL).
- Levitin, C. A., In Ren, J., Woods, A., Boesveldt, S., Chan, J. S., McKenzie, J., et al. (2014). Cross-cultural color-odor associations. *PLoS ONE* 9:e101651. doi: 10.1371/journal.pone.0101651
- Ludwig, V. U., and Simner, J. (2013). What colour does that feel? Tactile-visual mapping and the development of cross-modality. *Cortex* 59, 1089–1099. doi: 10.1016/j.cortex.2012.04.004
- Lupton, E., and Miller, J. A. (1991). *The ABC's of Design: The Bauhaus and Design Theory*. New York, NY: Princeton Architectural Press.
- Madison, G. (2011). "Cause and effect: a functional perspective on music and emotion," in *Art and the Senses*, eds F. Bacci and D. Melcher (New York, NY: Oxford University Press), 329–350.
- Maes, P.-J., Leman, M., Palmer, C., and Wanderley, M. M. (2014). Action-based effects on music perception. *Front. Psychol.* 4:1008. doi: 10.3389/fpsyg.2013.01008
- Makin, A. D. J., and Wuerger, S. (2013). The IAT shows no evidence for Kandinsky's color-shape associations. *Front. Psychol.* 4:616. doi: 10.3389/fpsyg.2013.00616
- Marks, L. E. (2004). "Cross-modal interactions in speeded classification," in *Cross-Modal Interactions in Speeded Classification*, eds G. A. Calvert, C. Spence, and B. E. Stein (Cambridge, MA: MIT Press), 85–105.
- Martino, G., and Marks, L. E. (2001). Synesthesia: strong and weak. *Curr. Dir. Psychol. Sci.* 10, 61–65. doi: 10.1111/1467-8721.00116
- Maurer, D., and Mondloch, C. (2005). "Neonatal synesthesia: A reevaluation," in *Perspective from Cognitive Neuroscience*, eds L. Robertson and N. Sagiv (Oxford: Oxford University Press), 193–213.
- Meier, B. (2013). Semantic representation of synesthesia. *Theor. Hist. Sci.* 10, 125–134. doi: 10.12775/ths-2013-0006
- Melara, R. D., and O'Brien, T. P. (1987). Interactions between synesthetically corresponding dimensions. *J. Exp. Psychol. Gen.* 116, 323–336.
- Moos, A., Simmons, D., Simner, J., and Smith, R. (2013). Color and texture associations in voice-induced synaesthesia. *Front. Psychol.* 4:568. doi: 10.3389/fpsyg.2013.00568
- Moos, A., Smith, R., Miller, S. R., and Simmons, D. P. (2014). Cross-modal associations in synesthesia: vowel colours in the ear of the beholder. *i-perception* 5, 132–142. doi: 10.1068/i0626
- Mroczo-Wąsowicz, A., and Nikolić, D. (2014). Semantic mechanisms may be responsible for developing synesthesia. *Front. Hum. Neurosci.* 8:509. doi: 10.3389/fnhum.2014.00509
- Mroczo-Wąsowicz, A., and Werning, M. (2012). Synesthesia, sensory-motor contingency, and semantic emulation: how swimming style-color synesthesia challenges the traditional view of synesthesia. *Front. Psychol.* 3:279. doi: 10.3389/fpsyg.2012.00279
- Nielsen, A., and Rendall, D. (2011). The sound of round: evaluating the sound-symbolic role of consonants in the classic Takete-Maluma phenomenon. *Can. J. Exp. Psychol.* 65, 115–124. doi: 10.1037/a0022268
- Osgood, C. E. (1956). *Method and Theory in Experimental Psychology*. Oxford: Oxford University Press.
- Palmer, S. E., and Schloss, K. B. (2010). An ecological valence theory of color preferences. *Proc. Natl. Acad. Sci. U.S.A.* 107, 8877–8882. doi: 10.1073/pnas.0906172107
- Palmer, S. E., Schloss, K. B., Xu, Z., and Prado-León, L. R. (2013). Music-colour associations are mediated by emotion. *Proc. Natl. Acad. Sci. U.S.A.* 110, 8836–8841. doi: 10.1073/pnas.1212562110
- Parise, C., and Spence, C. (2009). 'When birds of a feather flock together': synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PLoS ONE* 4:e5664. doi: 10.1371/journal.pone.0005664
- R. Core Team. (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Reybrouck, M., Verschaffel, L., and Lauwerier, S. (2009). Children's graphical notations as representational tools for musical sense-making in a music-listening task. *Br. J. Music Educ.* 26, 189–211. doi: 10.1017/S0265051709008432
- Sagiv, N., Simner, J., Collins, J., Butterworth, B., and Ward, J. (2006). What is the relationship between synaesthesia and visuo-spatial number forms? *Cognition* 101, 114–128. doi: 10.1016/j.cognition.2005.09.004
- Sagiv, N., and Ward, J. (2006). Crossmodal interactions: lessons from synesthesia. *Prog. Brain Res.* 155, 259–271. doi: 10.1016/S0079-6123(06)55015-0
- Schönberg, A., and Kandinsky, W. (1980). *Briefe, Bilder und Dokumente einer aussergewöhnlichen Begegnung*. St. Pölten: Residenz Verlag.
- Simner, J. (2012). Defining synaesthesia. *Br. J. Psychol.* 103, 1–15. doi: 10.1348/000712610X528305
- Simner, J., Gartner, O., and Taylor, M. D. (2011). Cross-modal personality attributions in synaesthetes and non-synaesthetes. *J. Neuropsychol.* 5, 283–301. doi: 10.1111/j.1748-6653.2011.02009.x
- Simner, J., Lanz, M., Jansari, A., Noonan, K., Glover, L., Oakley, D. A., et al. (2005). Nonrandom associations of graphemes to colours in synaesthetic and normal populations. *Cogn. Neuropsychol.* 22, 1069–1085. doi: 10.1080/02643290500200122
- Simner, J., Mulvenna, C., Sagiv, N., Tsakanikos, E., Witherby, S. A., and Fraser, C. (2006). Synaesthesia: the prevalence of atypical cross-modal experiences. *Perception* 35, 1024–1033. doi: 10.1068/p5469
- Simner, J., and Ward, J. (2006). Synesthesia: the taste of words on the tip of the tongue. *Nature* 444:438. doi: 10.1038/444438a
- Spector, F., and Maurer, D. (2008). The colour of Os: naturally biased associations between shape and colour. *Perception* 37, 841–847. doi: 10.1068/p5830
- Spector, F., and Maurer, D. (2011). The colours of the alphabet: naturally biased associations between shape and colour. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 484–495. doi: 10.1037/a0021437
- Spence, C. (2011). Crossmodal correspondences: a tutorial review. *Atten. Percept. Psychophys.* 73, 971–995. doi: 10.3758/s13414-010-0073-7
- Tan, S., and Kelly, M. (2004). Graphic representations of short musical compositions. *Psychol. Music* 32, 191. doi: 10.1177/0305735604041494
- Tomson, S. N., Narayan, M., Allen, G. I., and Eagleman, D. (2013). Neural networks of colored sequence synesthesia. *J. Neurosci.* 33, 14098–14106. doi: 10.1523/JNEUROSCI.5131-12.2013
- Walker, R. (1987). The effects of culture, environment, age, and musical training on choices of visual metaphors for sound. *Percept. Psychophys.* 42, 491–502. doi: 10.3758/BF03209757
- Ward, J. (2013). Synesthesia. *Annu. Rev. Psychol.* 64, 49–75. doi: 10.1146/annurev-psych-113011-143840
- Ward, J., Huckstep, B., and Tsakanikos, E. (2006a). Sound-colour synaesthesia: to what extent does it use cross-modal mechanisms common to us all? *Cortex* 42, 264–280. doi: 10.1016/S0010-9452(08)70352-6
- Ward, J., Li, R., Salih, S., and Sagiv, N. (2007). Varieties of grapheme-colour synaesthesia: a new theory of phenomenological and behavioural differences. *Conscious. Cogn.* 16, 913–931. doi: 10.1016/j.concog.2006.09.012
- Ward, J., and Simner, J. (2003). Lexical-gustatory synaesthesia. Linguistic and conceptual factors. *Cognition* 89, 237–261. doi: 10.1016/S0010-0277(03)00122-7
- Ward, J., Tsakanikos, E., and Bray, A. (2006b). Synaesthesia for reading and playing musical notes. *Neurocase* 12, 27–34. doi: 10.1080/13554790500473672
- Whiteford, K., Schloss, K. B., and Palmer, S. E. (2013). "Music-Color associations from Bach to the Blues: Emotional mediation in synesthetes and non-synesthetes," in *Poster Presented at the 13th Annual Meeting of the Vision Science Society* (Neaples).
- Wicker, F. W. (1968). Mapping the intersensory regions of perceptual space. *Am. J. Psychol.* 81, 178–188. doi: 10.2307/1421262
- Zaidel, D. W. (2010). Art and brain: insights from neuropsychology, biology and evolution. *J. Anat.* 216, 177–183. doi: 10.1111/j.1469-7580.2009.01099.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.

Copyright © 2015 Albertazzi, Canal and Micciolo. 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 coherent and fluent mind: how unified consciousness is constructed from cross-modal inputs via integrated processing experiences

Piotr Winkielman^{1,2,3*}, Michał Ziembowicz³ and Andrzej Nowak⁴

¹ Psychology Department, University of California, San Diego, San Diego, CA, USA

² Behavioural Science Group, Warwick Business School, University of Warwick, Coventry, UK

³ Department of Psychology, University of Social Sciences and Humanities, Warsaw, Poland

⁴ Center for Complex Systems and New Technologies, Institute for Social Studies, University of Warsaw, Warsaw, Poland

*Correspondence: pwinkiel@ucsd.edu

Edited by:

Aleksandra Mroczko-Wasowicz, National Yang Ming University, Taiwan

Reviewed by:

Sascha Topolinski, University of Cologne, Germany

Teresa Garcia-Marques, Instituto Superior de Psicologia Aplicada-Instituto Universitário, Portugal

Keywords: consciousness, coherence, fluency, subjective experience, cross-modal perception

Many philosophical approaches hypothesize that one function of consciousness is the creation of a unified subjective experience (Baars, 2005; Bayne, 2010). Such unified experience links different processing streams, originating in separate perceptual modules, thus enabling common access and generation of integrated decisions. All of this presumably occurs via a mechanism that blends information from different modalities into a single, multidimensional representation. But what exactly is unified in conscious experience? Prevailing explanations focus on integration of specific stimulus features at a perceptual or decisional level. In this opinion piece we discuss a simple but underappreciated explanation that focuses on processing dynamics. Specifically, we propose that cross-modal integration is facilitated by different modalities having a similar effect on the global subjective experience of processing quality. This integrated experience can then enter into decisional processes concerned with its source and relevance for the current behavior. As such, our account combines “experiential” and “decisional” process. Below we place this argument in the context of research on cross-modal integration and processing experiences, and discuss some implications.

Traditionally, research on multisensory experiences focuses on integration of information from different perceptual and

conceptual cues. Some classic examples of such phenomena include the McGurk effect (changes in audition as function of vision; McGurk and MacDonald, 1976) and the double-flash illusion (changes in vision as a function of audition; e.g., Shams et al., 2000). Other classic examples of low-level cross-modal interactions include influences between pitch and brightness, loudness and size, or pitch and elevation. On a more conceptual level, cross-modal influences include shape or sound symbolism, such as the “bouba/kiki” effect (Ramachandran and Hubbard, 2001) and semantically-driven cases of synesthesia (Mroczko-Wasowicz and Nikolich, 2014). A lively debate concerns when individual perceptual components from one modality are mandatorily modified by another modality, undergoing a low-level fusion that produces a single integrated percept, or when they are separate and integrated in high-level, post-perceptual stages via decisional processes (Spence, 2011; for an empirical example, see Hillis et al., 2002). Importantly, what such studies investigate are cross-modal influences on the representational content related to specific stimulus features.

Here we propose that cross-modal influences can also occur via processes that care less about the specific representational content but more about general representational quality, yielding global processing experiences. This proposal

is grounded in several theoretical and empirical considerations.

Historically, the basic idea of processing experiences goes back to William James (1890) who spoke of “fringe consciousness” as experience that communicates a vague, unarticulated sense of peripheral contents relevant to the main task. Some “fringe experiences” include the feelings of familiarity and knowing, tip-of-the-tongue phenomena, and the sense of ease, rightness or coherence. Initially neglected by cognitive science, processing experiences and global “quality signals” are now of interest as a computationally efficient way of representing rich relational information (Mangan, 1993; Reber et al., 2002).

Empirically, the initial evidence for processing experiences came from research on fluency and familiarity (Whittlesea, 2002). For example, a pioneering study observed that people judge variable background noise as less loud when they hear a target word that was previously studied (Jacoby et al., 1988). Apparently, the ease (fluency) of target processing, deriving from previous exposure, gets misattributed to the loudness judgment. A related study reported misattributions of previous exposure to visual blur judgments (Whittlesea et al., 1990). Subsequent memory research documented similar influences using changes in perceptual format between stimuli appearing in the study and test phase of the

experiment, including crossing words and pictures (e.g., *Fazendeiro et al., 2005*). Critically for the present argument, similar effects can occur for changes in modality, such as crossing auditory and visual stimulus presentation at study and test (e.g., *Curran and Dien, 2003; Miller et al., 2008*). These studies suggest that subjective experiences such as “fluency” and “familiarity” can be amodal and reflect joint influences from separate modalities. As a result, people cannot easily separate the processing quality associated with the target stimulus from contextual influences.

An important inspiration for our proposal are findings that a similar subjective experience can derive from processing facilitation at different processing stages. For example, factors that objectively facilitate visual detection and visual identification have similar effects on feelings of processing ease (*Reber et al., 2004; Wurtz et al., 2008*; but see *Reber et al., 2014*). As such, our proposal basically adds that experiential integration into a unified subjective feeling can occur even when the sources of processing experiences originate in different sensory modalities (e.g., quality signals from auditory processes can combine with quality signals from visual processes).

Importantly, our proposal assumes that experiential signals of processing quality can originate in processing of abstract, conceptual material, and extend beyond fluency (sense of ease) and familiarity (sense of oldness) to “structural experiences,” such as a sense of coherence, integrity, or rightness (*Whittlesea, 2002*). One example comes from research using artificial grammars and shows that decisions about grammaticality in one modality are influenced by previously learned grammatical rules in another modality, and that this influence involves non-analytical processes (*Dienes et al., 2011*). Importantly, this effect may not involve a feeling of fluency or familiarity, but rather a sense of structural coherence (*Scott and Dienes, 2010*). Stressing the breadth of such effects, our recent research shows that decisions about patterns in one modality can be influenced by the coherence of completely unrelated patterns from another modality (*Ziembowicz et al., 2013*). Let us elaborate as this research illustrates our

core argument. In three experiments participants judged targets in one sensory modality while being incidentally exposed to regular or irregular background stimuli from a different modality. For example, targets were auditory melodies and backgrounds were visual figures, or vice versa. Critically, the specific regularity of targets and backgrounds was unrelated—auditory regularity was tone sequence grammar, visual regularity was 3D realizability. We explored the effect of cross-modal coherence with different types of subjective judgments: “regularity” (Experiment 1), “familiarity” (Experiment 2), and “possibility” (Experiment 3). All three experiments showed similar results: the coherence of the background stimulus influenced the target judgment, regardless of judgment type and target modality. That is, visual and auditory targets were judged as more “regular,” “familiar,” and “possible” when the incidental cross-modal backgrounds were coherent.

What are the implications of such findings? As mentioned, the standard explanation of cross-modal phenomena assumes changes in representation of stimulus features, whether driven by perceptual processes or decisional processes that integrate cues from different modalities. In contrast, we argue that cross-modal influences also reflect integration at the level of processing experiences. We admit the need for direct evidence that the just discussed cross-modal studies (including *Ziembowicz et al., 2013*) involve changes in subjective experiences and that their integration is causally responsible for the obtained behavioral effects. However, there is good evidence that related phenomena do involve “experiences”—i.e., cognitive or affective feelings. First, participants in many (though not all) experiments actually report changes in the feeling of “ease,” “effort,” “familiarity” or “regularity” associated with processing (*Schwarz, 2015*). Second, various physiological measures pick up indicators of changes in experience, such as positivity associated with fluent processing (e.g., *Winkielman et al., 2003, 2012*). Third, many experiments show that “bleed-over” or “misattribution” effects vanish once a person is provided with an explanation targeting subjective experience, not unlike classic

studies on misattribution and discounting of affect or arousal (e.g., *Dutton and Aron, 1974*). For example in the previously mentioned cross-format study of *Fazendeiro et al. (2005)*, participants were asked to recognize (old/new) words and pictures, some of which appeared earlier as related cross-format stimuli (essentially serving as semantic primes). During this recognition task, background music was played, which for some participants was explained as influencing their “sense of familiarity.” In this condition, participants showed reduced false recognition judgments for the cross-format stimuli, presumably reflecting their discounting of familiarity experience. Additional evidence for the notion that participants consciously experience changes in processing quality comes from research on hidden semantic coherence and the intuitive basis of such judgments (*Topolinski and Strack, 2009a,b*). Interestingly, this work shows that participants cannot report and re-attributed changing levels of fluency (facilitation due to semantic coherence) but are only aware of affective (hedonic) consequences of changed fluency. This suggests that what specifically is “felt” about objective processing quality varies depending on the details of the task. Finally, the just mentioned studies again highlight that the integration at the level of subjective experience interacts with high-level decisional processes. That is, the exact impact of experience on stimulus judgments depends on the perceiver’s beliefs about the sources and relevance of the experience for the task at hand (*Schwarz, 2015*).

Neuroscientifically, our “joint quality signal” explanation for cross-modal integration matches evidence for global conflict signals or global prediction error (e.g., *Fernandez-Duque et al., 2000; Friston, 2010; Shackman et al., 2011; Botvinick and Braver, 2015*). Computationally, our account fits with connectionist models using global signals of processing quality (*Lewenstein and Nowak, 1989; Norman and O’Reilly, 2003; Cleeremans and Dienes, 2008*). Critically, these signals are non-specific, with different sources of coherence, ease, or familiarity generating a similar signal. Further, these signals are free-floating—not tightly bound to

the original representation, and thus transferable across contents. Still, the signals are useful. They highlight abstract correspondences across patterns (e.g., regularity). They also regulate the network's own behavior, terminating the recognition process (preventing pattern discovery) when coherence is low and letting recognition continue when coherence is high (Rychwalska et al., 2005). The specifics of the mechanisms can be illustrated using a model by Lewenstein and Nowak (1989). It is a Hopfield type neural network enhanced with a mechanism allowing the network to control its own processing dynamics. The controlling system is implemented as a feedback loop that draws on one of a set of parameters: coherence, volatility, signal strength, etc. Based on the momentary values of this "order parameter," the system can distinguish between known and unknown stimuli, but also react differently to primed, prototypical, regular, coherent, and distorted material. This model applies well to behavioral data, as seen in simulations of the mere exposure effect, which involves changes in fluency (Drogosz and Nowak, 2006). Consistent with behavioral data, the network reproduces the asymmetrical effect of "mere-exposed" stimuli on non-analytic, implicit, fluency-dependent judgments (preferences, familiarity) and analytic, explicit memory judgments. That is, the network results show that implicit measures of recognition (using the dynamic order parameter) can be faster than explicit measures, recreating a paradoxical phenomenon of somehow "knowing" the valence or familiarity of a stimulus before actually recognizing it.

In sum, we propose that some cross-modal phenomena involve integration via common experiences, including fluency, familiarity, and coherence, grounded in global signals about network dynamics. As a result, even when the modal origins of such signals differ, individuals experience integrated feelings of processing quality. Such feelings can then enter meta-cognitive processes and inform fundamental cognitive and social judgments (Winkielman and Schooler, 2011; Schwarz, 2015). Future research may explore cross-modal influences on experience-based judgments (risk, frequency, truth, fame, beauty, etc.). It should

also determine when such effects are pre- and post-decisional. One question in this regard concerns the level at which processing signals are combined. It could be pre-experiential (e.g., fluency signals could blend before any experience) or experiential (e.g., with one already blended, or two blendable feeling signals appearing in the experience). A related question is whether experiences from different sources are genuinely fused (i.e., their origin information is lost) or potentially separable. Research should also explore the specificity of experiences. That is, sometimes experiences act broadly, allowing for conflation of drastically different inputs such as physical arousal with familiarity (Goldinger and Hansen, 2005) or physical effort with retrieval difficulty (Stepper and Strack, 1993). But, individual processing experiences are also unique in subjective quality (e.g., feelings of coherence differ from familiarity or ease, not unlike different emotions). This should constrain possible experiential fusion (genuine blending) and judgmental misattributions (source errors).

In conclusion, it appears that the creation of a unified consciousness is facilitated by an experiential mechanism that combines signals of processing quality. This mechanism links diverse contents in the mind and allows people to experience the multi-modal world as integrated—though also sometimes as more (or less) unified than it actually is.

ACKNOWLEDGMENTS

Andrzej Nowak was partially supported by Polish National Science Centre (project no. DEC-2011/02/A/HS6/00231). Piotr Winkielman was supported by the UCSD Academic Senate Bridge Grant. We thank Evan Carr, Liam Kavanagh, Robert St. Louis, and Shlomi Sher for helpful comments.

REFERENCES

- Baars, B. J. (2005). Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. *Prog. Brain Res.* 150, 45–53. doi: 10.1016/S0079-6123(05)50004-9
- Bayne, T. (2010). *The Unity of Consciousness*. Oxford: Oxford University Press.
- Botvinick, M., and Braver, T. (2015). Motivation and cognitive control: from behavior to neural mechanisms. *Annu. Rev. Psychol.* 66, 83–113. doi: 10.1146/annurev-psych-010814-015044

- Cleeremans, A., and Dienes, Z. (2008). "Computational models of implicit learning," in *Handbook of Computational Cognitive Modeling*, ed R. Sun (Cambridge: University Press), 396–421.
- Curran, T., and Dien, J. (2003). Differentiating amodal familiarity from modality-specific memory processes: an ERP study. *Psychophysiology* 40, 979–988. doi: 10.1111/1469-8986.00116
- Dienes, Z., Scott, R. B., and Wan, L. L. (2011). "The role of familiarity in implicit learning," in *Constructions of Remembering and Metacognition: Essays in honour of Bruce Whittlesea*, eds P. Higham and J. Leboe (New York, NY: Algrave Macmillan), 51–62.
- Drogosz, M., and Nowak, A. (2006). A neural model of mere exposure: the EXAC mechanism. *Polish Psychol. Bull.* 37, 7–15.
- Dutton, D. G., and Aron, A. P. (1974). Some evidence for heightened sexual attraction under conditions of high anxiety. *J. Pers. Soc. Psychol.* 30, 510–517. doi: 10.1037/h0037031
- Fazendeiro, T., Winkielman, P., Luo, C., and Lorah, C. (2005). False recognition across meaning, language, and stimulus format: conceptual relatedness and the feeling of familiarity. *Mem. Cogn.* 33, 249–260. doi: 10.3758/BF03195314
- Fernandez-Duque, D., Baird, J. A., and Posner, M. I. (2000). Executive attention and metacognitive regulation. *Conscious. Cogn.* 9, 288–307. doi: 10.1006/ccog.2000.0447
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787
- Goldinger, S. D., and Hansen, W. A. (2005). Remembering by the seat of your pants. *Psychol. Sci.* 16, 525–529. doi: 10.1111/j.0956-7976.2005.01569.x
- Hillis, J. M., Ernst, M. O., Banks, M. S., and Landy, M. S. (2002). Combining sensory information: mandatory fusion within, but not between, senses. *Science* 298, 627–630. doi: 10.1126/science.1075396
- Jacoby, L. L., Allan, L. G., Collins, J. C., and Larwill, L. K. (1988). Memory influences subjective experience: noise judgments. *J. Exp. Psychol. Learn. Mem. Cogn.* 14, 240–247. doi: 10.1037/0278-7393.14.2.240
- James, W. (1890). *The Principles of Psychology*. New York, NY: Holt.
- Lewenstein, M., and Nowak, A. (1989). Recognition with self-control in neural networks. *Phys. Rev.* 40, 4652–4664. doi: 10.1103/PhysRevA.40.4652
- Mangan, B. (1993). Taking phenomenology seriously: the "fringe" and its implications for cognitive research. *Conscious. Cogn.* 2, 89–108. doi: 10.1006/ccog.1993.1008
- McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748. doi: 10.1038/264746a0
- Miller, J. K., Lloyd, M. E., and Westerman, D. L. (2008). When does modality matter? Perceptual versus conceptual fluency-based illusions in memory. *J. Mem. Lang.* 58, 1080–1094. doi: 10.1016/j.jml.2007.12.006
- Mroczko-Wąsowicz, A., and Nikolich, D. (2014). Semantic mechanisms may be responsible for developing synesthesia. *Front. Hum. Neurosci.* 8:509. doi: 10.3389/fnhum.2014.00509

- Norman, K. A., and O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: a complementary learning systems approach. *Psychol. Rev.* 110, 611–646. doi: 10.1037/0033-295X.110.4.611
- Ramachandran, V. S., and Hubbard, E. M. (2001). Synaesthesia—A window into perception, thought and language. *J. Conscious. Stud.* 8, 3–34.
- Reber, R., Christensen, B., and Meier, B. (2014). Effects of meaning and symmetry on judgments of size: an amount of information account. *Front. Psychol.* 5:1270. doi: 10.3389/fpsyg.2014.01270
- Reber, R., Fazendeiro, T. A., and Winkielman, P. (2002). Processing fluency as the source of experiences at the fringe of consciousness: commentary on Mangan. *Psyche* 8.
- Reber, R., Wurtz, P., and Zimmermann, T. D. (2004). Exploring “fringe” consciousness: The subjective experience of perceptual fluency and its objective bases. *Conscious. Cogn.* 13, 47–60. doi: 10.1016/S1053-8100(03)00049-7
- Rychwalska, A., Jablajski, P., Zochowski, M., and Nowak, A. (2005). Novelty based feedback regulation in artificial neural networks. *Acta Neurobiol. Exp.* 65, 453–463.
- Schwarz, N. (2015). “Metacognition,” in *APA Handbook of Personality and Social Psychology: Attitudes and Social Cognition*, eds E. Borgida and J. A. Bargh (Washington, DC: APA), 203–229.
- Scott, R. B., and Dienes, Z. (2010). Fluency does not express implicit knowledge of artificial grammars. *Cognition* 114, 372–388. doi: 10.1016/j.cognition.2009.10.010
- Shackman, A. J., Salomons, T. V., Slagter, H. A., Fox, A. S., Winter, J. J., and Davidson, R. J. (2011). The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nat. Rev. Neurosci.* 12, 154–167. doi: 10.1038/nrn2994
- Shams, L., Kamitani, Y., and Shimojo, S. (2000). Illusions. What you see is what you hear. *Nature* 408, 788. doi: 10.1038/35048669
- Spence, C. (2011). Crossmodal correspondences: a tutorial review. *Atten. Percept. Psychophys.* 73, 971–995. doi: 10.3758/s13414-010-0073-7
- Stepper, S., and Strack, F. (1993). Proprioceptive determinants of emotional and nonemotional feelings. *J. Pers. Soc. Psychol.* 64, 211–220. doi: 10.1037/0022-3514.64.2.211
- Topolinski, S., and Strack, F. (2009a). Scanning the “fringe” of consciousness: what is felt and what is not felt in intuitions about semantic coherence. *Conscious. Cogn.* 18, 608–618. doi: 10.1016/j.concog.2008.06.002
- Topolinski, S., and Strack, F. (2009b). The analysis of intuition: processing fluency and affect in judgements of semantic coherence. *Cogn. Emot.* 23, 1465–1503. doi: 10.1080/02699930802420745
- Whittlesea, B. W. A., Jacoby, L. L., and Girard, K. (1990). Illusions of immediate memory: Evidence of an attributional basis for feelings of familiarity and perceptual quality. *J. Mem. Lang.* 29, 716–732. doi: 10.1016/0749-596X(90)90045-2
- Whittlesea, B. W. A. (2002). “On the construction of behavior and subjective experience: the production and evaluation of performance,” in *Rethinking Implicit Memory*, eds J. Bowers and C. Marsolek (Oxford: Oxford University Press), 239–260.
- Winkielman, P., Huber, D. E., Kavanagh, L., and Schwarz, N. (2012). “Fluency of consistency: when thoughts fit nicely and flow smoothly” in *Cognitive consistency: A Fundamental Principle in Social Cognition*, eds B. Gawronski and F. Strack (New York, NY: Guilford Press), 89–111
- Winkielman, P., and Schooler, J. W. (2011). Splitting consciousness: unconscious, conscious, and metaconscious processes in social cognition. *Eur. Rev. Soc. Psychol.* 22, 1–35. doi: 10.1080/10463283.2011.576580
- Winkielman, P., Schwarz, N., Fazendeiro, T., and Reber, R. (2003). “The hedonic marking of processing fluency: implications for evaluative judgment,” in *The Psychology of Evaluation: Affective Processes in Cognition and Emotion*, eds J. Musch and K. C. Klauer (Mahwah, NJ: Lawrence Erlbaum), 189–217.
- Wurtz, P., Reber, R., and Zimmermann, T. D. (2008). The feeling of fluent perception: a single experience from multiple asynchronous sources. *Conscious. Cogn.* 17, 171–184. doi: 10.1016/j.concog.2007.07.001
- Ziembowicz, M., Nowak, A., and Winkielman, P. (2013). When sounds look right and images sound correct: cross-modal coherence enhances claims of pattern presence. *Cognition* 127, 273–278. doi: 10.1016/j.cognition.2013.07.010

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 October 2014; accepted: 15 January 2015; published online: 09 February 2015.

Citation: Winkielman P, Ziembowicz M and Nowak A (2015) The coherent and fluent mind: how unified consciousness is constructed from cross-modal inputs via integrated processing experiences. *Front. Psychol.* 6:83. doi: 10.3389/fpsyg.2015.00083

This article was submitted to *Consciousness Research*, a section of the journal *Frontiers in Psychology*.

Copyright © 2015 Winkielman, Ziembowicz and Nowak. 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.

