THE REASONING BRAIN: THE INTERPLAY BETWEEN COGNITIVE NEUROSCIENCE AND THEORIES OF REASONING

EDITED BY: Vinod Goel, Gorka Navarrete, Ira A. Noveck and Jérôme Prado

PUBLISHED IN: Frontiers in Human Neuroscience

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88945-118-0
DOI 10.3389/978-2-88945-118-0

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

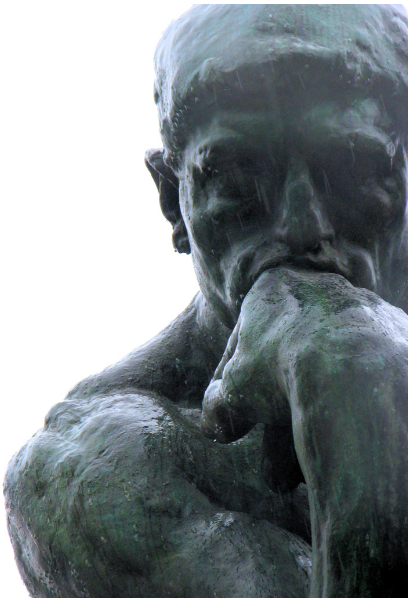
Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public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

THE REASONING BRAIN: THE INTERPLAY BETWEEN COGNITIVE NEUROSCIENCE AND THEORIES OF REASONING

Topic Editors:
Vinod Goel, York University, Canada
Gorka Navarrete, Universidad Adolfo Ibañez, Chile
Ira A. Noveck, Centre National de la Recherche Scientifique and Université de Lyon, France
Jérôme Prado, Centre National de la Recherche Scientifique and Université de Lyon, France



Cover portrait by innoxiuss (available at https://www.flickr.com/photos/46922409@N00/308920352); CC-BY-2.0

Despite the centrality of rationality to our identity as a species (let alone the scientific endeavour), and the fact that it has been studied for several millennia, the present state of our knowledge of the mechanisms underlying logical reasoning remains highly fragmented. For example, a recent review concluded that none of the extant (12!) theories provide an adequate account (Khemlani & Johnson- Laird, 2011), while other authors argue that we are on the brink of a paradigm change, where the old binary logic framework will be washed away and replaced by more modern (and correct) probabilistic and Bayesian approaches (see for example Elqayam & Over, 2012; Oaksford & Chater, 2009; Over, 2009).

Over the past 15 years neuroscience brain imaging techniques and patient studies have been used to map out the functional neuroanatomy of reasoning processes. The aim of this research topic is to discuss whether this line of research has facilitated, hindered, or has been largely irrelevant for understanding of reasoning processes. The answer is neither obvious nor uncontroversial. We would like to engage both the cognitive and the neuroscience community in this discussion. Some of the questions of interest are:

How have the data generated by the patient and neuroimaging studies:

- influenced our thinking about modularity of deductive reasoning
- impacted the debate between mental logic theory, mental model theory and the dual mechanism accounts
- affected our thinking about dual mechanism theories
- informed discussion of the relationship between induction and deduction
- illuminated the relationship between language, visual spatial processing and reasoning
- affected our thinking about the unity of deductive reasoning processes

Have any of the cognitive theories of reasoning helped us explain deficits in certain patient populations?

Do certain theories do a better job of this than others?

Is there any value to localizing cognitive processes and identifying dissociations (for reasoning and other cognitive processes)?

What challenges have neuroimaging data raised for cognitive theories of reasoning?

How can cognitive theory inform interpretation of patient data or neuroimaging data?

How can patient data or neuroimaging data best inform cognitive theory?

This list of questions is not exhaustive. Manuscripts addressing other related questions are welcome. We are interested in hearing from skeptics, agnostics and believers, and welcome original research contributions as well as reviews, methods, hypothesis & theory papers that contribute to the discussion of the current state of our knowledge of how neuroscience is (or is not) helping us to deepen our understanding of the mechanisms underlying logical reasoning processes.

References

Elqayam, S., & Over, D. E. (2012). Probabilities, beliefs, and dual processing: the paradigm shift in the psychology of reasoning. Mind & Society, 11(1), 27–40. doi:10.1007/s11299-012-0102-4

Khemlani, S. S., & Johnson-Laird, P. N. (2011). Theories of the syllogism: A meta-analysis, (571).

Oaksford, M., & Chater, N. (2009). Précis of bayesian rationality: The probabilistic approach to human reasoning. The Behavioral and brain sciences, 32(1), 69–84; discussion 85–120. doi:10.1017/S0140525X09000284

Over, D. E. (2009). New paradigm psychology of reasoning. Thinking & Reasoning, 15(4), 431–438. doi:10.1080/13546780903266188

Citation: Goel, V., Navarrete, G., Noveck, I. A., Prado, J., eds. (2017). The Reasoning Brain: The Interplay between Cognitive Neuroscience and Theories of Reasoning. Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-118-0

Table of Contents

Original Research Articles

Review & Methodological Articles

frontiers in Human Neuroscience

EDITORIAL
published: 05 January 2017
doi: 10.3389/fnhum.2016.00673

Editorial: The Reasoning Brain: The Interplay between Cognitive Neuroscience and Theories of Reasoning

Vinod Goel[1], Gorka Navarrete[2], Ira A. Noveck[3] and Jérôme Prado[3]*

[1] Psychology Department, York University, Toronto, ON, Canada, [2] Center for Social and Cognitive Neuroscience, School of Psychology, Universidad Adolfo Ibáñez, Santiago de Chile, Chile, [3] Institut des Sciences Cognitives Marc Jeannerod, Centre National de la Recherche Scientifique and Université de Lyon, Bron, France

Keywords: logic, rationality, inference, neuroimaging, deduction, induction, emotions, brain

Editorial on the Research Topic

The Reasoning Brain: The Interplay between Cognitive Neuroscience and Theories of Reasoning

The ability to reach logical conclusions on the basis of prior information is central to human cognition. Yet, it is generally agreed that the state of our knowledge regarding the mechanisms underlying logical reasoning remains incomplete and highly fragmented (e.g., Khemlani and Johnson-Laird, 2012). The emergence of functional neuroimaging over the past 20 years—and its ability to examine reasoning at the level of recruitment of cortical systems—provides an additional source of data to, not only better understand reasoning as a phenomenon, but to test different theoretical approaches. This has the potential to both prune the number of theoretical explanations of reasoning, but also to expand the space of possibilities in directions unanticipated by behavioral data. This Research Topic explores the extent to which neuroimaging and brain-lesion studies have informed cognitive theories of reasoning. It includes a selection of 20 empirical and theoretical papers from 69 authors. Below we briefly review these papers by breaking them down into two types of contribution, (i) original research articles, and (ii) review and methodological articles.

OPEN ACCESS

Edited and reviewed by:
Hauke R. Heekeren,
Freie Universität Berlin, Germany

****Correspondence:***
Vinod Goel
vgoel@yorku.ca

Received: *11 May 2016*
Accepted: *16 December 2016*
Published: *05 January 2017*

Citation:
Goel V, Navarrete G, Noveck IA and Prado J (2017) Editorial: The Reasoning Brain: The Interplay between Cognitive Neuroscience and Theories of Reasoning.
Front. Hum. Neurosci. 10:673.
doi: 10.3389/fnhum.2016.00673

ORIGINAL RESEARCH ARTICLES

Most contributions are original research articles that further our understanding of the reasoning brain in several important ways. Perhaps the main finding from these studies is that reasoning relies on a heterogeneous cerebral network that is task-dependent, as can be seen from functional neuroimaging, brain-lesion, and behavioral studies. For example, Liang et al. use neuroimaging data to show that different neural systems contribute to semantic bias and conflict detection in the inclusion fallacy task. Smith et al. and Smith et al. further demonstrate that the neural bases of logical syllogisms can be modulated by the emotional context of the task. Pamplona et al. also provide evidence that general intelligence modulates connectivity between brain regions underlying reasoning. Using a behavioral approach, Andrews et al. show that a frontal-based domain-general capacity for relational processing is particularly important for tasks that require planning, whereas Vendetti et al. find hemispheric differences in the encoding of ordered vs. out-of-order premises in relational reasoning tasks. Finally, Ye et al. demonstrate a causal relationship between activity in the temporo-parietal cortex and tasks relying on mental state attribution for moral judgment.

The fact that the brain network for reasoning is heterogeneous, however, does not imply that some regions are not more important than others for reasoning. This is notably the case for the Inferior Parietal Lobule (IPL), which is related to several different aspects of reasoning in perspective taking tasks (Arora et al.), and is consistently found activated in reasoning tasks (Wendelken). The importance of the IPL is also illustrated by Hinton et al. who show that enhanced activity in the parietal cortex may be critical for compensating reasoning deficits in sub-clinically depressed participants.

REVIEW AND METHODOLOGICAL ARTICLES

Other contributions to the Research Topic are reviews and opinions that speculate on the link between cognitive neuroscience research and theories of reasoning. For example, Oaksford reviews some of the brain imaging research on deductive reasoning and argues that this literature could benefit from adopting the probabilistic and dual-system frameworks of reasoning. Oaksford is notably challenged by Bonatti et al. who argue that neuroscience research has made clear progress within these last 15 years, and does not have much to gain from adopting such frameworks. Other important theoretical contributions are those of Khemlani et al. who illustrate how cognitive neuroscience research can inspire a novel computational theory of how individuals segment perceptual information into representations of events. In a similar vein, Houdé and Borst show how cognitive neuroscience can be used to test an inhibitory-control theory of the reasoning brain, which stresses the importance of inhibiting misleading heuristics when activating logical algorithms.

Six contributions are more methodologically driven and argue for changes in the way cognitive neuroscience research on reasoning is done. Papo argues that the study of reasoning in the brain must rely on the development of a new set of non-standard brain metrics, experimental designs, and analytical tools. Roser et al. propose that a useful way to advance investigations of the reasoning brain would be to integrate several neuroscience methods within a single study. Heit argues that a greater use of "forward inference" in interpreting cognitive neuroscience data may settle disputes between competing cognitive theories. Rotello and Heit caution how misinterpretation of behavioral data could lead to the wrong conclusions at the neuropsychological level. Cummins emphasizes the importance of taking into account how knowledge is activated and weighted in decision processes in the modeling of human causal inference. Finally, Beatty and Vartanian point out that cognitive research on reasoning might also have practical implications. For example, the fact that reasoning is intrinsically linked to working-memory suggests that working memory training could lead to important improvements in reasoning.

Have neuroimaging and brain-lesion studies enhanced our understanding of human reasoning? The main contribution of the augmentation of behavioral data with neuropsychological data has been to question unitary accounts and advocate for the engagement of multiple cognitive systems in reasoning. That is, rather than simply pruning the space of possibilities provided by mental models, mental logic, dual mechanism, and probabilistic account theories, the effect of the neuropsychological data has been to expand the search space in ways not foreseen by behavioral data. This does not make the contribution any less valuable. It identifies challenges, issues, and directions for future research. We hope that readers find this Research Topic informative, thought provoking, and helpful in moving forward the understanding of the cognitive and neural basis logical reasoning.

AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

This work was supported in parts by grants from Wellcome Trust (Grant # ABH00 FA032YBH064) and NSERC to VG, Comisión Nacional de Investigación Científica y Tecnológica (CONICYT/FONDECYT Regular 1150824) to GN, and Fondation de France (2012–00033701), European Union (Marie-Curie Carreer Integration Grant n° PCIG12-GA-2012-333602), and Agence Nationale de la Recherche (ANR-14-CE30-0002-01) to JP. IN was supported by the ESF Euro-XPrag network.

REFERENCES

Khemlani, S. S., and Johnson-Laird, P. N. (2012). Theories of the syllogism: a meta-analysis. *Psychol. Bull.* 138, 427–57. doi: 10.1037/a0026841

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.

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 20 October 2014
doi: 10.3389/fnhum.2014.00797

Different neural systems contribute to semantic bias and conflict detection in the inclusion fallacy task

Peipeng Liang*[1,2]*, Vinod Goel*[3,4]**, *Xiuqin Jia*[1,2] *and Kuncheng Li*[1,2]

[1] *Xuanwu Hospital, Capital Medical University, Beijing, China*
[2] *Brain Key Laboratory of Magnetic Resonance Imaging and Brain Informatics, Beijing, China*
[3] *Department of Psychology, York University, Toronto, ON, Canada*
[4] *IRCCS Fondazione Ospedale San Camillo, Venice, Italy*

Edited by:
Jérôme Prado, Centre National de la Recherche Scientifique, France

Reviewed by:
Carlo Reverberi, Università Milano Bicocca, Italy
Soohyun Cho, Chung-Ang University, South Korea
Hugo Mercier, Université de Neuchâtel, Switzerland

***Correspondence:**
Vinod Goel, Department of Psychology, York University, Toronto, ON M3J 1P3, Canada
e-mail: vgoel@yorku.ca;
Kuncheng Li, Xuanwu Hospital, Capital Medical University, 45 Chang Chun Street, Xi Cheng District, Beijing 100053, China
e-mail: lkc1955@gmail.com

The inclusion fallacy is a phenomenon in which generalization from a specific premise category to a more general conclusion category is considered stronger than a generalization to a specific conclusion category nested within the more general set. Such inferences violate rational norms and are part of the reasoning fallacy literature that provides interesting tasks to explore cognitive and neural basis of reasoning. To explore the functional neuroanatomy of the inclusion fallacy, we used a 2 × 2 factorial design, with factors for quantification (explicit and implicit) and response (fallacious and non-fallacious). It was found that a left fronto-temporal system, along with a superior medial frontal system, was specifically activated in response to fallacious responses consistent with a semantic biasing of judgment explanation. A right fronto-parietal system was specifically recruited in response to detecting conflict associated with the heightened fallacy condition. These results are largely consistent with previous studies of reasoning fallacy and support a multiple systems model of reasoning.

Keywords: fMRI, inductive reasoning, prefrontal cortex, inclusion fallacy, category-based induction

INTRODUCTION

As rational beings, we look to reasons to motivate and justify our actions. However, a long series of cognitive studies suggest that we make systematic errors while reasoning. Perhaps the most pervasive errors have to do with the impact of our belief structures on logical reasoning (Wilkins, 1928; Evans et al., 1983). Several imaging studies have examined the neural basis of belief bias (i.e., the inclination to agree or disagree with an argument based upon whether we find the conclusion believable or unbelievable) in syllogistic reasoning (Goel et al., 2000; Goel and Dolan, 2003). The basic finding is that the left frontal–temporal system is recruited for logical reasoning in the presence of semantic content about which subjects have beliefs, and a right frontal and bilateral parietal system is engaged where such beliefs are absent (Goel et al., 2000) or need to be overcome to generate the logical response (Goel and Dolan, 2003). Where the beliefs are not overcome, a ventral medial frontal system is engaged (Goel and Dolan, 2003). The goal of the current study is to see if these mechanisms generalize to more informal reasoning domains, such as category-based induction.

Category-based induction is a reasoning process by which we project knowledge about certain classes of entities to other related classes of entities (e.g., inferring that ostriches have gene X from the fact that robins have gene X). Inductive generalization from the known to the unknown enables us to benefit from past instances and enlarge the scope of our knowledge. There is a phenomenon within this domain, known as the inclusion fallacy. The inclusion fallacy is a phenomenon in which generalization from a specific category to a more general category (e.g., from robin to bird) is considered to be stronger or more convincing than generalization to a more specific category (e.g., to ostrich) nested within the more general set. Consider the following examples from Osherson et al. (1990):

Robins secrete uric acid crystals
____
Birds secrete uric acid crystals

and

Robins secrete uric acid crystals
____
Ostriches secrete uric acid crystals.

Subjects are presented with pairs of arguments, such as these, and required to make a direct comparison of their relative strength. Many (but not all) people sometimes (but not always) fallaciously choose the first argument as stronger than the second, and thus, commit the inclusion fallacy (since the conclusion of the second argument is contained in the conclusion of the first, it can not be stronger).

The individual arguments are inductive and have no logically correct response. However, as typically administered (Osherson et al., 1990), the task forces subjects to make a direct comparison of the relative strength of the two arguments. There is a logically correct response to this critical component of the task. It is to say that the generalization to all birds cannot be stronger than the generalization to a specific bird (and vice versa). This response is, however, excluded by the task setup. Subjects must choose one or the other as being "stronger," there being no option to say "same strength"[1]. None the less, it seems to defy rational plausibility norms to assert a property to all birds but not a specific bird.

The inclusion fallacy seems to reflect the perceived relationship between the subjects in the premise and conclusion. The link between robin and bird is quite strong because robin is considered to be a typical/central member of the bird category. But an ostrich, despite being a bird, is an atypical/peripheral member of the bird category and is somewhat removed from the representation of robin. In this sense, the phenomenon of inclusion fallacy is similar to the conjunction fallacy in the Linda problem[2] (Tversky and Kahneman, 1983) and the belief-bias effect in deductive reasoning (Evans et al., 1983; Goel and Dolan, 2003; Evans and Curtis-Holmes, 2005; De Neys, 2006a,b), in that the fallacious response is biased by the organization of our world knowledge. However, participants will sometimes overlook the more constrained/logical response and answer on the basis of their knowledge about birds, robins, and ostriches. The inference is biased toward the more familiar/easily accessible category (bird over ostrich).

Not all participants are susceptible to the inclusion fallacy, and those that are do not fall prey to it on all occasions. One factor that may affect participants' susceptibility to the fallacy is the quantifier associated with the conclusion. In the stimuli used by Osherson et al. (1990), e.g., "birds secrete uric acid crystals," the quantifier is only implied, leaving room for ambiguity. If one assumes a strict universal quantifier (e.g., "all birds secrete uric acid crystals") then one should be more aware of the fact that the superordinate category (i.e., bird) subsumes the subordinate category (i.e., ostrich), which should in turn reduce the inclusion fallacy. However, if one does not assume strict universal quantification, then one may be less likely to subsume the subordinate category in the superordinate category. For example, the participant may reason that perhaps the sentence means "most birds or virtually all birds. And after all, ostriches are not real birds." Under such an interpretation one is more likely to make the inclusion fallacy. Thus, the absence of an explicit "all" should increase uncertainty and the inclusion fallacy while the presence of an explicit "all" should decrease uncertainty and the fallacy response. That the presence of an explicit or implicit quantifier should modulate the inclusion fallacy is consistent with the psychological literature on the interpretation of quantifiers (Collins and Quillian, 1969; Newstead and Griggs, 1984). It is also consistent with a related study (Sloman, 1998) that shows that fallacious inferences (specifically, the inclusion similarity)[3] can be modulated by making the category of inclusion relations explicit.

To understand the neural basis of the inclusion fallacy, and its modulation by explicit and implicit quantifiers, we undertook an fMRI study of healthy volunteers while they engaged in generalization inferences on material similar to Osherson et al. (1990). At the behavioral level, we anticipated that a subset of the participants would display the inclusion fallacy and that the fallacy would be displayed much more frequently in the implicit quantifier condition than the explicit quantifier condition. At the neural level, we were interested in the mechanisms underlying responses biased by beliefs and knowledge structures (i.e., the fallacious responses) versus responses in which these beliefs and knowledge structures were bypassed/suppressed to generate non-fallacious responses. We expected these systems to be modulated by the explicit/implicit quantifier condition. Based on the fact that fallacious responses are driven by the organization of our beliefs, we predicted involvement of a left hemisphere frontal–temporal system, including left middle/inferior frontal gyrus and middle temporal gyrus in this condition as seen in several previous reasoning studies (Goel et al., 2000; Goel and Dolan, 2004). Reasoning trials uninfluenced by beliefs (i.e., the non-fallacious responses in the present study), on the other hand, should activate a parietal system, often found in reasoning trials devoid of beliefs (Goel et al., 2000; Waechter et al., 2012). The task paradigm contains a tension/conflict between the fallacious and non-fallacious responses. This is exasperated in the implicit quantifier condition where the uncertain scope of the quantifier leaves room for doubt (see Discussion). In this situation, we predicted activation in right frontal PFC in response to conflict detection, particularly in the case of non-fallacy responses (Goel et al., 2000; Goel and Dolan, 2003; De Neys et al., 2008; Stollstorff et al., 2011).

MATERIALS AND METHODS

SUBJECTS

Sixty-two paid healthy undergraduate and postgraduate students participated in the experiment. All subjects were right-handed and had normal or corrected-to-normal vision. None of the subjects reported any history of neurological or psychiatric diseases. The study was approved by the Ethics Committee of Xuanwu Hospital, Capital Medical University. All participants gave written informed consent.

STIMULI AND DESIGN

One hundred twenty trials, modeled on the Osherson et al. (1990) stimuli, were included in the current study. Each trial was composed of pairs of arguments, one appearing above the other (see **Table 1**). The ordering of the arguments was counterbalanced. The subjects were instructed to judge, and indicate, which one of the two arguments was stronger.

The stimuli were divided into two conditions (see **Table 1**), explicit quantification (60), and implicit quantification (60). Subjects' responses to each trial were used to further divide the stimuli into fallacy or non-fallacy response trials. A fallacious response would be one where the participant chose the argument "robins secrete uric acid crystals, therefore, birds secrete uric acid crystals"

[1] It remains an open question whether the fallacious response would persist if a "same strength" option was made available to participants.

[2] Like the inclusion fallacy, the conjunction fallacy requires a contrivance whereby the one piece of information that appears individually and in the conjunct (i.e., Linda is a bank teller) is not in keeping with the description of Linda, whereas the other half of the conjunct is.

[3] Inclusion similarity is the phenomenon whereby the first argument below is considered stronger (or more convincing) than the second argument: (A) all animals use norepinephrine as a neurotransmitter. Therefore, all mammals use norepinephrine as a neurotransmitter. (B) All animals use norepinephrine as a neurotransmitter. Therefore, all reptiles use norepinephrine as a neurotransmitter. The rationale is that the class of mammals is considered to be more representative or similar to the class of animals than is the class of reptiles.

Table 1 | Example of experimental tasks.

	Explicit	Implicit
Argument 1 (typical to atypical)	All robins secrete uric acid crystals All ostriches secrete uric acid crystals	Robins secrete uric acid crystals Ostriches secrete uric acid crystals
Argument 2 (typical to general)	All robins secrete uric acid crystals All birds secrete uric acid crystals	Robins secrete uric acid crystals Birds secrete uric acid crystals

as being stronger or more convincing than "robins secrete uric acid crystals, therefore, ostriches secrete uric acid crystals." The reverse selection (i.e., where the latter is stronger or more convincing than the former) would be the non-fallacious correct selection. This yielded a 2 × 2 factorial design, with factors for quantification (explicit and implicit) and response (fallacious or non-fallacious), resulting in the following four cells: implicit fallacy (I_F), implicit non-fallacy (I_NF), explicit fallacy (E_F), and explicit non-fallacy (E_NF).

STIMULI PRESENTATION

Stimuli from all conditions were organized into two sessions and presented randomly in an event related design. The order of sessions was counterbalanced among subjects. Trials began with the presentation of one of the arguments (premise plus conclusion). Two seconds later, the second argument (premise plus conclusion) was presented and subjects were given 8 s to respond. Half of the participants used a left button press to indicate that the first argument was stronger and the right button press to indicate that the second argument was stronger. The other half of the participants used the reverse. The two arguments remained on the screen until the end of the trial or the subjects' button-press response. Subjects were instructed to respond as accurately and quickly as possible and move to the next trial if the stimuli advanced before they could respond. The length of trials varied from 9 to 11 s (with a TR/2 jitter), i.e., the length of the trials may be 9, 10, or 11 s with the same probability, randomly. This was determined by pilot data indicating that the range of the inter-trial interval was 7–9 s, with a reaction time of around 3 s. There were 60 event presentations during a session and each session lasted 10 min.

MRI DATA ACQUISITION

Scanning was performed on a 3.0-T MRI system (Siemens Trio Tim; Siemens Medical System, Erlanger, Germany) and with a 12-channel phased array head coil. Foam padding and headphones were used to limit head motion and reduce scanning noise. High-resolution structural images were acquired using a T1 weighted 3D MPRAGE sequence (TR/TE = 1600/2.25 ms, TI = 800 ms, 192 sagittal slices, FOV = 256 mm, 9° flip angle, voxel size = 1 mm × 1 mm × 1 mm). Functional images were obtained using a T2* gradient-echo EPI sequence (TR/TE = 2000/31 ms, 90° flip angle, 64 × 64 matrix size in 240 mm × 240 mm FOV). Thirty axial slices with a thickness of 4 mm and an interslice gap of 0.8 mm were acquired and paralleled to the AC–PC line. The scanner was synchronized with the presentation of every trial.

DATA PREPROCESSING

Data were analyzed using SPM5 software[4]. The first four images for each session were discarded to allow for T1 equilibration effects. The remaining fMRI images were first corrected for within-scan acquisition time differences between slices and then realigned to the first volume to correct for inter-scan head motions (head movements were <1 voxel in all cases). The structural image was co-registered to the mean functional image created from the realigned images using a linear transformation. The transformed structural images were then segmented into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) by using a unified segmentation algorithm (Ashburner and Friston, 2005). The realigned functional volumes were spatially normalized to the Montreal Neurological Institute (MNI) space and re-sampled to 3 mm isotropic voxels using the normalization parameters estimated during unified segmentation. The registration of the functional data to the template was checked for each individual subject. Subsequently, the functional images were spatially smoothed with a Gaussian kernel of 8 mm × 8 mm × 8 mm full width at half maximum (FWHM) to decrease spatial noise.

fMRI ANALYSIS

For all trials, the epoch of interest extends from the presentation of the first argument to the response. The BOLD signal was modeled using canonical HRF with temporal derivative implemented in SPM5. Condition effects at each voxel were estimated according to the general linear model and regionally specific effects were compared using linear contrasts. Each contrast produced a statistical parametric map (SPM) of the t-statistic, which was subsequently transformed to a unit normal Z-distribution. The contrast images were then used in a random effect analysis to determine the regions most consistently activated across subjects. The contrasts of primary interest in the present study are the main effect of fallacy (F–NF, NF–F), explicitness (I–E and E–I), and the interaction effects [(I_F–I_NF)–(E_F–E_NF) and (E_F–E_NF)–(I_F–I_NF)]. The activations reported survived a voxel-level threshold of $p < 0.001$ and a cluster size comprised of a minimum of eight contiguous voxels, which corresponded to a corrected $p < 0.05$ using the AlphaSim program[5] (parameters: FWHMx = 12.23 mm, FWHMy = 10.39 mm, FWHMz = 9.67 mm, within the GM mask). The real smoothness in the three directions was estimated by using 3dFWHMx.

RESULTS

BEHAVIORAL PERFORMANCE

Of the 62 subjects, 58 exhibited the fallacy at least once in the implicit condition and 54 exhibited the fallacy at least once in the explicit condition. To ensure adequate signal-to-noise ratio, and to allow for within subject analyses, we used a cut off of at least 12 trials in the fallacy and logical response conditions to select participants for fMRI analyses. Fifteen subjects (7 females) with a

[4] http://www.fil.ion.ucl.ac.uk

[5] http://afni.nimh.nih.gov/pub/dist/doc/manual/AlphaSim.pdf

mean age of 23.6 ± 3.1 years met this criterion and were included in the subsequent fMRI data analysis. The initial behavioral analysis, below, includes all 62 participants. The subsequent analysis is limited to 15 participants used in the fMRI analysis. The pattern of results in the two cases is identical.

Behavioral scores were in keeping with expectations (see **Figure 1**). In terms of responses from all 62 participants, we found a main effect of response [$F(1,61) = 3.81$, $p = 0.05$], such that the number of non-fallacious responses were greater than the number of fallacious responses. There was also a quantification (explicit, implicit) by response (fallacy, non-fallacy) interaction [$F(1,61) = 23.97$, $p = 0.00$] (see **Figure 1A**), driven by the fact that there were more non-fallacious responses than fallacious responses in the explicit quantifier trials [$F(1,61) = 15.54$, $p = 0.00$], but there was no difference in the number of non-fallacious and fallacious responses in the implicit trials [$F(1,61) = 0.02$, $p = 0.90$].

In terms of reaction times, there was a main effect of response [$F(1,49) = 6.15$, $p = 0.017$], with participants taking longer to respond in trials in which they commit the inclusion fallacy (see **Figure 1A**). The main effect of quantification [$F(1,49) = 0.24$, $p = 0.62$] and the quantification by response interaction [$F(1,49) = 2.68$, $p = 0.11$] were not significant. The *post hoc* analysis of RTs also showed that the RT for fallacy trials was significantly longer than that for non-fallacy response trials in the explicit condition [$F(1,52) = 4.20$, $p = 0.046$] but not in the implicit condition [$F(1,55) = 2.28$, $p = 0.14$]. (Note: as there are NULL values for RT in some conditions for several subjects, the degrees of freedom are not always 61, but variable).

We then analyzed the results of the 15 subjects that will be included in the fMRI analyses (see **Figure 1B**). In terms of accuracy responses, we found a main effect of response [$F(1,14) = 24.47$, $p = 0.00$], such that the number of non-fallacious responses was greater than the number of fallacious responses, and a quantification (explicit, implicit) by response (fallacy, non-fallacy) interaction [$F(1,14) = 11.70$, $p = 0.004$], again driven by the fact that the difference between non-fallacious and fallacious responses was greater in the explicit trials than the implicit trials. In terms of reaction times, the effects were not significant, but the pattern was similar to that of the 62 subjects.

fMRI RESULTS

As noted above, the fMRI results are based on 15 of the 62 participants who had at least 12 trials in each of the 4 conditions.

The main effect of response (**Table 2**), derived from comparisons of trials with fallacious and non-fallacious responses (F–NF), revealed activation of bilateral superior/medial frontal gyrus (BA 8), left inferior frontal gyrus/insula (BA45, 13), and left middle temporal gyrus (BA 21, 22) in the fallacy trials (**Table 2**; **Figure 2**). The reverse comparison, of the main effect of

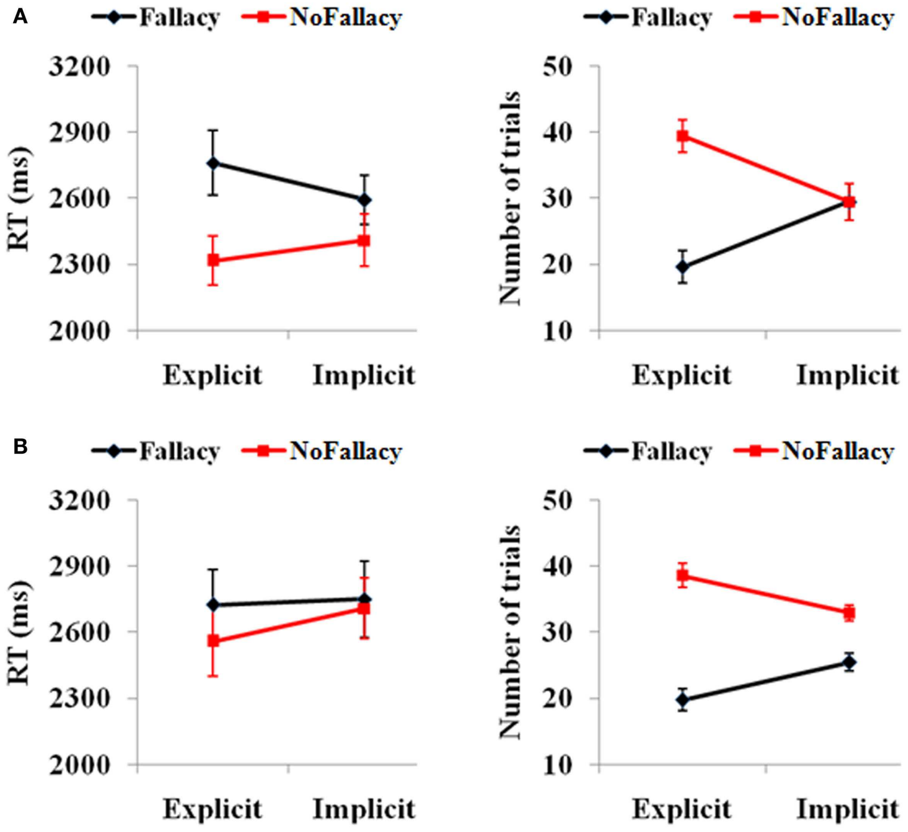


FIGURE 1 | (A) Behavioral performance of 62 subjects and **(B)** the 15 subjects with enough trials for the further fMRI data analysis. The error bars represent the SEM.

Table 2 | Main effect of fallacy and explicitness and the interaction effect of fallacy by explicitness.

Brain regions	MNI coordinate			BA	Cluster size	*T*-score
	x	*y*	*z*			
F–NF						
Medial·Superior frontal gyrus	3	33	48	8	17	5.23
Lt. middle temporal gyrus	−66	−39	−6	21	28	5.21
Lt. middle temporal gyrus	−63	−36	3	22		4.58
Lt. inferior frontal gyrus/insula	−39	15	9	45/13	17	4.81
	−30	24	6	45		4.64
Lt. medial frontal gyrus	−3	36	48	8	12	4.60
NF–F						
No significant activation						
E–I						
No significant activation						
I–E						
Rt. inferior parietal lobule	42	−54	48	40	10	5.02
Rt. superior parietal lobule	36	−57	54	7		3.87
(I_F–I_NF)–(E_F–E_NF)						
Rt. superior parietal lobule	27	−57	45	7	33	4.74
Rt. precuneus	24	−72	51	7		3.37
Lt. fusiform gyrus	−48	−57	−15	37	10	4.22
Rt. middle frontal gyrus	48	33	18	46	10	3.70
(E_F–E_NF)–(I_F–I_NF)						
No significant activation						

response, non-fallacious versus fallacious trials (NF–F), revealed no significant activations.

The main effect of quantification, derived from comparisons of implicit minus explicit trials, revealed activation of right superior/inferior parietal lobule (BA 40, 7) (**Table 2**; **Figure 3**). The reverse comparison, explicit minus implicit quantifiers, revealed no significant activations.

We next examined the interaction between response and quantification. The difference between fallacious and non-fallacious responses in implicit condition trials [(I_F–I_NF)–(E_F–E_NF)], resulted in greater activation in right middle frontal gyrus (BA 46), right superior parietal lobule (BA 7), and left fusiform gyrus (BA 37) than the difference between fallacious and non-fallacious responses in the explicit condition trials (**Table 2**; **Figure 4**). No regions of significant activation were found in the reverse direction [(I_NF–I_F)–(E_NF–E_F)].

Additionally, in order to exclude the potential effect of task difficulty on the activations, we performed another analysis using RT of each trial as covariates. These results are reported in Table S1 in Supplementary Material. It was found that almost all activations survived the supplementary analysis, indicating that the results were not driven by task difficulty differences between trial types.

DISCUSSION

Consistent with previous literature (Osherson et al., 1990; Shafir et al., 1990), our results demonstrate susceptibility to the inclusion fallacy in a subset of participants. Furthermore, we demonstrate that the fallacy is indeed modulated by the explicitness of the quantifier. The presence of an explicit universal quantifier significantly reduces the rate of fallacious responses. This may be because the explicit quantifier eliminates ambiguity regarding the scope of the general category and increases the likelihood that the general category will subsume the more specific category.

Our main aim is to explore the neural basis of this fallacy and its modulation by explicit quantification. Consistent with our first neural prediction we found that committing the fallacy was associated with a predominantly left hemisphere frontal–temporal system, including the left inferior frontal gyrus/insula and middle temporal gyrus. This is a semantic system found to be involved in inductive reasoning and belief-based deductive reasoning (Goel et al., 2000; Goel and Dolan, 2004). The involvement of this system in the fallacious response trials is consistent with the possibility that fallacious responses in this paradigm are driven by a combination of the organization of our knowledge base (i.e., typicality/centrality effects), which sometimes exclude ostriches from the class of birds, and an overweighting of the resulting belief-based response over the more rationally plausible response. The activity in bilateral medial/superior frontal cortex may be associated with attentional orientation response (Hopfinger et al., 2000; Rushworth et al., 2004; Woldorff et al., 2004; Taylor et al., 2008).

Despite our prediction of parietal activation, we did not find significant activation in the reverse condition (non-fallacious responses versus fallacious responses). One possible explanation for the lack of finding in this comparison is that, unlike the syllogistic reasoning paradigm, where the logical response is much more complex and effortful, in the present paradigm the non-fallacious response is trivial, so activations associated with it may have been subsumed by the fallacy condition.

In terms of the quantification factor, the absence of the explicit quantifier significantly increased the number of fallacious responses and decreased the number of non-fallacious responses. The neural correlates of this can be seen in the activation of right inferior and superior parietal lobule in the comparison of implicit versus explicit conditions. The implicit condition introduces some uncertainty into the task by increasing ambiguity. Parameter estimates (**Figure 4**) indicate that this activation is driven by the difference in implicit fallacious versus implicit non-fallacious responses. We consider this activation below, in the discussion of the interaction results.

The explicit minus implicit comparison, on the other hand, revealed no significant activation. As above, it is possible that, given the explicit condition had a preponderance of non-fallacious responses, and that the non-fallacious condition is quite trivial (if the fallacious response is never considered), activations associated with the explicit quantifier condition may be subsumed by activations in the implicit quantifier condition.

Focusing on the response by quantifier interaction highlights the critical role of right lateral prefrontal cortex and parietal lobule system in reasoning. As this is an interaction analysis, and controls for the presence of fallacy and non-fallacy responses, one can interpret the result as being driven by the greater uncertainty in the implicit condition rather than general semantic requirements of the fallacy responses (as in the main effect). (Examination of the parameter estimates clearly indicates that the effect is driven

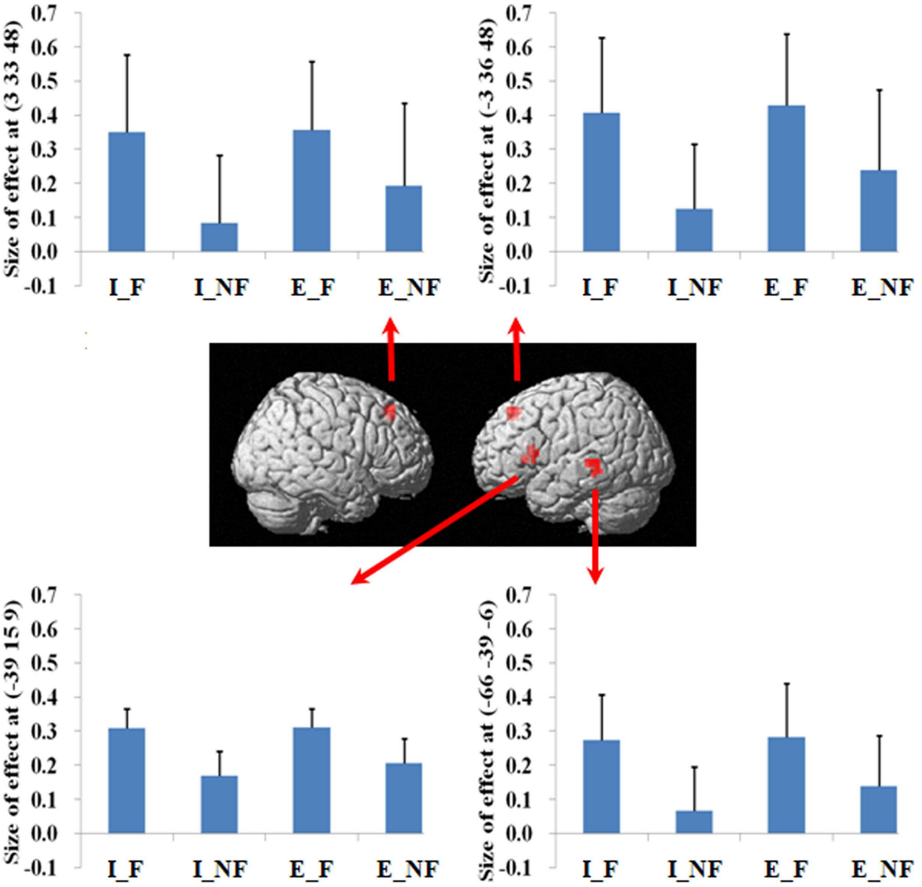


FIGURE 2 | A statistical parametric map (SPM) rendered into standard stereotactic space. A comparison of fallacy trials versus non-fallacy trials (F_NF) results in activation in left inferior frontal gyrus/insula (MNI: −39, 15, 9; $T = 4.81$) (BA 45/13), left middle temporal gyrus (MNI: −66, −39, −6; $T = 5.21$) (BA 21/22), left medial frontal gyrus (MNI: −3, 36, 48; $T = 4.60$) (BA 8), and right superior frontal gyrus (MNI: 3, 33, 48; $T = 5.23$) (BA 8) [also see the main effect of (F–NF) in **Table 2**]. Condition specific parameter (beta) estimates show that the left fronto-temporal system and bilateral mesial frontal gyrus are specifically responding to fallacy trials in both implicit and explicit conditions. The error bars represent the SEM. The activations reported survived an uncorrected voxel-level intensity threshold of $p < 0.001$ with a minimum cluster size of 10 contiguous voxels, which corresponds to a corrected $p < 0.05$ (using the AlphaSim program as described in Section Materials and Methods).

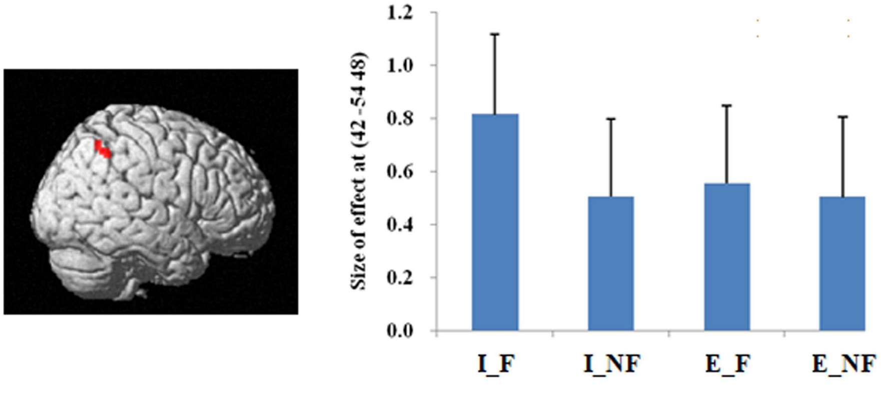


FIGURE 3 | A statistical parametric map (SPM) rendered into standard stereotactic space. A comparison of implicit trials versus explicit trials (I–E) results in activation in right inferior/superior parietal lobule (MNI: 42, −54, 48/36, −57, 54; $T = 5.02/3.87$) (BA 40/7) [also see the main effect of (I–E) in **Table 2**]. Condition specific parameter (beta) estimates show that the right parietal area is responding to fallacy trials in both implicit and explicit conditions, but the main effect in this region is mainly driven by the implicit fallacy trials. The error bars represent the SEM. The activations reported survived an uncorrected voxel-level intensity threshold of $p < 0.001$ with a minimum cluster size of 10 contiguous voxels, which corresponds to a corrected $p < 0.05$ (using the AlphaSim program as described in Section Materials and Methods).

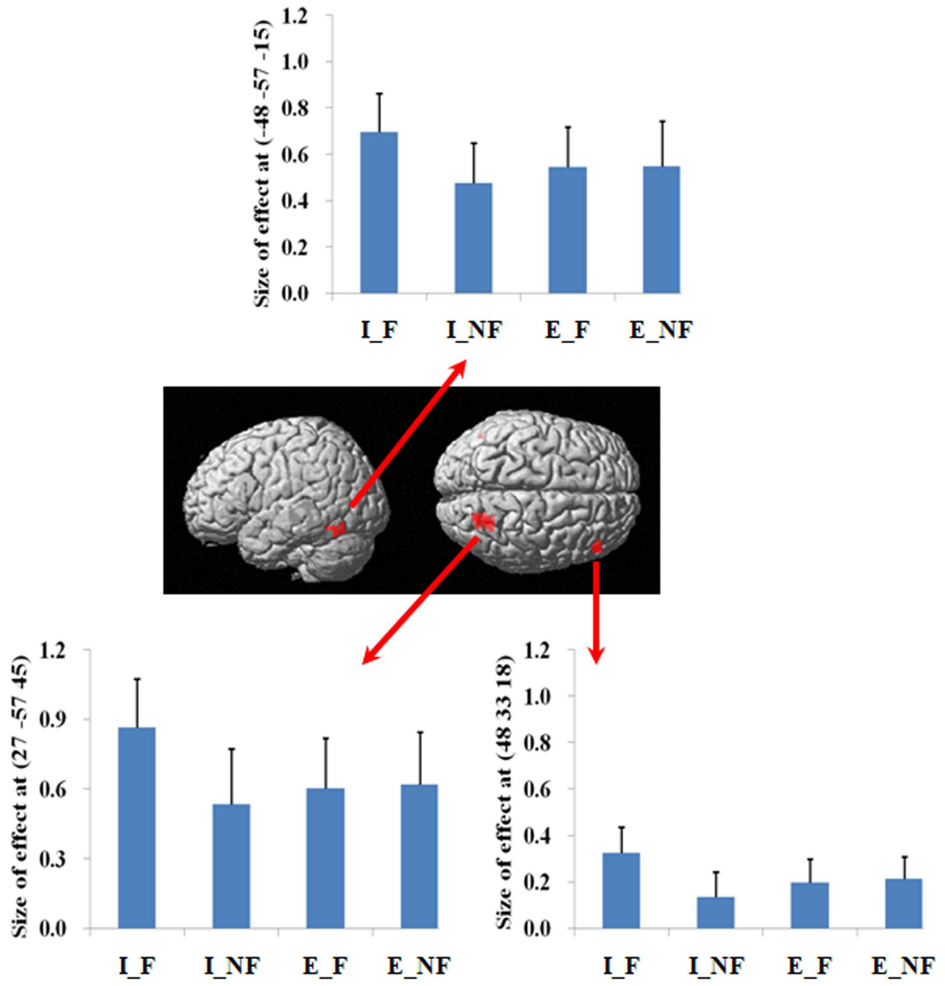


FIGURE 4 | A statistical parametric map (SPM) rendered into standard stereotactic space. The quantification (explicit, implicit) by response (fallacious, non-fallacious) interaction, i.e., a comparison of the difference between implicit fallacy trials versus implicit non-fallacy trials with the difference between explicit fallacy trials versus explicit non-fallacy trials [(I_F–I_NF)–(E_F–E_NF)], results in activation in right middle frontal gyrus (MNI: 48, 33, 18; $T = 3.70$) (BA 46) and superior parietal lobule (MNI: 27, −57, 45; $T = 4.74$) (BA 7) [also see the interaction effect of (I_F–I_NF)–(E_F–E_NF) in **Table 2**]. Condition specific parameter (beta) estimates show that the right fronto-parietal system is specifically responding to fallacies with implicit items, but not to fallacies with explicit items. The error bars represent the SEM. The activations reported survived an uncorrected voxel-level intensity threshold of $p < 0.01$ with a minimum cluster size of 10 contiguous voxels, which corresponds to a corrected $p < 0.05$ (using the AlphaSim program as described in Section Materials and Methods).

by differential response of this system to the fallacious versus non-fallacious responses in the implicit condition. This right hemisphere frontal parietal system shows no differential sensitivity to the explicit condition trials.) When one exhibits the fallacy in the explicit condition (i.e., after being told that *All* birds have X) it may be a function of oversight, or simply believing that the property of the superordinate category does not generalize to this specific subordinate category (e.g., believing that most properties of robins do not generalize to ostriches). However, the implicit condition facilitates the fallacy by introduction of uncertainty and ambiguity. In the absence of an explicit quantifier, one may be less likely to subsume the subordinate category in the superordinate category. For example, the participant may reason that perhaps the sentence means "most birds or virtually all birds. And after all, ostriches are not real birds." Under such an ambiguous interpretation, one is more likely to make the inclusion fallacy.

These results differ in two important respects from our expectations. First, the activation was not specific to the non-fallacious condition (i.e., where the fallacious response is suppressed), as we had predicted. Previous studies have reported right PFC activation in detecting and/or overcoming conflict in reasoning (Goel et al., 2000; Goel and Dolan, 2003; Aron et al., 2004; Prado and Noveck, 2007; De Neys et al., 2008; Stollstorff et al., 2011). However, there is evidence that fallacious responses are accompanied by an awareness of the conflict between the more logical response and the belief cued response, even when the fallacious response is not suppressed (De Neys, 2006a,b). The present results suggest that detection of conflict may be sufficient to activate this system. Second, while several previous studies report right PFC activation for conflict detection, Goel and Dolan (2003) also noted accompanying activation in parietal cortex, even though it did not survive correction. The present results suggest a role of the parietal system in conflict detection. Finally, the recruitment of the left fusiform gyrus is consistent with semantic processing and retrieval (Thompson-Schill et al., 1999; Devlin et al., 2006; Mion et al., 2010).

In summary, our results show that a left fronto-temporal system, along with bilateral medial superior frontal system, is specifically activated in the main effect of fallacy in response to biasing of reasoning judgment by the semantic organization of knowledge. A right fronto-parietal system, along with left fusiform gyrus, is specifically recruited in the absence of explicit quantifiers, where fallacious responses increase, as a function of increased uncertainty and ambiguity. These activations may reflect an awareness of the conflict between the selected response and logical response. More generally, these results reinforce the involvement of multiple systems in logical reasoning.

ACKNOWLEDGMENTS

This work was supported by the Natural Science Foundation of China (61473196, 61105118 to Peipeng Liang); Beijing Nova Program (Z12111000250000 to Peipeng Liang); and Open Research Fund of the State Key Laboratory of Cognitive Neuroscience and Learning (CNLZD1302 to Peipeng Liang).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at http://www.frontiersin.org/Journal/10.3389/fnhum.2014.00797/abstract

REFERENCES

Aron, A. R., Robbins, T. W., and Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends Cogn. Sci.* 8, 170–177. doi:10.1016/j.tics.2004.02.010

Ashburner, J., and Friston, K. J. (2005). Unified segmentation. *Neuroimage* 26, 839–851. doi:10.1016/j.neuroimage.2005.02.018

Collins, A., and Quillian, M. (1969). Retrieval time from semantic memory. *J. Verbal Learn. Verbal Behav.* 8, 240–247. doi:10.1016/S0022-5371(69)80069-1

De Neys, W. (2006a). Dual processing in reasoning: two systems but one reasoner. *Psychol. Sci.* 17, 428–433. doi:10.1111/j.1467-9280.2006.01723.x

De Neys, W. (2006b). Automatic-heuristic and executive-analytic processing during reasoning: chronometric and dual-task considerations. *Q. J. Exp. Psychol.* 59, 1070–1100. doi:10.1080/02724980543000123

De Neys, W., Vartanian, O., and Goel, V. (2008). Smarter than we think: when our brains detect that we are biased. *Psychol. Sci.* 19, 483–489. doi:10.1111/j.1467-9280.2008.02113.x

Devlin, J. T., Jamison, H. L., Gonnerman, L. M., and Matthews, P. M. (2006). The role of the posterior fusiform gyrus in reading. *J. Cogn. Neurosci.* 18, 911–922. doi:10.1162/jocn.2006.18.6.911

Evans, J. S., Barston, J. L., and Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Mem. Cognit.* 11, 295–306. doi:10.3758/BF03196976

Evans, J. S., and Curtis-Holmes, J. (2005). Rapid responding increases belief bias: evidence for the dual-process theory of reasoning. *Think. Reason.* 11, 382–389. doi:10.1080/13546780542000005

Goel, V., Buchel, C., Frith, C., and Dolan, R. (2000). Dissociation of mechanisms underlying syllogistic reasoning. *Neuroimage* 12, 504–514. doi:10.1006/nimg.2000.0636

Goel, V., and Dolan, R. J. (2003). Explaining modulation of reasoning by belief. *Cognition* 87, B11–B22. doi:10.1016/S0010-0277(02)00185-3

Goel, V., and Dolan, R. J. (2004). Differential involvement of left prefrontal cortex in inductive and deductive reasoning. *Cognition* 93, B109–B121. doi:10.1016/j.cognition.2004.03.001

Hopfinger, J. B., Buonocore, M. H., and Mangun, G. R. (2000). The neural mechanisms of top down attentional control. *Nat. Neurosci.* 3, 284–291. doi:10.1038/72999

Mion, M., Patterson, K., Acosta-Cabronero, J., Pengas, G., Izquierdo-Garcia, D., Hong, Y. T., et al. (2010). What the left and right anterior fusiform gyri tell us about semantic memory. *Brain* 133, 3256–3268. doi:10.1093/brain/awq272

Newstead, S. E., and Griggs, R. A. (1984). Fuzzy quantifiers as an explanation of set inclusion performance. *Psychol. Res.* 46, 377–388. doi:10.1007/BF00309070

Osherson, D. N., Smith, E. E., Wilkie, O., López, A., and Shafir, E. (1990). Category-based induction. *Psychol. Rev.* 97, 185–200. doi:10.1037/0033-295X.97.2.185

Prado, J., and Noveck, I. A. (2007). Overcoming perceptual features in logical reasoning: a parametric fMRI study. *J. Cogn. Neurosci.* 19, 642–657. doi:10.1162/jocn.2007.19.4.642

Rushworth, M. F., Walton, M. E., Kennerley, S. W., and Bannerman, D. M. (2004). Action sets and decisions in the medial frontal cortex. *Trends Cogn. Sci.* 8, 410–417. doi:10.1016/j.tics.2004.07.009

Shafir, E. B., Smith, E. E., and Osherson, D. N. (1990). Typicality and reasoning fallacies. *Mem. Cognit.* 18, 229–239. doi:10.3758/BF03213877

Sloman, S. A. (1998). Categorical inference is not a tree: the myth of inheritance hierarchies. *Cogn. Psychol.* 35, 1–33. doi:10.1006/cogp.1997.0672

Stollstorff, M., Vartanian, O., and Goel, V. (2011). Levels of conflict in reasoning modulate right lateral prefrontal cortex. *Brain Res.* 1428, 24–32. doi:10.1016/j.brainres.2011.05.045

Taylor, P. C. J., Rushworth, M. F. S., and Nobre, A. C. (2008). Choosing where to attend and the medial frontal cortex: an fMRI study. *J. Neurophysiol.* 100, 1397–1406. doi:10.1152/jn.90241.2008

Thompson-Schill, S. L., Aguirre, G. K., D'Esposito, M., and Farah, M. J. (1999). A neural basis for category and modality specificity of semantic knowledge. *Neuropsychologia* 37, 671–676. doi:10.1016/S0028-3932(98)00126-2

Tversky, A., and Kahneman, D. (1983). Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. *Psychol. Rev.* 90, 293–315. doi:10.1037/0033-295X.90.4.293

Waechter, R., Goel, V., Raymont, V., Kruger, F., and Grafman, J. (2012). Transitive inference reasoning is impaired by focal lesions in parietal cortex rather than rostrolateral prefrontal cortex. *Neuropsychologia* 51, 464–471. doi:10.1016/j.neuropsychologia.2012.11.026

Wilkins, M. C. (1928). The effect of changed material on the ability to do formal syllogistic reasoning. *Arch. Psychol.* 16, 1–83.

Woldorff, M. G., Hazlett, C. J., Fichtenholtz, H. M., Weissman, D. H., Dale, A. M., and Song, A. W. (2004). Functional parcellation of attentional control regions of the brain. *J. Cogn. Neurosci.* 16, 149–165. doi:10.1162/089892904322755638

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: 31 May 2014; accepted: 18 September 2014; published online: 20 October 2014.

Citation: Liang P, Goel V, Jia X and Li K (2014) Different neural systems contribute to semantic bias and conflict detection in the inclusion fallacy task. Front. Hum. Neurosci. ***8****:797. doi: 10.3389/fnhum.2014.00797*

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

ORIGINAL RESEARCH
published: 12 May 2015
doi: 10.3389/fnhum.2015.00273

Syllogisms delivered in an angry voice lead to improved performance and engagement of a different neural system compared to neutral voice

Kathleen W. Smith[1], *Laura-Lee Balkwill*[2], *Oshin Vartanian*[3] *and Vinod Goel*[1,4]*

[1] *Department of Psychology, Faculty of Health, York University, Toronto, ON, Canada,* [2] *Humanist Canada, Ottawa, ON, Canada,* [3] *Department of Psychology, University of Toronto at Scarborough, Toronto, ON, Canada,* [4] *IRCCS Fondazione Ospedale San Camillo, Venice, Italy*

Despite the fact that most real-world reasoning occurs in some emotional context, very little is known about the underlying behavioral and neural implications of such context. To further understand the role of emotional context in logical reasoning we scanned 15 participants with fMRI while they engaged in logical reasoning about neutral syllogisms presented through the auditory channel in a sad, angry, or neutral tone of voice. Exposure to angry voice led to improved reasoning performance compared to exposure to sad and neutral voice. A likely explanation for this effect is that exposure to expressions of anger increases selective attention toward the relevant features of target stimuli, in this case the reasoning task. Supporting this interpretation, reasoning in the context of angry voice was accompanied by activation in the superior frontal gyrus—a region known to be associated with selective attention. Our findings contribute to a greater understanding of the neural processes that underlie reasoning in an emotional context by demonstrating that two emotional contexts, despite being of the same (negative) valence, have different effects on reasoning.

Keywords: reasoning, emotion, fMRI, anger, sadness, auditory

Edited by:
Srikantan S. Nagarajan,
University of California,
San Francisco, USA

Reviewed by:
Matt Roser,
Plymouth University, UK
Bastien Trémolière,
Université du Québec à Trois-Rivières,
Canada

***Correspondence:**
Vinod Goel,
Department of Psychology, Faculty of Health, York University, 4700 Keele Street, Toronto, ON M3J 1P3, Canada
vgoel@yorku.ca

Received: *06 February 2015*
Accepted: *27 April 2015*
Published: *12 May 2015*

Citation:
Smith KW, Balkwill L-L, Vartanian O and Goel V (2015) Syllogisms delivered in an angry voice lead to improved performance and engagement of a different neural system compared to neutral voice. Front. Hum. Neurosci. 9:273. doi: 10.3389/fnhum.2015.00273

Introduction

It has been demonstrated that whereas reasoning with neutral material was associated with activation in left dorsolateral prefrontal cortex, reasoning with negatively charged (provocative) emotional material was associated with activation in ventromedial prefrontal cortex; furthermore, these neural mechanisms were activated in a reciprocal manner (Goel and Dolan, 2003b). Smith et al. (2014) found that, when emotion was induced by positively or negatively valenced pictorial stimuli prior to the introduction of the reasoning task, reasoning about neutral material led to dissociable neural patterns depending on whether the induction had been positive, negative, or neutral. For example, direct comparison of neural activation in the reasoning time windows in the positive and negative conditions, after controlling for baseline effects, yielded activation in cerebellar vermis and right inferior frontal gyrus (orbitalis) after positive emotion induction but activation in left caudate nucleus and left inferior frontal gyrus (opercularis) after negative emotion induction.

In the current study, we continue our investigation of the effect that emotion has on reasoning. Whereas the previous studies examined the effects of visually presented emotional syllogism content, and visually presented emotional valence (positive and negative), here our interest is to discover whether reasoning and its neural underpinnings will be affected differently by exposure to the expression of two different emotions in the auditory channel.

There is support from various theoretical models in the literature for the existence of different specific emotions, each with its own neural and/or physiological signature (Friedman, 2010); moreover, individuals in therapy can be guided to switch from one specific emotion to another by methods designed to alter their underlying physiology and therefore their current emotional experience (Smith and Greenberg, 2007). Appraisal models likewise consider the differential effects of specific emotions such as dispositional fear and anger on the evaluation of subsequently occurring events (Lerner and Keltner, 2001; DeSteno et al., 2004; Dunn and Schweitzer, 2005).

Our interest in testing the effects of specific emotions (rather than emotional valence) is that we hope to show that reasoning and its neural underpinnings are affected differently by expression of different specific emotions. We chose anger and sadness as the specific emotions because there is literature (to be presented next) suggesting that these emotions are characterized differently.

The neuroimaging literature provides evidence that sadness and anger are characterized differently. A meta-analysis of neuroimaging of emotion (Murphy et al., 2003) reported that whereas anger has been associated with the lateral orbitofrontal cortex, happiness and sadness have been associated with supracallosal anterior cingulate and dorsomedial prefrontal cortex.

Neural activation associated with hearing the voice of an angry speaker (Sander et al., 2005) was noted in bilateral superior temporal sulcus (right BA 42, bilateral BA 22) and right amygdala. Grandjean et al. (2005) demonstrated that superior temporal lobe activation associated with anger prosody is associated with the angry emotion itself, and not with low-level acoustical properties of the stimulus. Other activations found by Sander et al. (2005) include cuneus, left superior frontal gyrus (BA 8), right medial orbitofrontal cortex, left lateral frontal pole (BA 10), right superior temporal sulcus (BA 39), and bilateral ventrolateral prefrontal cortex (BA 47). Ethofer et al. (2009) investigated whether neural activation to angry versus neutral prosody would depend on the relevance of the prosody to the task; tasks were to judge the affective prosody (angry, neutral) or word class (adjective, noun) of semantically neutral spoken words. Neural activation associated with angry versus neutral prosody was reported in bilateral superior temporal gyrus, bilateral inferior frontal/orbitofrontal cortex, bilateral insula, mediodorsal thalamus, and bilateral amygdala, regardless of task, suggesting that these activations occur automatically when processing emotional information in the voice. Neural activation was greater during judgment of emotion than word classification in bilateral inferior frontal/orbitofrontal cortex, right dorsomedial prefrontal cortex, and right posterior middle and superior temporal cortex. Quadflieg et al. (2008) found that neural activation associated with angry versus neutral prosody was noted in fronto-temporal regions, amygdala, insula, and striatum. Identification of the prosody as emotional was additionally associated with activation in orbitofrontal cortex. Individuals with social phobia, compared to healthy controls, demonstrated a larger response in orbitofrontal cortex in response to angry prosody, regardless of whether the task related to the prosody (identify prosody as emotional or neutral) or not (identify the gender of the speaker).

Neural correlates of sadness invoked by re-experiencing of sad autobiographical episodes (Liotti et al., 2000) were reported in the subgenual anterior cingulate (BA 24/25), right posterior insula, and left anterior insula. Relative deactivation was noted in right dorsolateral prefrontal cortex (BA 9), bilateral inferior temporal gyrus (left BA 20, right BA 20/37), right posterior cingulate/retrosplenial cortex, and bilateral parietal lobes.

A second reason for choosing anger and sadness is that these emotions have been posited to have different effects on attention, memory, and categorization (Gable and Harmon-Jones, 2010b) and therefore may have different effects on reasoning.

In theoretical terms, anger is an important emotion because despite its negative valence it is an ‘approach-related’ emotion, and this observation has prompted a reconsideration of theoretical models of emotion (Carver and Harmon-Jones, 2009). Carver and Harmon-Jones (2009) proposed a model incorporating discrete emotions such as joy, anger, calm, and fear into a dimensional model combining approach/withdrawal with system functioning (i.e., events going well or poorly). In this model, anger is classified as an approach emotion activated when system functioning is going poorly.

Following on this, Gable and Harmon-Jones (2010b) proposed a model outlining the consequences for attention, memory, and categorization of emotions classified on the dimensions of approach/withdrawal in relation to an object or goal, coupled with the strength of that motivation. Specifically, disgust and fear may be strong motivators to avoid an object or goal whereas sadness may be a mild motivator to withdraw from an object or goal. Anger, in contrast, may be a strong motivator to approach an object or goal, despite being negative in valence (Carver and Harmon-Jones, 2009). Regarding the consequences of a strong motivator (such as anger) and a weak motivator (such as sadness) on attention, converging evidence (see Gable and Harmon-Jones, 2010b for a review) suggests that strong motivation to either approach or avoid an object or goal is associated with narrowed attention toward that object or goal, and a lack of attention to other stimuli in the environment that are not relevant to that goal. In contrast, weak motivation, which may occur post-goal-attainment, is associated with broadened attention toward more information from the environment beyond the goal itself.

Consistent with the Gable and Harmon-Jones (2010b) model, lab-induced anger and fear have (separately) led to selective attention to targets at the expense of non-target information (Finucane, 2011); so has disgust (Gable and Harmon-Jones, 2010a). Brosch et al. (2008) reported that angry prosody facilitated selective attention to a concurrently presented visual stimulus.

In contrast, sadness has led to a broadening of attention to global rather than local features of stimuli (Gable and Harmon-Jones, 2010a).

As has been noted above, anger is often studied using an auditory paradigm. Accordingly, we decided to use an auditory paradigm in the current study. Auditory paradigms have been used previously to study reasoning in the absence of emotion (Knauff et al., 2002, 2003; Fangmeier and Knauff, 2009).

Finally, we chose to deliver the reasoning material concurrently with the emotive (and neutral) tones of voice, rather than subsequent to the different tones of voice. Our choice was pragmatic: the latter design would have resulted in a longer experiment, and therefore longer scanning time.

Therefore, our study investigated whether reasoning about neutral material would be affected if the content were presented in sad, neutral, or angry tone of voice. To address this issue, we constructed a 3 (Emotion) × 2 (Task) within-subjects design, where the three levels of the Emotion factor were sad, neutral, and angry, and the two levels of the Task factor were reasoning and baseline.

In Smith et al. (2014), the negative and positive valence inductions were each comprised of a mix of emotions, and we found that reasoning tended to be impaired after each valence of emotion. In the current study, our choice of two specific negative emotive tones of voice, anger and sadness, was motivated by the expectation that each of these specific expressions of emotion would lead to different reasoning performance and different underlying neural characteristics. Thus, our hypothesis was that the neural systems underlying reasoning (involving syllogisms with neutral content) following exposure to each of angry and sad emotion expression would differ from the neural underpinnings of reasoning in the neutral condition, and would thereby elucidate the mechanisms underlying differences in reasoning performance in the two emotional contexts.

Materials and Methods

Participants

Data were acquired from 17 participants (10 males, 7 females). Education levels ranged from partially completed undergraduate study to completed graduate degrees, with a mean of 16 years (SD = 2.04) of education. Ages ranged from 20 to 38 (mean 26.5 years, SD 5.95).

The study was approved by the York University Research Human Participants Ethics Committee. All participants gave informed consent.

Stimuli

Reasoning stimuli consisted of 80 syllogisms that were emotionally neutral in content. The arguments in 39 of these syllogisms were logically valid whereas the arguments in the remaining 41 were logically invalid. Examples of syllogisms are "All gentle pets are canines. Some kittens are gentle pets. Some kittens are canines" (which is valid), and "No fruits are fungi. All mushrooms are fungi. Some mushrooms are fruits" (which is invalid).

As well, there were 40 baseline "syllogisms," in which the concluding sentence was taken from a different syllogism in the dataset, thereby ensuring that the conclusion of the baseline would be unrelated to the content of the two premises. An example of a baseline trial is "Some movie-goers are men. All men are French. No people are priests." Thus, the baseline trials provide a control for the reasoning trials, in that the following processes are held constant across both types of trials: hearing the speaker deliver sentences with neutral semantics, hearing the emotion in the tone of voice (constant within each condition), learning the two premises of each argument, and preparing to engage in reasoning. Crucially, what is *not* held constant is that, in a baseline trial, the participant would disengage from the reasoning process instead of making any attempt to integrate the "conclusion" into the premises.

We controlled for the effect of belief-bias (Evans, 2003; Goel and Dolan, 2003a) by ensuring the reasoning syllogisms were balanced overall for validity and for congruence between logic and beliefs. Congruence occurs when the argument logic is valid and the conclusion is believable or when the argument logic is invalid and the conclusion is unbelievable. Incongruence occurs when the argument logic is valid and the conclusion is unbelievable or when the argument logic is invalid and the conclusion is believable.

Congruent syllogisms, incongruent syllogisms, and baselines were chosen (during study design) for each level of the Emotion factor (Sad, Neutral, and Angry). Then the order of the 120 trials was randomized. Finally, the trials were segregated into three presentation sets of 40 trials each. The order of presentation of these three sets was counterbalanced among participants, one set for each session ("run") in the scanner.

All stimuli had been pre-recorded by the same female speaker (Laura-Lee Balkwill). Among the 80 reasoning syllogisms, the tone of voice was sad for 20, angry for 20, and neutral for 40 stimuli. Among the 40 baseline "syllogisms," the tone of voice was sad for 10, angry for 10, and neutral for 20 stimuli. Please refer to the Supplementary Material for a discussion concerning the frequency of baseline trials. The intended expression of emotion of all of the stimuli was determined by a separate pilot test involving 15 participants who did not participate in the main experiment. See Appendix A for details.

Study Design

Each trial involved the following presentation sequence (see **Figure 1**): On each trial, the participant listened to a syllogism through earphones; the task was to press one of two keys to indicate whether or not the conclusion followed logically from the two previous statements. Each participant used one hand for both responses; choice of hand was counterbalanced among participants. Soundfiles varied in length from 7.4 to 15.6 s (mean 10.74 s, SD 1.77 s). However, presentation of the next sound stimulus was not entrained to the preceding response but was timed to be in synchrony with the acquisition of the brain scans. Therefore, trials varied in length from 16.53 to 16.74 s (mean 16.65 s, SD 0.024 s).

*f*MRI Scanning Technique

A 1.5T Siemens VISION system (Siemens, Erlangen, Germany) was used to acquire T1 anatomical volume images (1 mm × 1 mm × 1.5 mm voxels) and T2*-weighted images (64 × 64, 3 × 3-mm pixels, TE = 40 ms), obtained with a

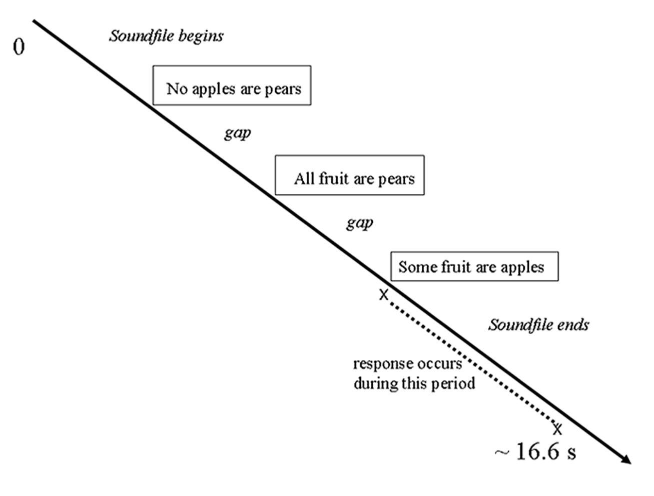


FIGURE 1 | Design of each trial.

gradient echo-planar sequence using blood oxygenation level-dependent (BOLD) contrast. Echo-planar images (2-mm thick) were acquired axially every 3 mm, positioned to cover the whole brain. Each volume (scanning of the entire brain) was partitioned into 36 slices, obtained at 90 ms per slice. Data were recorded during a single acquisition period. Volume (vol) images, 215 volumes per session, were acquired continuously, for a total of 645 volume images over three sessions, with a repetition time (TR) of 3.24 s/vol. The first six volumes in each session were discarded (leaving 209 volumes per session) to allow for T1 equilibration effects.

Data Analysis

Behavior

Behavioral data were analyzed using SPSS, version 16.0 (SPSS Inc., Chicago, IL, USA).

Note that we shall refer to the conditions as 'anger,' 'sad,' and 'neutral,' for ease of reading, rather than repeating 'expression of.'

Data from 15 of the original 17 participants were usable in the neuroimaging analysis (data from two participants were discarded because of head movement greater than 2 mm during scanning); therefore, the behavioral analyses are based on 15 participants. As well, one person's data for the third run (session) were discarded because of lack of engagement in the task. There were a total of 1760 trials remaining: 1175 reasoning (66.76%) and 585 baselines (33.24%). Fifty percentage of trials were neutral; 25% were sad, and 25% were angry. Thus, half of all trials were neutral and half were emotional.

Neuroimaging

The functional imaging data were preprocessed and subsequently analyzed using Statistical Parametric Mapping SPM8 (Friston et al., 1994; Wellcome Department of Imaging Neuroscience[1]).

All functional volumes were spatially realigned to the first volume. All volumes were temporally realigned to the AC–PC slice, to account for different sampling times of different slices. A mean image created from the realigned volumes was coregistered with the structural T1 volume and the structural volumes spatially normalized to the Montreal Neurological Institute brain template (Evans et al., 1993) using non-linear basis functions (Ashburner and Friston, 1999). The derived spatial transformation was then applied to the realigned T2* volumes, which were finally spatially smoothed with a 12 mm FWHM isotropic Gaussian kernel in order to make comparisons across subjects and to permit application of random field theory for corrected statistical inference (Worsley and Friston, 1995). The resulting time series across each voxel were high-pass filtered with a cut-off of 128 s, using cosine functions to remove section-specific low frequency drifts in the BOLD signal. Global means were normalized by proportional scaling to a grand mean of 100, and the time series temporally smoothed with a canonical hemodynamic response function to swamp small temporal autocorrelations with a known filter.

During each trial, the participant listened to the aural delivery of premise one, premise two, and the conclusion of the syllogism. This was followed by a period of silence during which the participant could indicate, by a keypress, whether or not the conclusion logically followed from the first two statements. During neuroimaging data analysis, the emotion expression time window was defined as "listening to premise one and premise two, plus the gap following premise two." The reasoning time window was defined as "the gap from offset of the conclusion up to but not including the actual motor response." Each of these time windows was analyzed separately.

Within each stimulus soundfile, the mean decibel level was calculated for the time segment corresponding to each brain scan that had been acquired. During the first level of neuroimaging analysis, described below, the potential confound of mean decibel level was covaried out.

Condition effects at each voxel were estimated according to the general linear model and regionally specific effects compared using linear contrasts. Each contrast produced a statistical parametric map of the t-statistic for each voxel, which was subsequently transformed to a unit normal Z-distribution. The BOLD signal was modeled as a canonical hemodynamic response function with time derivative.

Emotion Expression Time Window

All events from the emotion expression time window (sad, angry, and neutral listening) were modeled in the design matrix as epochs, and events of no interest (conclusion, thinking, and motor response) were modeled out. Sad, angry, and neutral listening were each modeled as an epoch from onset of premise one, with duration being the length of the syllogism *minus* the length of the conclusion. Onset for the conclusion condition was the start of hearing the conclusion; onset for the thinking condition was the end of hearing the conclusion; and onset for the motor response was the scan being acquired at the onset time of each motor response for each participant for each trial. Mean decibel level for each scan was covaried out during this first level analysis.

[1] http://www.fil.ion.ucl.ac.uk/spm

Contrast images were subsequently analyzed at the group level. A one-way univariate analysis of variance (ANOVA), within-subjects, was conducted with three conditions of interest (sad, angry, and neutral) and 15 subject conditions, with correction for non-sphericity. The analysis generates one F test for the effects of interest. The F test generated a statistical parametric map of the *F-ratio* for each voxel. The subsequent comparisons each generated a statistical parametric map of the t-statistic for each voxel, which was subsequently transformed to a unit normal Z-distribution. The activations reported in Supplementary Table S1 survived a threshold of $p < 0.005$ using a random effect model and an extent of 180 voxels. This choice of threshold and extent corresponds to a corrected $p < 0.05$ using the AlphaSim program[2] with parameters (FWHMx = 8.35 mm, FWHMy = 6.59 mm, FWHMz = 7.74 mm, within the avg152T2.nii mask from the SPM toolbox). The real smoothness in the three directions was estimated from the residuals by using 3dFWHMx. (This AlphaSim procedure was also used during the reasoning time-window, with the following parameters: FWHMx = 8.33 mm, FWHMy = 6.58 mm, FWHMz = 7.71 mm.)

Reasoning Time Window

For first-level analysis of the reasoning window, the scans acquired while the participant was engaged in reasoning were modeled as epochs by task (reasoning, baseline) and emotion (sad, angry, neutral) whereas all other conditions (Premise 1, Premise 2, Conclusion, motor response) were modeled out as events of no interest.

Onset for the six Emotion × Task conditions was the end of the conclusion sentence. Duration was from that moment until the individual participants' motor response within each trial. However, for those trials where there was no response, or the response occurred after the start of the next trial, the duration was set as "start of the next soundfile *minus* 200 ms." For those trials where participants responded during the concluding sentence (6% of trials), the duration was set as 100/3240 (that is, 0.03 TR); this strategy allowed us to include the contrast image (rather than having an unbalanced design) while ensuring minimal contribution of the activations to the analysis. Onset for each premise and the conclusion was the beginning of the relevant sentence; onset of the motor response was the millisecond at which that response occurred. Thus, altogether, 10 (conditions) × 3 (sessions) contrast images were generated for each participant. Mean decibel level for each scan was covaried out.

Contrast images were subsequently analyzed at the group level. A one-way univariate ANOVA was conducted, within-subjects, with six conditions of interest (sad reasoning, sad baseline, angry reasoning, angry baseline, neutral reasoning, neutral baseline) and 15 subject conditions, with correction for non-sphericity. The analysis generates one F test for the effects of interest.

The F test and the subsequent *a priori* comparisons each generated a statistical parametric map of the t-statistic for each voxel, which was subsequently transformed to a unit normal Z-distribution. The activations reported in Supplementary Table S2 survived a threshold of $p < 0.005$ using a random effect model and an extent of 180 voxels. (See the above description regarding the emotion expression time-window for details.)

Results

Behavioral Results

The overall percentage of correct responses on the reasoning trials was 66.9%. For baselines (where the correct response would always be "not valid"), the percentage of correct responses was 99.3%. Mean reaction time, after presentation of the third sentence, on reasoning trials was 2211 ms (SD 1121), and on baseline trials it was 472 ms (SD 112). This difference was significant: paired $t(14) = -6.366, p = 0.001$.

For each participant, the percentage of correct responses was calculated within each level of the Emotion factor. A repeated-measures analysis was conducted, using the multivariate approach; the omnibus test was significant: $F(2,13) = 4.084$, $p = 0.042$. The Emotion factor (tone of voice) accounted for 38.6% of the total variance in the percentage of correct responses. The percentage of correct responses was significantly higher in the Angry condition than in the Neutral condition ($p = 0.031$, corrected for multiple comparisons using Bonferroni). See **Figure 2**.

Mean percentages of correct responses were as follows: neutral 64.4% (SD 14.9); sad 66.1% (SD 16.5); angry 72.6% (SD 16.7).

A repeated-measures analysis of response time on correct responses was conducted across the Emotion factor. There was no significant difference among the means ($p = 0.818$). Mean reaction times were as follows: neutral 1599 ms (SD 480); sad 1626 ms (SD 672); angry 1671 ms (SD 573).

Neuroimaging Results

Emotion Expression Time Window

As indicated in Supplementary Table S1, in the contrast (Emotion − Neutral), relative deactivation was found in left hippocampus extending into left insula and relative activation was found in right posterior insula extending into

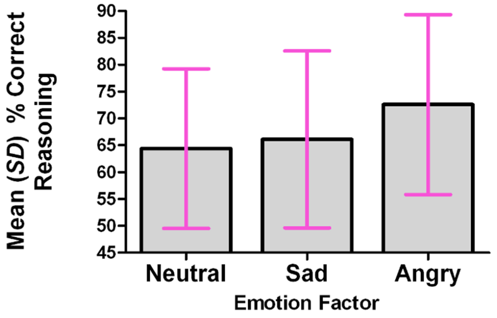


FIGURE 2 | The percentage of correct reasoning responses was significantly higher in the angry condition than in the neutral or sad conditions.

[2] http://afni.nimh.nih.gov/pub/dist/doc/manual/AlphaSim.pdf

right inferior temporal gyrus. The reverse contrast, namely (Neutral − Emotion), yielded relative deactivation in left inferior frontal gyrus (opercularis, extending into triangularis area 45) and in left precentral gyrus extending into left superior frontal gyrus. The contrast (Sad − Neutral, masked inclusively with Emotion − Neutral at $p = 0.05$) yielded relative activation in left hippocampus extending into left precuneus, in right hippocampus extending into right inferior temporal gyrus and right fusiform, in left inferior temporal gyrus extending into left hippocampus and fusiform, and in right primary somatosensory cortex extending into right precentral gyrus (area 6; see **Figure 3**). The reverse contrast (Neutral – Sad) yielded relative activation in left superior temporal gyrus extending into middle temporal gyrus, in right superior temporal gyrus, relative deactivation in left cerebellum extending into right cerebellar vermis, in left inferior frontal gyrus (opercularis: area 44), in left calcarine gyrus (area 17), and in right cerebellum. The contrast (Angry − Neutral, masked inclusively with Emotion − Neutral at $p = 0.05$) yielded relative activation in left superior temporal gyrus, in right superior temporal gyrus, and in right supramarginal gyrus extending into right superior temporal gyrus (see **Figure 4**). The reverse contrast (Neutral − Angry) yielded relative deactivation in left superior frontal gyrus (area 6), in left supramarginal gyrus, and in right angular gyrus. The contrast (Sad − Angry, masked inclusively with Emotion − Neutral at $p = 0.05$) yielded relative activation in left hippocampus extending into left cuneus, and in right hippocampus extending into right inferior temporal gyrus. The reverse contrast (Angry − Sad, masked inclusively with Emotion − Neutral at $p = 0.05$) yielded relative activation in left superior temporal gyrus extending into secondary somatosensory cortex, and in right superior temporal gyrus.

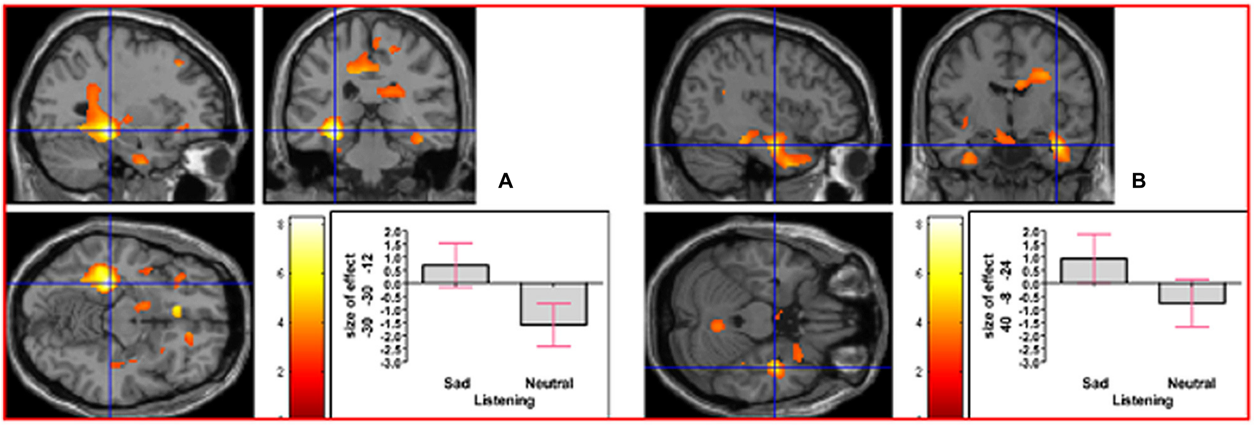


FIGURE 3 | The contrast (Sad − Neutral) elicited activation in (A) left hippocampus (MNI co-ordinates: −30, −30, −12, cluster size 6766 voxels, *Z* = 5.83), and in (B) right hippocampus (MNI co-ordinates: 40, −8, −24, cluster size 1135 voxels, *Z* = 4.51). There was also activation in left inferior temporal gyrus and in right primary somatosensory cortex (not shown).

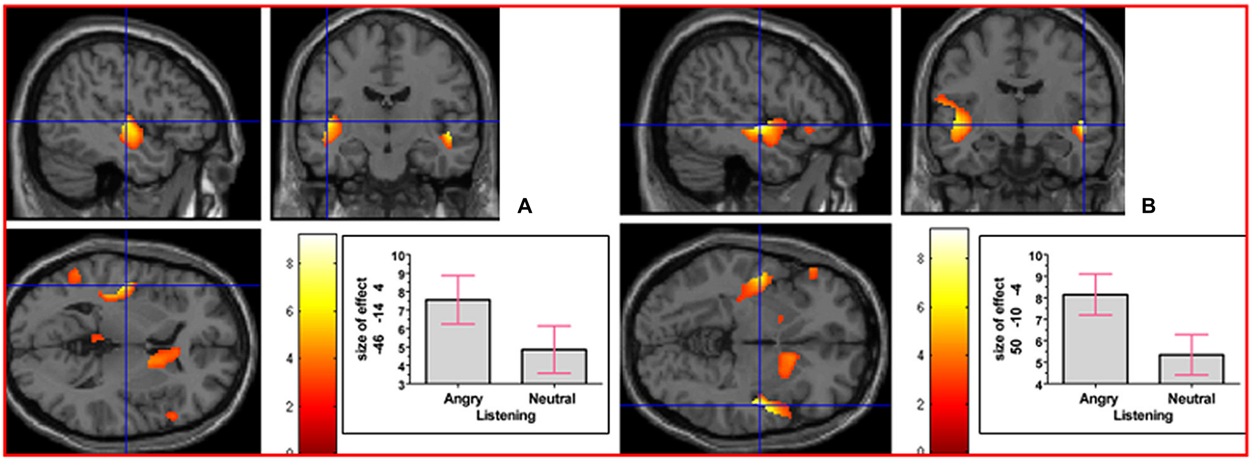


FIGURE 4 | The contrast (Angry − Neutral) elicited activation in (A) left superior temporal gyrus (MNI co-ordinates: −46, −14, 4, cluster size 746 voxels, *Z* = 5.01), and in (B) right superior temporal gyrus (MNI co-ordinates: 50, −10, −4, cluster size 463 voxels, *Z* = 6.20). There was also activation in right supramarginal gyrus (not shown).

Reasoning Time Window

As indicated in Supplementary Table S2, analysis of the main effect of (Reasoning – Baseline) yielded relative activation in right insula extending into right caudate nucleus, in left precentral gyrus extending into left primary somatosensory cortex, and in left insula extending into left inferior frontal gyrus (triangularis). Analysis of the main effect (Emotional Reasoning – Emotional Baseline) yielded relative activation in right thalamus (temporal) extending into right insula, in left precentral gyrus extending into left primary somatosensory cortex, and in right middle cingulate cortex.

For results of simple effect analyses please refer to the Supplementary Material including Supplementary Table S2.

We next addressed the question of whether neural activation underlying reasoning in an emotional context, collapsed across the emotion factor, would differ from that underlying neutral reasoning. The interaction contrast [(Emotional Reasoning – Emotional Baseline) – (Neutral Reasoning – Neutral Baseline)] yielded relative activation in left thalamus (temporal) extending into right thalamus (temporal) and right caudate nucleus, and in right middle cingulate cortex (see **Figure 5**). For details of the reverse interaction contrast, see the Supplementary Material including Supplementary Table S2.

To determine whether neural activation underlying reasoning in the sad and neutral time windows would differ, we analyzed the interaction contrast [(Sad Reasoning – Sad Baseline) – (Neutral Reasoning – Neutral Baseline)]; this analysis yielded no clusters surviving the specified extent. For details of the reverse interaction contrast, see the Supplementary Material including Supplementary Table S2.

To determine whether neural activation underlying reasoning in the angry and neutral time windows would differ, we analyzed the interaction contrast [(Angry Reasoning – Angry Baseline) – (Neutral Reasoning – Neutral Baseline)]; this analysis yielded relative activation in right superior frontal gyrus and in right thalamus (prefrontal; see **Figure 6**). For details of the reverse interaction contrast, see the Supplementary Material including Supplementary Table S2.

To determine whether neural activation underlying reasoning in the sad and angry time windows would differ, we analyzed the interaction contrast [(Sad Reasoning – Sad Baseline) – (Angry Reasoning – Angry Baseline)] and also the reverse interaction contrast [(Angry Reasoning – Angry Baseline) – (Sad Reasoning – Sad Baseline)]; neither of these interaction contrasts yielded any clusters surviving the specified extent.

To determine whether there would be any activations in common between sad reasoning and angry reasoning after accounting for their respective baselines, we conducted a conjunction analysis of the two interaction contrasts [(Sad Reasoning – Sad Baseline) – (Neutral Reasoning – Neutral Baseline)] and [(Angry Reasoning – Angry Baseline) – (Neutral Reasoning – Neutral Baseline)]; however, there were no suprathreshold clusters.

Discussion

Engagement with the Task

First, we consider whether participants were engaged in the reasoning task, by looking first at the behavioral and then at the neural results. Behaviorally, we note that accuracy levels were above chance. At the neural level, we have reported caudate nucleus involvement in several reasoning contrasts, including the main effect of reasoning. Such findings are consistent with the important role of basal ganglia in the reasoning process, as reported in the literature (Goel et al., 2000; Christoff et al., 2001; Melrose et al., 2007; Smith et al., 2014).

Success of Tone of Voice Manipulations

Second, we consider whether our tone of voice manipulations were successful. Reasoning performance in the sad condition

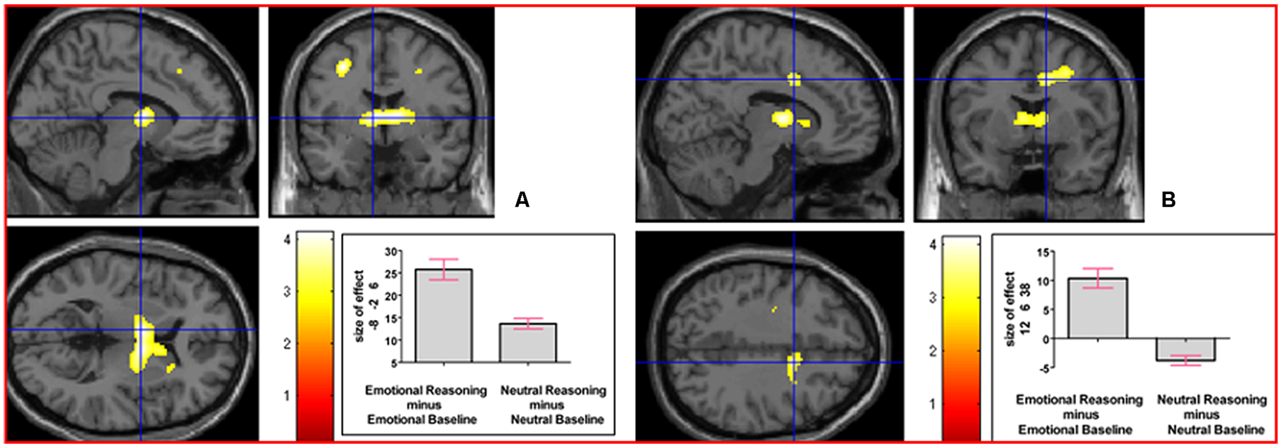

A
4
3
2
1
30
25
20
15
10
5
size of effect -8 -2 6
Emotional Reasoning minus Emotional Baseline
Neutral Reasoning minus Neutral Baseline
B
4
3
2
1
15
10
5
0
-5
size of effect 12 6 38
Emotional Reasoning minus Emotional Baseline
Neutral Reasoning minus Neutral Baseline

FIGURE 5 | The interaction contrast [(Emotional Reasoning – Emotional Baseline) – (Neutral Reasoning – Neutral Baseline)] elicited activation in (A) left thalamus (MNI co-ordinates: –8, –2, 6, cluster size 832 voxels, *Z* = 3.88), and in (B) right middle cingulate cortex (MNI co-ordinates: 12, 6, 38, cluster size 311 voxels, *Z* = 3.47).

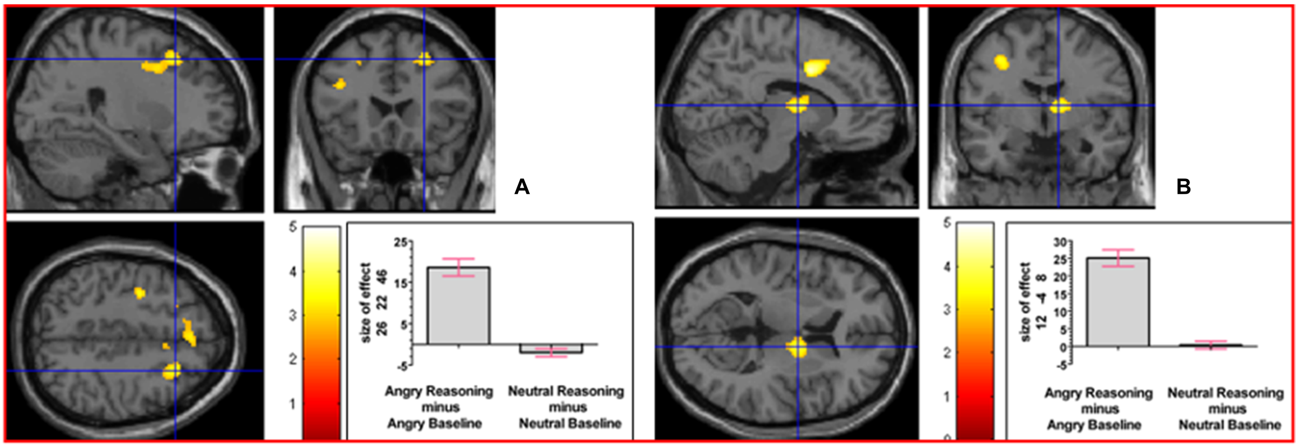

A
5
4
3
2
1
25
15
5
-5
size of effect
26 22 46
Angry Reasoning minus Angry Baseline
Neutral Reasoning minus Neutral Baseline
B
5
4
3
2
1
0
30
25
20
15
10
5
0
-5
size of effect
12 -4 8
Angry Reasoning minus Angry Baseline
Neutral Reasoning minus Neutral Baseline

FIGURE 6 | The interaction contrast [(Angry Reasoning − Angry Baseline) − (Neutral Reasoning − Neutral Baseline)] elicited activation in (A) right superior frontal gyrus (MNI co-ordinates: 26, 22, 46, cluster size 611 voxels, $Z = 3.43$), and in (B) right thalamus (MNI co-ordinates: 12, −4, 8, cluster size 220 voxels, $Z = 3.67$).

was neither impaired nor improved compared to reasoning in the neutral condition. However, reasoning performance in the angry condition was better than in the neutral tone of voice condition. If we were to consider only the behavioral results, we might conclude that the sad tone of voice was ineffective. However, the pattern of neural results indicates that each of the two tones of voice were successful: During the listening time window, each emotive tone of voice condition yielded a different pattern of neural activation. Specifically, the contrast "sad *minus* neutral" activated a different neural pattern than did the contrast "anger *minus* neutral." As well, the contrasts "sad *minus* angry" and "angry *minus* sad" yielded different patterns of neural activation. Thus, evidence shows that while participants were listening to the syllogism, they were being affected, concurrently, by the emotion expression, whether in the sad or in the angry condition.

The field of emotion research still has much to learn about the decoding and interpretation of auditory anger; thus, we should consider the possibility that our 'anger' stimuli invoked responses in the participants that would be more associated with fearful expression than expression of anger. We did not obtain emotion ratings during scanning, nor did we acquire peripheral psychophysical measurements from study participants. However, converging evidence from the pilot study of stimuli ratings and from other sources points more toward 'anger' than toward 'fear.'

During the pilot study, participants had the opportunity on 50% of trials to reject both 'sad' and 'angry' as ratings in favor of writing down a preferred term; nevertheless no participant wrote 'fear' for any stimulus. On the other 50% of trials, participants were asked to rate stimuli in terms of being active (goal-oriented) or passive (no goal) rather than choosing an emotion term. Only one participant rated one 'angry' stimulus as passive. On 100% of trials, participants indicated how sure they were of each rating; for each of sad and angry, people indicated 'yes' or 'definitely' (rather than 'maybe') on 29 out of 30 stimuli being rated. Please refer to Appendix A for details. Secondly (see below), neural activation associated with anger expression in the current study was similar to that reported by Grandjean et al. (2005). We did not find any neural activation in amygdala, a neural region often associated with fear (LeDoux, 1996; van Well et al., 2012; Adolphs, 2013).

Interpretation of Findings Regarding Reasoning in an Angry Context

We now consider how the findings regarding reasoning in an angry context should be interpreted. In two separate studies, induced anger has been shown to enhance heuristic rather than analytical processing (Bodenhausen et al., 1994; Tiedens and Linton, 2001). In contrast, Gable and Harmon-Jones (2010b) proposed that emotions such as anger that are associated with high motivation toward a goal should promote selective attention toward a target and away from irrelevant distraction. Indeed, that model fits well with our behavioral findings, which were that reasoning (the target task) improved after angry tone of voice (which was not the focus of the assigned task) compared to reasoning after neutral tone of voice.

As reported above, neural activation associated with hearing the voice of an angry speaker (Sander et al., 2005) was noted in bilateral superior temporal sulcus (right BA 42, bilateral BA 22), and right amygdala; Grandjean et al. (2005) demonstrated that superior temporal lobe activation associated with anger prosody is associated with the angry emotion itself, and not with low-level acoustical properties of the stimulus. Sander et al. (2005) utilized a dichotic listening task, which was to attend to the left- or right-ear presentation and identify the gender of the speaker; there was no instruction associated with the speaker's angry or neutral tone of voice. The above findings (in Sander et al., 2005) were for angry prosody regardless of whether attended or not; however, neural data were also analyzed separately for the attended and unattended ear of presentation. There was a tendency (in Sander et al., 2005) for activation in orbito-frontal cortex to increase in the attended-side angry prosody condition and to decrease in the unattended-side neutral prosody condition. Also, there was a tendency for activation in bilateral ventro-lateral prefrontal cortex to increase in the attended-side angry prosody condition. There was also activation in right cuneus associated with attended anger, but this activation did not survive correction for multiple comparisons. In the current study, we noted activation in *left* cuneus associated with the angry *reasoning* condition, but we did not find any activations in orbito-frontal cortex, ventro-lateral prefrontal cortex, right cuneus, or amygdala, in either the angry listening time window or the angry reasoning time window. Thus, neural activations previously associated with attention to the anger prosody were not apparent among our findings.

Selective attention has often been associated with neural activation in right superior frontal gyrus (see the review by Corbetta and Shulman, 2002). In the current study, reasoning in the angry condition was found to be associated with significant activation in right superior frontal gyrus and in right thalamus.

Thus, converging behavioral and imaging evidence suggests that, during the listening time window, angry tone of voice led to activation of neural regions previously associated with unattended anger; subsequently, during the (silent) reasoning time window, a neural region previously associated with selective attention toward the main task (in this case, reasoning) was recruited and participants' level of reasoning performance was sharper than it was after neutral tone of voice.

Interpretation of Findings Regarding Reasoning in a Sad Context

Clearly, a different mechanism was at work as a result of the expression of sad tone of voice. As we indicated above, the expressed sadness itself was effective, leading to a differentiated pattern of neural activation during the listening time window. Looking at past literature, we note that auditory induction of sadness, using sad classical music, led to activation in hippocampus/amygdala and auditory association areas (Mitterschiffthaler et al., 2007); as in that study, our use of sad expression led to extensive activation in hippocampus during the listening time window. However, in Mitterschiffthaler et al. (2007) participants were directed to pay attention to their emotional experience during scanning. A different study showed that emotional memories, but not neutral memories, have been associated with hippocampal and amygdala activation (Dolcos et al., 2004). Therefore, we propose that in the current study, participants were attending to the sad tone of voice while simultaneously learning the syllogism. However, given that reasoning performance in the sad condition was comparable to that in the neutral condition, we conclude that sad emotive tone of voice did not significantly impact the reasoning process itself.

Conclusion

We have contributed to a deeper understanding of the characterization of specific emotions, by demonstrating that two contexts of expressed emotion, each being of negative valence, have nevertheless different effects on reasoning. Unlike sad auditory context, logical reasoning in an angry auditory context is characterized by increased accuracy, and is accompanied by recruitment of an underlying neural system known to be associated with selective attention. These results increase our understanding of the neural processes that underlie reasoning in the context of auditory emotion.

Acknowledgment

This study was funded by a Wellcome Trust Grant (ABH00 FA032YBH064) to VG.

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fnhum.2015.00273/abstract

References

Adolphs, R. (2013). The biology of fear. *Curr. Biol.* 23, R79–R93. doi: 10.1016/j.cub.2012.11.055

Ashburner, J., and Friston, K. J. (1999). Nonlinear spatial normalization using basis functions. *Hum. Brain Mapp.* 7, 254–266. doi: 10.1002/(SICI)1097-0193(1999)7:4<254::AID-HBM4>3.0.CO;2-G

Bodenhausen, G. V., Sheppard, L. A., and Kramer, G. P. (1994). Negative affect and social judgment: the differential impact of anger and sadness. *Eur. J. Soc. Psychol.* 24, 45–62. doi: 10.1002/ejsp.2420240104

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

Carver, C. S., and Harmon-Jones, E. (2009). Anger is an approach-related affect: evidence and implications. *Psychol. Bull.* 135, 183–204. doi: 10.1037/a0013965

Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J. K., Holyoak, K. J., et al. (2001). Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *Neuroimage* 14, 1136–1149. doi: 10.1006/nimg.2001.0922

Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215. doi: 10.1038/nrn755

DeSteno, D., Petty, R. E., Rucker, D. D., Wegener, D. T., and Braverman, J. (2004). Discrete emotions and persuasion: the role of emotion-induced expectancies. *J. Pers. Soc. Psychol.* 86, 43–56. doi: 10.1037/0022-3514.86.1.43

Dolcos, F., LaBar, K. S., and Cabeza, R. (2004). Interaction between the amygdala and the medial temporal lobe memory system predicts better memory for emotional events. *Neuron* 42, 855–863. doi: 10.1016/S0896-6273(04)00289-2

Dunn, J. R., and Schweitzer, M. E. (2005). Feeling and believing: the influence of emotion on trust. *J. Pers. Soc. Psychol.* 88, 736–748. doi: 10.1037/0022-3514.88.5.736

Egner, T., Monti, J. M., and Summerfield, C. (2010). Expectation and surprise determine neural population responses in the ventral visual stream. *J. Neurosci.* 30, 16601–16608. doi: 10.1523/JNEUROSCI.2770-10.2010

Eickhoff, S., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, Z., et al. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 25, 1325–1335. doi: 10.1016/j.neuroimage.2004.12.034

Ethofer, T., Kreifelts, B., Wiethoff, S., Wolf, J., Grodd, W., Vuilleumier, P., et al. (2009). Differential influences of emotion, task, and novelty on brain regions underlying the processing of speech melody. *J. Cogn. Neurosci.* 21, 1255–1268. doi: 10.1162/jocn.2009.21099

Evans, A. C., Collins, D. L., Mills, S. R., Brown, E. D., Kelly, R. L., and Peters, T. M. (1993). "3D statistical neuroanatomical models from 305 MRI volumes," in *Proceedings of the Nuclear Science Symposium and Medical Imaging Conference 1993*, *1993 IEEE Conference Record* (San Francisco, CA: IEEE), 3, 1813–1817. Avaialable at: http://www.ece.uvic.ca/∼btill/papers/learning/Evans_etal_1993.pdf [accessed February 26, 2010].

Evans, J. S. (2003). In two minds: dual-process accounts of reasoning. *Trends Cogn. Sci.* 7, 454–459. doi: 10.1016/j.tics.2003.08.012

Fangmeier, T., and Knauff, M. (2009). Neural correlates of acoustic reasoning. *Brain Res.* 1249, 181–190. doi: 10.1016/j.brainres.2008.10.025

Finucane, A. M. (2011). The effect of fear and anger on selective attention. *Emotion* 11, 970–974. doi: 10.1037/a0022574

Fletcher, P. C., Anderson, J. M., Shanks, D. R., Honey, R., Carpenter, T. A., Donovan, T., et al. (2001). Responses of human frontal cortex to surprising events are predicted by formal associative learning theory. *Nat. Neurosci.* 4, 1043–1048. doi: 10.1038/nn733

Friedman, B. H. (2010). Feelings and the body: the Jamesian perspective on autonomic specificity of emotion. *Biol. Psychol.* 84, 383–393. doi: 10.1016/j.biopsycho.2009.10.006

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., and Frackowiak, R. S. J. (1994). Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* 2, 189–210. doi: 10.1002/hbm.460020402

Gable, P., and Harmon-Jones, E. (2010a). The blues broaden, but the nasty narrows: attentional consequences of negative affects low and high in motivational intensity. *Psychol. Sci.* 21, 211–215. doi: 10.1177/0956797609359622

Gable, P., and Harmon-Jones, E. (2010b). The motivational dimensional model of affect: implications for breadth of attention, memory, and cognitive categorisation. *Cogn. Emot.* 24, 322–337. doi: 10.1080/02699930903378305

Genovese, C. R., Lazar, N. A., and Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* 15, 870–878. doi: 10.1006/nimg.2001.1037

Goel, V., Buchel, C., Frith, C., and Dolan, R. (2000). Dissociation of mechanisms underlying syllogistic reasoning. *Neuroimage* 12, 504–514. doi: 10.1006/nimg.2000.0636

Goel, V., and Dolan, R. J. (2003a). Explaining modulation of reasoning by belief. *Cognition* 87, B11–B22. doi: 10.1016/S0010-0277(02)00185-3

Goel, V., and Dolan, R. J. (2003b). Reciprocal neural response within lateral and ventral medial prefrontal cortex during hot and cold reasoning. *Neuroimage* 20, 2314–2341. doi: 10.1016/j.neuroimage.2003.07.027

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., et al. (2005). The voices of wrath: brain responses to angry prosody in meaningless speech. *Nat. Neurosci.* 8, 145–146. doi: 10.1038/nn1392

Knauff, M., Fangmeier, T., Ruff, C. C., and Johnson-Laird, P. N. (2003). Reasoning, models, and images: behavioral measures and cortical activity. *J. Cogn. Neurosci.* 15, 559–573. doi: 10.1162/089892903321662949

Knauff, M., Mulack, T., Kassubek, J., Salih, H. R., and Greenlee, M. W. (2002). Spatial imagery in deductive reasoning: a functional MRI study. *Cogn. Brain Res.* 13, 203–212. doi: 10.1016/S0926-6410(01)00116-1

Langner, R., Kellermann, T., Boers, F., Sturm, W., Willmes, K., and Eickhoff, S. B. (2011). Modality-specific perceptual expectations selectively modulate baseline activity in auditory, somatosensory, and visual cortices. *Cereb. Cortex* 21, 2850–2862. doi: 10.1093/cercor/bhr083

LeDoux, J. (1996). *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York: Simon & Schuster.

Lerner, J. S., and Keltner, D. (2001). Fear, anger, and risk. *J. Pers. Soc. Psychol.* 81, 146–159. doi: 10.1037/0022-3514.81.1.146

Liotti, M., Mayberg, H. S., Brannan, S. K., McGinnis, S., Jerabek, P., and Fox, P. T. (2000). Differential limbic-cortical correlates of sadness and anxiety in healthy subjects: implications for affective disorders. *Biol. Psychiatry* 48, 30–42. doi: 10.1016/S0006-3223(00)00874-X

Melrose, R. J., Poulin, R. M., and Stern, C. E. (2007). An fMRI investigation of the role of the basal ganglia in reasoning. *Brain Res.* 1142, 146–158. doi: 10.1016/j.brainres.2007.01.060

Mitterschiffthaler, M. T., Fu, C. H. Y., Dalton, J. A., Andrew, C. M., and Williams, S. C. R. (2007). A functional MRI study of happy and sad affective states induced by classical music. *Hum. Brain Mapp.* 28, 1150–1162. doi: 10.1002/hbm.20337

Murphy, F. C., Nimmo-Smith, I., and Lawrence, A. D. (2003). Functional neuroanatomy of emotions: a meta-analysis. *Cogn. Affect. Behav. Neurosci.* 3, 207–233. doi: 10.3758/CABN.3.3.207

Quadflieg, S., Mohr, A., Mentzel, H.-J., Miltner, W. H. R., and Straube, T. (2008). Modulation of the neural network involved in the processing of anger prosody: the role of task-relevance and social phobia. *Biol. Psychol.* 78, 129–137. doi: 10.1016/j.biopsycho.2008.01.014

Sander, D., Grandjean, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., et al. (2005). Emotion and attention interactions in social cognition: brain regions involved in processing anger prosody. *Neuroimage* 28, 848–858. doi: 10.1016/j.neuroimage.2005.06.023

Smith, K. W., and Greenberg, L. S. (2007). Internal multiplicity in emotion-focused psychotherapy. *J. Clin. Psychol.* 63, 175–186. doi: 10.1002/jclp.20340

Smith, K. W., Vartanian, O., and Goel, V. (2014). Dissociable neural systems underwrite logical reasoning in the context of induced emotions with positive and negative valence. *Front. Hum. Neurosci.* 8:736. doi: 10.3389/fnhum.2014.00736

Tiedens, L. Z., and Linton, S. (2001). Judgment under emotional certainty and uncertainty: the effects of specific emotions on information processing. *J. Pers. Soc. Psychol.* 81, 973–988. doi: 10.1037/OO22-3514.81.6.973

van Well, S., Visser, R. M., Scholte, H. S., and Kindt, M. (2012). Neural substrates of individual differences in human fear learning: evidence from concurrent fMRI, fear- potentiated startle, and US-expectancy data. *Cogn. Affect. Behav. Neurosci.* 12, 499–512. doi: 10.3758/s13415-012-0089-7

Wessel, J. R., Danielmeier, C., Morton, J. B., and Ullsperger, M. (2012). Surprise and error: common neuronal architecture for the processing of errors and novelty. *J. Neurosci.* 32, 7528–7537. doi: 10.1523/JNEUROSCI.6352-11.2012

Worsley, K. J., and Friston, K. J. (1995). Analysis of fMRI time-series revisited – again. *Neuroimage* 2, 173–181. doi: 10.1006/nimg.1995.1023

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.

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 23 September 2014
doi: 10.3389/fnhum.2014.00736

Dissociable neural systems underwrite logical reasoning in the context of induced emotions with positive and negative valence

Kathleen W. Smith[1], Oshin Vartanian[2] and Vinod Goel[1,3,4]*

[1] York University, Toronto, ON, Canada
[2] University of Toronto Scarborough, Toronto, ON, Canada
[3] University of Hull, Hull, UK
[4] IRCCS Fondazione Ospedale San Camillo, Venice, Italy

Edited by:
Jérôme Prado, Centre National de la Recherche Scientifique, France

Reviewed by:
Isabelle Blanchette, Université du Québec à Trois-Rivières, Canada
Mathieu Cassotti, Université Paris Descartes-Sorbonne Paris Cité, France
Michael Vendetti, University of California Berkeley, USA

***Correspondence:**
Vinod Goel, Faculty of Health, Department of Psychology, York University, 4700 Keele Street, Toronto, ON M3J 1P3, Canada
e-mail: vgoel@yorku.ca

How emotions influence syllogistic reasoning is not well understood. fMRI was employed to investigate the effects of induced positive or negative emotion on syllogistic reasoning. Specifically, on a trial-by-trial basis participants were exposed to a positive, negative, or neutral picture, immediately prior to engagement in a reasoning task. After viewing and rating the valence and intensity of each picture, participants indicated by keypress whether or not the conclusion of the syllogism followed logically from the premises. The content of all syllogisms was neutral, and the influence of belief-bias was controlled for in the study design. Emotion did not affect reasoning performance, although there was a trend in the expected direction based on accuracy rates for the positive (63%) and negative (64%) versus neutral (70%) condition. Nevertheless, exposure to positive and negative pictures led to dissociable patterns of neural activation during reasoning. Therefore, the neural basis of deductive reasoning differs as a function of the valence of the context.

Keywords: reasoning, emotion, fMRI, IAPS, belief-bias, positive, negative

INTRODUCTION

Although the empirical literature examining the effects of emotion on cognition is very large, relatively few studies have investigated the effect of emotion on logical reasoning. Behavioral studies that have investigated this effect have usually found that compared to neutral valence, positive and negative valence result in impaired accuracy in logical reasoning. This has been shown to be true regardless of whether the emotions are manipulated via the content of the logical arguments (Lefford, 1946), mood of the participants (Melton, 1995; Oaksford et al., 1996), or both (Blanchette and Richards, 2004; Blanchette, 2006). See also the review by Blanchette and Richards (2010).

However, other studies have reported no impairment in cognitive processing associated with negative emotion. In fact, sadness and depression have been found to promote systematic cognitive processing (Alloy and Abramson, 1979; Schwarz and Bless, 1991; Bless et al., 1992; Bohner et al., 1992; Edwards and Weary, 1993). Blanchette et al. (2007) found that reasoning in the negative condition improved logical reasoning by reducing belief-bias, but only when the material referred to participants' actual exposure to terrorist activity; otherwise, reasoning in the negative condition was impaired, both for other participant groups on all negative material and for the group exposed to actual terrorist activity on non-terror-related negative material. Goel and Vartanian (2011) found that, when argument logic and beliefs about the material itself required opposite responses (incongruence) on a given trial, reasoning performance was better when the reasoning material was politically incorrect than when otherwise. These results suggest that under some conditions negative content can improve reasoning performance.

The inconsistency in the literature on the effect of emotion on cognitive processes could arise from various sources, such as variations in the type of stimulus materials, incongruence between argument logic and one's beliefs about the content, or presentation of the emotion as either part of the content or separately, as part of the context.

To extend this literature, we explored whether the effects of emotion on underlying reasoning processes differ depending on whether the emotion is positive or negative. This exploration was motivated by evidence suggesting that positive and negative emotions may exert different effects on cognition. Positive emotion promotes creativity (Isen et al., 1987) and facilitates noticing more relations among concepts (Isen and Daubman, 1984). It also promotes a reliance on such heuristic shortcuts as source expertise and stereotyping instead of considering the evidence when making evaluations (Schwarz and Clore, 1983; Bless et al., 1992; Bodenhausen et al., 1994). Positive emotion also impairs working memory (Martin and Kerns, 2011), and distracts attention toward task-irrelevant information (Biss and Hasher, 2011) at the level of early sensory encoding (Vanlessen et al., 2013). The bulk of available evidence suggests that positive emotion might exert its deleterious effects on reasoning by taxing working memory with induced bottom-up task-irrelevant information and by promoting a top-down heuristic processing mode.

There is now good evidence to suggest that positive and negative emotion induction have different effects on the brain. Using a gender identification task (to reduce attention to the emotion manipulation), Schmitz et al. (2009) found that positive emotion broadened focus to peripherally presented stimuli (houses) and was accompanied by neural activation in right lateral frontal pole (BA 10), lateral orbitofrontal cortex (BA 11), as well as by correlated activity in parahippocampal place area and primary visual cortex. In contrast, negative emotion narrowed focus to targets (faces) only, and was accompanied by neural activation in amygdala, as well as by inversely correlated activity in parahippocampal place area and primary visual cortex. In Schmitz et al. (2009), emotion had been induced by means of pictures from the International Affective Picture System (IAPS; Lang et al., 1997). In Dolcos et al. (2004), valence ratings of positive and negative IAPS pictures during scanning were accompanied by different patterns of neural activation; positive evaluations were associated with activation in left dorsolateral prefrontal cortex (BA 8/9), whereas negative evaluations were associated with activation in bilateral dorsolateral prefrontal cortex (BA 8/9) and right ventrolateral prefrontal cortex (BA 47). Using only negative IAPS pictures, Taylor et al. (2000) found that activation in the amygdala, uncus, and anterior parahippocampal gyrus was positively correlated with increasingly aversive ratings of pictures; as well, mildly aversive ratings were associated with activation in left-hemisphere posterior and subcortical regions, whereas strongly aversive ratings were associated with activation in bilateral posterior and subcortical regions and lateral orbitofrontal cortex. In general, the above reports suggest that, apart from activation in orbitofrontal cortex, positive and negative emotion induction lead to differentiated underlying patterns of neural activity; positive emotion is accompanied by medial frontal and left frontal activation, whereas negative emotion is accompanied by activation in amygdala and bilateral or right frontal activation. Patterns of activation in posterior cortical and in subcortical regions (apart from amygdala) vary depending on the task but, within these studies, differ by valence or intensity of emotion.

In the first neuroimaging study to examine the effect of emotion on deductive reasoning, Goel and Dolan (2003b) demonstrated that reasoning with negatively charged material was associated with activation in ventromedial prefrontal cortex, whereas reasoning with neutral material was associated with activation in left dorsolateral prefrontal cortex; furthermore, these neural mechanisms were activated in a reciprocal manner. In that study, emotion was manipulated using the content of the syllogism such that, depending on the condition, content was either emotionally provocative or neutral. The results demonstrated that the pattern of neural activation during reasoning varies as a function of emotional content.

In the present study, we sought to extend the findings of Goel and Dolan (2003b) by making an important change to the paradigm. Whereas Goel and Dolan varied the emotionality of the content itself, we chose to manipulate the emotionality of the context in which reasoning about neutral material would take place. Specifically, on each trial, participants first viewed and rated a picture on valence and intensity, and after the picture was removed from view, they engaged in a syllogistic reasoning task involving visually presented syllogisms with non-emotional content. This design feature enabled us to analyze the neural correlates of reasoning separately from those acquired during emotion induction itself. Secondly, whereas the emotional content in Goel and Dolan was negative and provocative, in the current study, we chose to induce not only negative but also positive emotion.

Therefore, the current study utilized a 3 (Emotion) × 2 (Task) within-subjects design, where the three levels of the Emotion factor were positive, neutral, and negative, and the two levels of the Task factor were reasoning and baseline. Also, because it is known that reasoning is subject to a belief-bias effect (Evans, 2003), we controlled for belief-bias in the study design.

Because of the more common findings in the literature, that is, that reasoning is impaired by positive or negative emotion manipulation, we hypothesized that each of positive and negative emotion would be detrimental to reasoning. Additionally, we hypothesized that the neural systems underlying reasoning under those two conditions would differ from that in the neutral condition.

MATERIALS AND METHODS

PARTICIPANTS

Data were acquired from 16 participants (7 males, 9 females). Education levels ranged from partially completed undergraduate study to completed graduate degrees, with a mean of 17.54 (SD = 3.82) years of education. Ages ranged from 19 to 56 (mean age was 28, SD = 10 years). All participants gave informed consent. The study was approved by the York University Research Human Participants Ethics Committee.

STIMULI

Pictures, normed as to emotional valence, were taken from the IAPS system (Lang et al., 1997). The valence categories from the IAPS were used to choose 40 positive and 40 negative pictures for the experiment. In addition, 40 pictures of furniture were added, to serve as neutral pictures.

Reasoning stimuli consisted of 75 syllogisms that were emotionally neutral in content. The arguments in 38 of these syllogisms were logically valid, whereas the arguments in the remaining 37 were logically invalid. An example of a valid syllogism is "All dogs are pets; All poodles are dogs; All poodles are pets," and an example of an invalid syllogism is "All paper is absorbent; All napkins are paper; No napkins are absorbent."

As well, there were 45 baseline "syllogisms," in which the concluding sentence was taken from a different syllogism in the dataset, thereby ensuring that the conclusion of the baseline would be unrelated to the content of the two premises. Thus, in a baseline trial, the participant would prepare to respond to what was expected to be a syllogism; however, the unrelated conclusion would indicate that the stimulus is not an argument and can be rejected without integrating the conclusion into the premises.

STUDY DESIGN

The study involved 120 trials delivered over 3 sessions (or "runs") in the scanner. Each trial involved the following sequence (see **Figure 1**): first, the participant saw a slide with the fixation point (xxx) for 500 ms; then the fixation point disappeared. Next, the participant viewed a picture and pressed one of eight keys to indicate simultaneously the rating of positive or negative valence and the intensity of the picture's emotional content. The specific

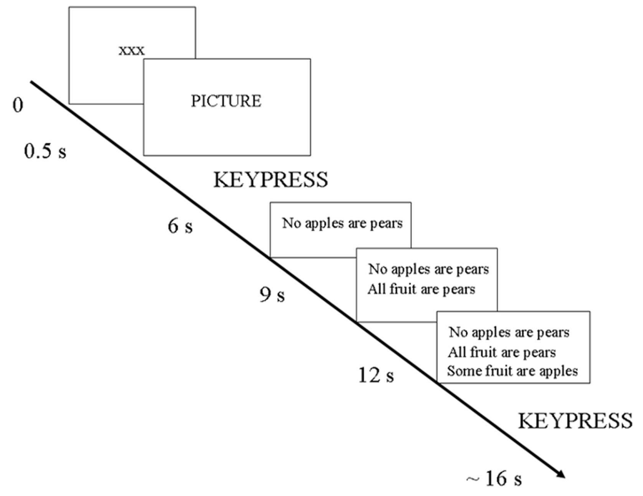


FIGURE 1 | Design of one trial.

meaning of the keys will be explained below. Then, the picture disappeared and a syllogism was presented over three consecutive slides (slide one: first premise alone; slide two: first two premises together; slide three: the two premises plus the conclusion). The syllogism remained in view during the reasoning period. The participant pressed a key to indicate whether the conclusion followed or not from the two statements (premises). Disappearance of the picture and syllogism slides was not entrained to the responses but was timed to be in synchrony with the acquisition of the brain scans. Trials varied in length and were approximately 16–20 s.

The specific meaning of the eight picture-rating keys is as follows: valence and intensity were captured in the same keypress. There were four keys in one direction for "increasingly negative" and four in the other direction for "increasingly positive." The side was counterbalanced among participants. Participants used the index finger of each hand to respond. All participants were declared as right-handed.

The effect of belief-bias was controlled for. That is, the reasoning syllogisms were balanced overall for validity and for congruence between logic and beliefs. Congruence occurs when the argument logic is valid and the conclusion is believable or when the argument logic is invalid and the conclusion is unbelievable. Incongruence occurs when the argument logic is valid and the conclusion is unbelievable or when the argument logic is invalid and the conclusion is believable.

Thus, syllogisms and baseline trials were matched to pictures so that there were equivalent numbers of congruent syllogisms, incongruent syllogisms, and baselines within each level of the emotion factor (positive, negative, and neutral). Then the order of the 120 trials was randomized. Finally, the trials were segregated into three presentation sets of 40 trials each (see Supplementary Material). Thus, pictures were not presented in blocks by valence; the valences (positive, neutral, and negative) were quasi-randomly intermixed. The order of presentation of these three sets was counterbalanced among participants, one set for each session ("run") in the scanner.

fMRI SCANNING TECHNIQUE

A 1.5-T Siemens VISION system (Siemens, Erlangen, Germany) was used to acquire T1 anatomical volume images (1 mm × 1 mm × 1.5 mm voxels) and T2*-weighted images (64 × 64, 3 mm × 3 mm pixels, TE = 40 ms), obtained with a gradient echo-planar sequence using blood oxygenation level-dependent (BOLD) contrast. Echo-planar images (2 mm thick) were acquired axially every 3 mm, positioned to cover the whole brain. Each volume was partitioned into 36 slices, obtained at 90 ms per slice. Data were recorded during a single acquisition period. Volume (vol) images, 243 per session, were acquired continuously, for a total of 729 images over three sessions, with a repetition time (TR) of 3.24 s/vol. The first six volumes in each session were discarded (leaving 237 per session) to allow for T1 equilibration effects.

DATA ANALYSIS

Behavior

Behavioral data were analyzed using SPSS, version 16.0 (SPSS Inc., Chicago, IL, USA).

In the design there were 120 trials, 75 (62.5%) involving reasoning and 45 (37.5%) baselines. Data from two participants were discarded because of movement artifacts in the neuroimaging data. Therefore, the behavioral analyses are based on 14 participants. Twelve participants completed all three sessions of 40 trials each. One participant completed two sessions. One other participant completed all three sessions, but because some of the scan volumes were missing from the data, it was necessary to excise three trials from the middle of Session 1 and one trial from the middle of Session 2. Thus, there were a total of 12 × 120 + 80 + 116 = 1636 trials. Of these, 1021 (62.4%) were reasoning trials and 615 (37.6%) were baselines. The participants' valence ratings were sorted into three categories: positive, negative, and neutral. Ratings of −2, −3, or −4 were classified as "negative"; ratings of +2, +3, or +4 were classified as "positive." Ratings of −1 or +1 were considered "neutral."

Neuroimaging

The functional imaging data were preprocessed and subsequently analyzed using Statistical Parametric Mapping SPM8 (Friston et al., 1994; Wellcome Department of Imaging Neuroscience; http://www.fil.ion.ucl.ac.uk/spm/).

All functional volumes were spatially realigned to the first volume. Data from two participants with head movement >2 mm were discarded. All volumes were temporally realigned to the AC–PC slice, to account for different sampling times of different slices. A mean image created from the realigned volumes was co-registered with the structural T1 volume and the structural volumes spatially normalized to the Montreal Neurological Institute brain template (Evans et al., 1993) using non-linear basis functions (Ashburner and Friston, 1999). The derived spatial transformation was then applied to the realigned T2* volumes, which were finally spatially smoothed with a 12 mm FWHM isotropic Gaussian kernel in order to make comparisons across subjects and to permit application of random field theory for corrected statistical inference (Worsley and Friston, 1995). The resulting time series across each voxel were high-pass filtered with a cut-off of 128 s, using cosine functions to remove section-specific low-frequency drifts in the BOLD signal. Global means were normalized by proportional scaling to a grand mean of 100, and the time series temporally smoothed with a canonical hemodynamic response function to swamp small temporal autocorrelations with a known filter.

Condition effects at each voxel were estimated according to the general linear model and regionally specific effects compared using linear contrasts. Each contrast produced a statistical parametric map of the *t* statistic for each voxel, which was subsequently transformed to a unit normal *Z* distribution. The BOLD signal was modeled as a canonical hemodynamic response function with time derivative. All events were modeled in the design matrix, but events of no interest (the first two sentences, and the two motor responses on a trial-by-trial basis) were modeled out. Positive, neutral, and negative picture viewing/rating were each modeled as an epoch from picture onset up to but excluding the motor response. Positive, neutral, and negative reasoning, and positive, neutral, and negative baseline were each modeled as an event. The onset of the event was the halfway point between presentation of the concluding sentence and the motor response.

Parametric (correlational) analyses were conducted to determine neural regions associated with increasingly intense positive and negative picture ratings. The BOLD signal was modeled as a canonical hemodynamic response function. All events were modeled in the design matrix, but events of no interest (the three sentences, and the two motor responses on a trial-by-trial basis) were modeled out. Positive intensity and negative intensity were each modeled as an event from picture onset.

The individual-level analyses involving emotion induction were subsequently analyzed at the group level in a random effects model, using *t*-tests (see Table 1 in Supplementary Material). The individual-level analyses of the reasoning time window were analyzed at the group level in a random effects model, using a 2 (Task: Reasoning, Baseline) × 3 Emotion (positive, negative, neutral) factorial design, with correction for non-sphericity and with proportional overall grand mean scaling (see Table 2 in Supplementary Material).

All reported results survived a threshold of $p < 0.005$ and an extent of $k \geq 20$ voxels, a combination that has been demonstrated to produce a desirable balance between type I and type II error rates (Lieberman and Cunningham, 2009).

RESULTS

BEHAVIORAL RESULTS

For each participant, we computed the proportion of each of positive:total ratings, neutral:total ratings, and negative:total ratings. For example, one participant rated 119 of the 120 trials, of which 39 were rated neutral; therefore, for this participant, the proportion of neutral:total ratings is 0.33. A repeated-measures analysis, multivariate approach, was conducted; the within-subjects factor was choice of valence (positive, neutral, and negative) and the dependent variable was mean proportion. Participants rated a significantly greater proportion of pictures as positive than as negative ($F_{2,11} = 9.988$, $p = 0.003$, partial $\eta^2 = 0.645$).

The mean response time to rate the pictures was calculated for each participant, separately for each valence. A repeated-measures analysis, multivariate approach, was conducted; the within-subjects factor was Emotion (positive, neutral, and negative) and the dependent variable was mean picture-rating response time. Data were analyzed for 13 participants, as 1 participant had not rated any picture as "neutral." Participants took significantly longer to rate pictures as positive than as neutral ($F_{2,11} = 5.739$, $p = 0.02$, partial $\eta^2 = 0.511$).

The mean (SD) proportion of total picture ratings for each valence was as follows: positive 0.3859 (0.108), neutral 0.2731 (0.130), negative 0.2308 (0.085); the mean (SD) response time in milliseconds to rate the pictures was as follows: positive 2184 (483), neutral 1919 (623), negative 2092 (467). See "Behavioral Scores" in Supplementary Material.

For the reasoning trials, the overall proportion of correct:total responses was 0.630. For baselines (where the correct response would always be "not valid"), the proportion of correct:total responses was 0.972. Mean reaction time was 4185 (SD 789) ms on reasoning trials overall (that is, without regard to accuracy), and 1874 (SD 456) ms on baseline trials. This difference was significant: paired $t(13) = 8.567$, $p = 0.001$.

The proportion of correct reasoning responses to the total number of reasoning trials was computed for each participant within each valence. For instance, 1 participant rated 20 of the pictures (on reasoning trials) as positive, and reasoned logically on 15 of those trials; thus, the proportion of correct responses on positively valenced reasoning trials was 0.75 for that participant. Next, a repeated-measures analysis of variance ($n = 13$; the one participant who had not rated any pictures as neutral was excluded from this analysis), multivariate approach, was conducted to test whether the valence rating affected reasoning. The independent variable was the emotion factor (positive, neutral, and negative), and the dependent variable consisted of each participant's mean proportion of correct:total reasoning responses. The result was not significant ($p = 0.391$, partial $\eta^2 = 0.157$). Overall, the valence of the picture did not significantly influence subsequent reasoning. See "Behavioral Scores" in Supplementary Material.

A repeated-measures analysis of variance, multivariate approach, indicated that mean reaction time to reasoning syllogisms overall (that is, collapsed across accuracy) did not differ by Emotion (positive, neutral, and negative). Participants responded significantly more slowly on reasoning trials when their response was incorrect than when it was correct, regardless of the valence of the trial. The main effect of accuracy was significant: $F(1, 12) = 7.537$, $p = 0.018$, partial $\eta^2 = 0.386$; there was no main effect of Emotion (positive versus negative) and no significant interaction of Accuracy × Emotion). Mean (SD) reaction times in milliseconds to syllogisms, by valence and accuracy, were as follows: for correct responses ($n = 13$), mean (SD) was 3480 (574) for positive, 3759 (729) for neutral, and 3793 (461) for negative. For incorrect responses ($n = 9$), mean (SD) was 4215 (673) for positive, 4199 (691) for neutral, and 4008 (755) for negative. For the sake of consistency with the other results, we repeated this analysis using correct trials only (repeated-measures, multivariate approach), and found that mean reaction time when responding correctly to syllogisms did not differ significantly by Emotion ($p = 0.267$, partial $\eta^2 = 0.213$).

Manipulation check demonstrating the need to control for belief-bias

Instantiation of belief-bias in the current design would be as follows: on trials where there is incongruence between argument logic and beliefs (valid argument and false belief, or invalid argument and true belief), responses should be less logical and slower than on trials where there is congruence between argument logic and beliefs (valid argument and true belief, or invalid argument and false belief). We controlled for belief-bias in the study design, by ensuring equivalent numbers of congruent syllogisms, incongruent syllogisms, and baselines within each level of the emotion factor.

We thank a reviewer for suggesting that we should test directly this possible effect of belief-bias, at the behavioral level. The proportion of correct:total responses was analyzed for congruence with beliefs (congruent, incongruent) by Emotion (positive, neutral, and negative) using a repeated-measures analysis (multivariate approach). The main effect of Congruence was significant ($F_{1,12} = 6.835$, $p = 0.023$, partial $\eta^2 = 0.363$) and the Congruence × Emotion interaction approached significance ($F_{2,11} = 3.194$, $p = 0.081$, partial $\eta^2 = 0.367$). Thus, correct responding is significantly hindered when the logic of the argument conflicts with beliefs, tending to be more so (reduced to chance level) after positive and negative than after neutral picture ratings.

The mean proportions (SD) correct:total were as follows ($n = 13$): for congruent syllogisms, positive:total was 0.727 (0.252), neutral:total was 0.729 (0.174), and negative:total was 0.762 (0.233). For incongruent syllogisms, positive:total was 0.537 (0.174), neutral:total was 0.659 (0.267), and negative:total was 0.504 (0.305).

The mean reaction time (RT) to the syllogisms where the response was correct was analyzed for congruence with beliefs (congruent, incongruent) by Emotion (positive, neutral, and negative) using a repeated-measures analysis (multivariate approach). The main effect of Congruence was significant ($F_{1,11} = 39.740$, $p < 0.001$, partial $\eta^2 = 0.783$); the Congruence*Emotion interaction was not significant ($p = 0.151$, partial $\eta^2 = 0.315$). Thus, correct responses are significantly slower when the logic of the argument conflicts with beliefs, regardless of valence.

Mean reaction times ($n = 12$) when responding correctly were as follows: (a) congruent positive: 3097 ms (SD 530); (b) congruent neutral: 3437 ms (SD 532); (c) congruent negative: 3410 ms (SD 499); (d) incongruent positive: 3901 ms (SD 829); (e) incongruent neutral: 3585 ms (SD 1077); (f) incongruent negative: 4466 ms (SD 625).

NEUROIMAGING RESULTS

Neuroimaging analysis: emotion induction time window

As indicated in Table 1 of Supplementary Material, the contrast positive–neutral yielded neural activation in left thalamus, right cerebellum, occipital lobe bilaterally, left parietal (supramarginal gyrus and secondary somatosensory area), right inferior parietal lobe, and left fusiform gyrus. The contrast negative–neutral yielded neural activation in left putamen, right amygdala, occipital lobe bilaterally, left inferior parietal (secondary somatosensory cortex and supramarginal gyrus), right inferior parietal (supramarginal gyrus), and right inferior frontal gyrus (triangularis, area 45). The contrast positive–negative yielded neural activation in left cerebellum, right hippocampus, left postcentral gyrus, and superior temporal gyrus bilaterally. The contrast negative–positive yielded neural activation in left amygdala and insula, left middle cingulate, right hippocampus, left occipital lobe, inferior parietal (supramarginal gyrus) bilaterally, left superior parietal (area 7), right precuneus, right postcentral gyrus, inferior frontal gyrus (left opercularis area 44, right area 44), left frontal (supplementary motor area and area 4), right precentral gyrus (areas 44 and 6), and superior frontal gyrus bilaterally. See Table 1 in Supplementary Material.

Parametric (correlational) analyses were conducted to determine neural regions associated with increasingly intense positive and negative picture ratings. As positive intensity increased, significant neural activation was noted in cerebellum bilaterally, left thalamus, occipital lobe bilaterally, postcentral gyrus bilaterally, middle temporal gyrus bilaterally, right inferior temporal gyrus, right fusiform gyrus, and left inferior frontal gyrus. See Table 1 in Supplementary Material and **Figure 2A**. As negative intensity increased, significant neural activation was noted in right amygdala, right occipital lobe, and right inferior frontal gyrus. See Table 1 in Supplementary Material and **Figures 2B,C**.

Neuroimaging analysis: reasoning time window

Neural activations associated with the reasoning time window are listed in Table 2 in Supplementary Material.

The contrast positive reasoning–positive baseline yielded neural activation in right thalamus, right occipital lobe, left parietal (supramarginal gyrus), right middle temporal gyrus, and right precentral gyrus. The contrast negative reasoning–negative baseline yielded neural activation in occipital lobe bilaterally, left inferior parietal lobe (supramarginal gyrus), left postcentral gyrus, left middle temporal gyrus, and left inferior frontal gyrus (triangularis). The contrast positive reasoning–neutral reasoning yielded activation in right inferior parietal (supramarginal gyrus). The contrast negative reasoning–neutral reasoning yielded neural activation in inferior occipital lobe bilaterally, left superior parietal lobe, left postcentral gyrus, right supramarginal gyrus, left inferior temporal and right middle temporal gyrus, left hippocampus, left middle frontal gyrus, and right frontal gyrus area 6. The contrast positive reasoning–negative reasoning yielded neural activation in left insula, right thalamus, superior temporal gyrus bilaterally, and right inferior frontal gyrus (orbitalis). The contrast negative reasoning–positive reasoning yielded significant neural activation in caudate nucleus bilaterally, left insula, occipital lobe bilaterally, left precuneus, and left postcentral gyrus.

To determine whether neural activation underlying reasoning in the positive and neutral time windows would differ after removing baseline effects, we analyzed the interaction contrast [(positive reasoning–positive baseline) – (neutral reasoning–neutral baseline)]; this analysis yielded neural activation in left middle cingulate, occipital lobe bilaterally, left inferior parietal lobe (angular gyrus), left intraparietal sulcus, right postcentral gyrus, left precentral gyrus, and right supplementary motor area.

To determine whether neural activation underlying reasoning in the negative and neutral time windows would differ after

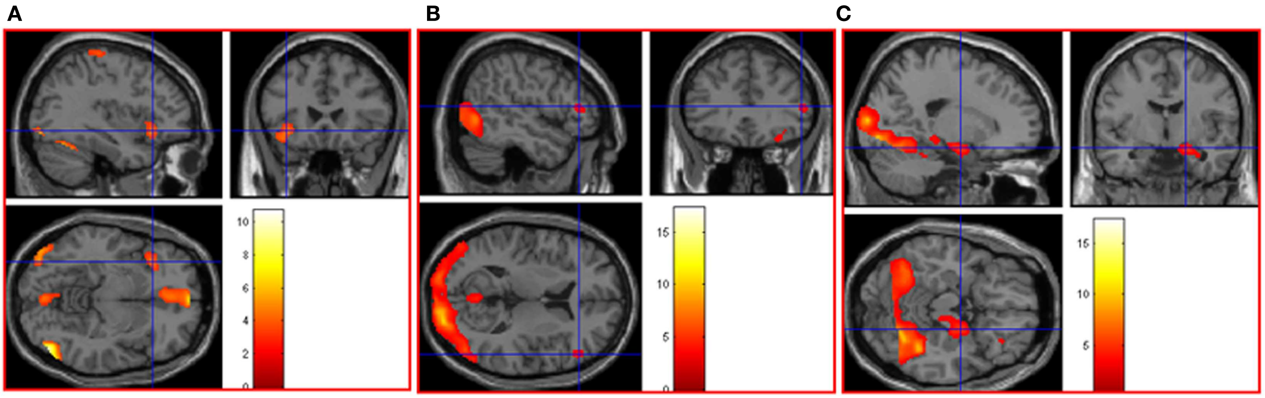


FIGURE 2 | (A) As picture ratings increase in positive intensity, activation increases in left inferior frontal gyrus (orbitalis) (MNI co-ordinates: −36, 24, −8, *k* = 310, *Z* = 3.54) and other areas (see Table 1 in Supplementary Material). As picture ratings increase in negative intensity, activation increases in **(B)** right inferior frontal gyrus (triangularis: area 45; MNI co-ordinates: 52, 32, 10, *k* = 57, *Z* = 3.31) and in **(C)** right amygdala (MNI co-ordinates: 20, −6, −16, *k* = 744, *Z* = 3.81), as well as other areas (see Table 1 in Supplementary Material).

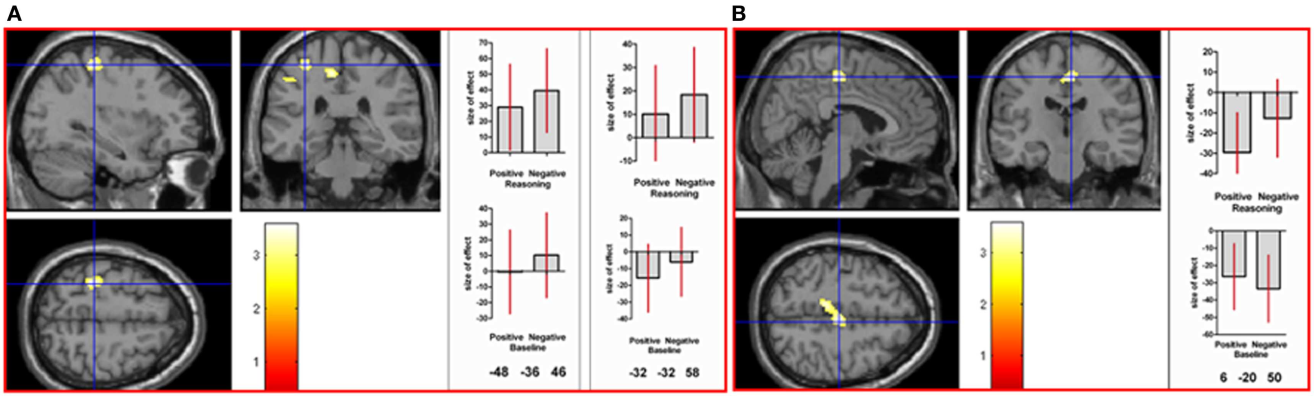


FIGURE 3 | A conjunction analysis demonstrated activation in common between the positive and negative reasoning time windows in (A) left postcentral gyrus (at the crosshair; MNI co-ordinates: −32, −32, 58, *k* = 122, *Z* = 3.43) and intraparietal sulcus (shown to the left of the crosshair in the coronal image; MNI co-ordinates: −48, −36, 46, *k* = 34, *Z* = 2.78), and in (B) right supplementary motor area (MNI co-ordinates: 6, −20, 50, *k* = 226, *Z* = 3.34), as well as other areas (see Table 2 in Supplementary Material). Graphs show size of effect (beta) with 5% confidence interval.

removing baseline effects, we analyzed the interaction contrast [(negative reasoning–negative baseline) − (neutral reasoning–neutral baseline)]; this analysis yielded neural activation in left superior parietal, inferior parietal lobe (angular gyrus) bilaterally, left inferior parietal (supramarginal gyrus), left postcentral gyrus, left inferior frontal gyrus (triangularis), and right supplementary motor area.

The interaction contrast [(neutral reasoning–neutral baseline) − (positive reasoning–positive baseline)] yielded neural activation in right fusiform gyrus. The interaction contrast [(neutral reasoning–neutral baseline) − (negative reasoning–negative baseline)] yielded neural activation in right hippocampus.

To determine areas activated in common in the positive and negative reasoning time window, we performed a conjunction analysis of two interaction contrasts: [(positive reasoning–positive baseline) − (neutral reasoning–neutral baseline)] and [(negative reasoning–negative baseline) − (neutral reasoning–neutral baseline)]. This conjunction analysis revealed neural activation in left superior parietal lobe, left inferior parietal lobe (angular gyrus, intraparietal sulcus, and supramarginal gyrus), left postcentral gyrus, and right supplementary motor area (see **Figure 3**).

To directly compare neural activations in the positive and negative reasoning time window, we conducted two interaction contrasts as follows. The interaction contrast [(positive reasoning–positive baseline) − (negative reasoning–negative baseline)] yielded neural activation in cerebellum (vermis), right superior parietal lobe, left fusiform gyrus, and right inferior frontal gyrus (orbitalis) (see **Figure 4**). The interaction contrast [(negative reasoning–negative baseline) − (positive reasoning–positive baseline)] yielded neural activation in left caudate nucleus, left

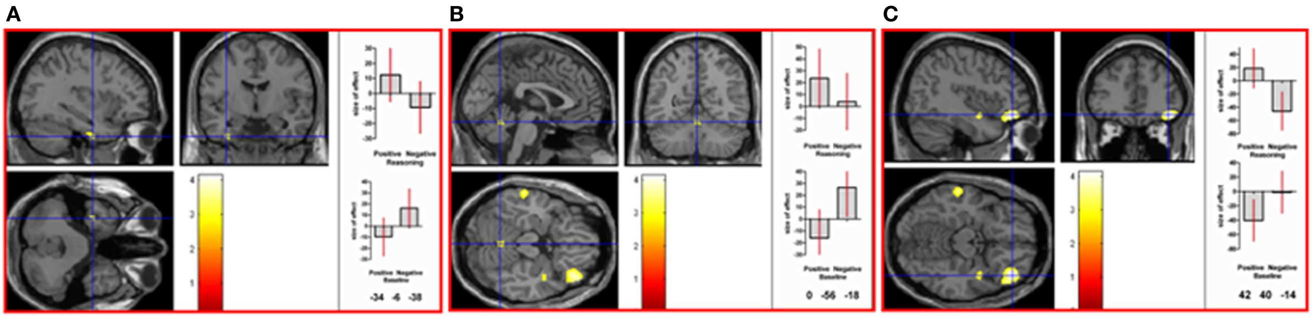


FIGURE 4 | Neural activation associated with the positive reasoning time window that is not shared with the negative reasoning time window occurs in (A) left fusiform gyrus (MNI co-ordinates: −34, −6, −38, *k* = 28, *Z* = 3.06), in (B) the vermis of the cerebellum (MNI co-ordinates: 0, −56, −18, *k* = 35, *Z* = 2.9), in (C) right inferior frontal gyrus (orbitalis; MNI co-ordinates: 42, 40, −14, *k* = 428, *Z* = 3.91), and in right superior parietal lobe (not shown) (see Table 2 in Supplementary Material). Graphs show size of effect (beta) with 5% confidence interval.

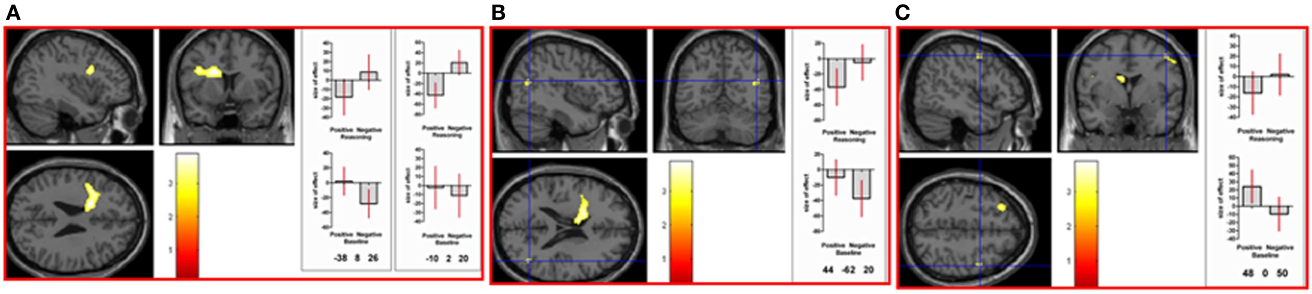


FIGURE 5 | Neural activation associated with the negative reasoning time window that is not shared with the positive reasoning time window occurs in (A) left caudate nucleus (MNI co-ordinates: −10, 2, 20, *k* = 594, *Z* = 3.39) extending into left inferior frontal gyrus (opercularis; MNI co-ordinates: −38, −8, 26, *Z* = 3.35), in (B) right middle temporal gyrus (relative deactivation; MNI co-ordinates: 44, −62, 20, *k* = 39, *Z* = 2.86), in (C) right precentral gyrus (area 6; MNI co-ordinates: 48, 0, 50, *k* = 38, *Z* = 2.85), as well as in left occipital lobe (not shown) (See Table 2 in Supplementary Material). Graphs show size of effect (beta) with 5% confidence interval.

occipital lobe, left inferior frontal gyrus (opercularis), and right precentral gyrus, as well as relative deactivation in right middle temporal gyrus (see **Figure 5**).

DISCUSSION

The above-chance reasoning accuracy levels indicate that participants were engaged in the task. The emotion manipulations were also successful, as indicated by the variation in participants' ratings of picture valence.

EMOTION INDUCTION

Patterns of neural responses during picture viewing/rating were consistent with those reported in the literature. As positive intensity increased, activation was noted in the left inferior frontal cortex. Likewise, Dolcos et al. (2004) reported neural activation in frontal cortex, left hemisphere only, in association with the rating of positive pictures. Furthermore, there is a trend in the neuroimaging literature (Wager et al., 2003) for left-lateralization in the frontal lobe associated with approach-related emotions[1].

During negative picture viewing/rating, activations in the contrast (negative picture–neutral picture) included right amygdala and right inferior frontal gyrus. Activations in the contrast (negative picture–positive picture) included left amygdala and inferior frontal gyrus bilaterally. As negative intensity increased, activations were in right occipital, right amygdala, and right inferior frontal gyrus. In Dolcos et al. (2004), rating of negative pictures was associated with neural activation in bilateral frontal regions. In Taylor et al. (2000), ratings of aversiveness of negative pictures were associated with neural activation in amygdala, uncus, and anterior parahippocampus. Neuroimaging studies of emotion perception (including studies using the IAPS) often report activation in amygdala, parahippocampal cortex, pregenual anterior cingulate, dorsal inferior frontal gyrus, inferior temporal and occipital cortex, and lateral cerebellum (Wager et al., 2008); withdrawal-related emotions[2] are generally correlated with bilateral frontal activation (Murphy et al., 2003) and with amygdala activation (Wager et al., 2003).

[1] Approach emotions include anger but are otherwise positive; none of our stimuli were designed to induce anger.

[2] Withdrawal emotions are negative in valence.

REASONING

Based on existing literature, we had hypothesized that both positive and negative emotion would be detrimental to subsequent reasoning. We did not find a significant difference in either reasoning accuracy or mean reaction time among the positive, neutral, and negative conditions. The Congruence*Emotion manipulation check indicated that reasoning was impaired when beliefs and logic were incongruent; however, we did not have the power to explore this at the neural level, because of design choices we made at the outset. Further study of this issue may be warranted (see Supplementary Material).

There have been other studies showing that emotion does not necessarily impair reasoning. Specifically, negative emotions have not invariably been associated in the literature with impaired reasoning. Goel and Vartanian (2011) conducted a behavioral study in which they manipulated the conflict between argument logic and beliefs about the conclusion by introducing politically incorrect material; on incongruent trials (a valid argument with an unbelievable conclusion, or an invalid argument with a believable conclusion), reasoning performance was better when the statement was politically incorrect than when otherwise. Blanchette et al. (2007) found that reasoning in the negative condition (compared to neutral) improved only when the reasoning material was related to participants' actual exposure to terrorist activity, whereas reasoning about other negative material was impaired.

Blanchette and Leese (2011) found no relation between reasoning performance and participant ratings of the intensity of negative and neutral stimuli. It is intriguing to note a similarity between their study and ours; Blanchette and Leese's study may be the first to link deductive reasoning with physiological arousal (measured with transient skin conductance response) underlying negative emotion induction, and ours may be the first study using pictures from the IAPS to link deductive reasoning with neural activation (measured using fMRI) underlying positive and negative emotion induction. Blanchette and Leese found no relation between reasoning performance and participant ratings of the intensity of negative and neutral stimuli, whereas our study found no effect on reasoning performance of positive or negative emotion induction in a design that included participant ratings.

Our main interest, reflected in our hypotheses, was to show that the neural systems underlying reasoning in each of the positive and negative conditions would differ from those in the neutral condition. These hypotheses were supported.

First, results indicated a crossover interaction, or double dissociation, between the positive and neutral reasoning time windows at the neural level. Not only did the interaction contrast [(positive reasoning–positive baseline) − (neutral reasoning–neutral baseline)] reveal activations but so also did the reverse interaction contrast [(neutral reasoning–neutral baseline) − (positive reasoning–positive baseline)]. Thus, although reasoning after positive emotion induction is not impaired, it is implemented at the neural level differently than is neutral reasoning. The neural pattern associated with the positive reasoning time window involves increased activation in left middle cingulate, occipital lobes bilaterally, left inferior parietal (angular gyrus), left intraparietal sulcus, right postcentral gyrus, left precentral gyrus, and right supplementary motor area.

A double dissociation indicates those neural regions implicated in condition A but not in condition B, and simultaneously, those neural regions implicated in condition B but not in condition A. Therefore, it indicates that conditions A and B involve separable systems.

Activation in the left inferior parietal lobe has been associated with abstract reasoning (Goel et al., 2000; Goel, 2009; Kuo et al., 2009; Watson and Chatterjee, 2012). Activation in the left angular gyrus has been associated with semantic meaning (Seghier et al., 2010; Sharp et al., 2010), more so when there is a conflict involving implausible sentences (Ye and Zhou, 2009) or when the stimulus is emotional (Hervé et al., 2012); it is implicated also in problem identification (Dandan et al., 2013b), in problem solving (Dandan et al., 2013a; Grabner et al., 2013), and in cognitive flexibility (Jacobson et al., 2011). Activation in intraparietal sulcus has been associated with item-specific processing but not with relations among items (Ackerman and Courtney, 2012), with symbolic number processing (Bugden et al., 2012), with attention to items presented in the periphery (Gillebert et al., 2013), and with temporal orienting (that is, attention toward a specific moment in time; Davranche et al., 2011). Left frontal precentral gyrus has been associated with the interaction of attention and language comprehension (Kristensen et al., 2013), with syntax complexity and *post hoc* reanalysis of sentence comprehension (Meltzer et al., 2010), and with successful inhibitory control (Padmala and Pessoa, 2010). Activation in postcentral gyrus has been associated with the illusory perception of motion (Planetta and Servos, 2012), and with visceral stimulation (Hojo et al., 2012; Kaplan and Meyer, 2012). The right frontal supplementary motor area has been associated with speeded decision-making (Wenzlaff et al., 2011), with attention maintenance (Kristensen et al., 2013), and is considered to be part of a ventral attention network that mediates bottom-up capture of attention by memory (Burianová et al., 2012).

Secondly, results indicated a crossover interaction, or double dissociation, between the negative and neutral reasoning time windows at the neural level. Not only did the interaction contrast [(negative reasoning–negative baseline) − (neutral reasoning–neutral baseline)] reveal activations but so also did the reverse interaction contrast [(neutral reasoning–neutral baseline) − (negative reasoning–negative baseline)]. Thus, although reasoning after negative emotion induction is not impaired, it is implemented at the neural level differently than is neutral reasoning. The neural pattern associated with the negative reasoning time window involves left postcentral gyrus, left inferior parietal (supramarginal gyrus), left superior parietal lobe, inferior parietal (angular gyrus) bilaterally, left inferior frontal gyrus, and right supplementary motor area.

As mentioned above, activation in postcentral gyrus has been associated with the illusory perception of motion and with visceral stimulation. Left supramarginal gyrus is considered to be part of a ventral attention network (Corbetta et al., 2008) that mediates bottom-up capture of attention by memory (Burianová et al., 2012). Superior parietal lobe is involved in the interaction between language processing and the control of movement (Segal and Petrides, 2012); activation has been associated with syllogistic reasoning involving abstract or incongruent materials (Tsujii et al., 2011). As mentioned above, activation in the left inferior parietal lobe has been associated with abstract reasoning; activation in the left angular gyrus has been associated with semantic meaning, more so when there is a conflict involving implausible sentences or when the stimulus is emotional, with problem identification and problem solving, and with cognitive flexibility. Activation in the left inferior frontal region has been associated with semantic integration (Yu et al., 2011; Huang et al., 2012) and with categorization (Lupyan et al., 2012; Philipp et al., 2013). As mentioned above, activation in the right supplementary motor area has been associated with speeded decision-making and with attention maintenance, and is considered to be part of a ventral attention network that mediates bottom-up capture of attention by memory.

The positive and negative reasoning time windows yielded similar activation in left superior parietal, left inferior parietal (angular gyrus, intraparietal sulcus, and supramarginal gyrus), left postcentral gyrus, and right supplementary motor area. This finding emerged from a conjunction analysis of two interaction contrasts: [(positive reasoning–positive baseline) − (neutral reasoning–neutral baseline)] and [(negative reasoning–negative baseline) − (neutral reasoning–neutral baseline)].

Beyond these similarities, however, results indicated a crossover interaction, or double dissociation, between the positive and negative reasoning time windows at the neural level. Not only did the interaction contrast [(positive reasoning–positive baseline) − (negative reasoning–negative baseline)] reveal activations but so also did the reverse interaction contrast [(negative reasoning–negative baseline) − (positive reasoning–positive baseline)].

The interaction favoring the positive reasoning time window revealed activation in right inferior frontal (orbitalis, or BA 47), right superior parietal, cerebellar vermis, and left fusiform. In the literature, activation in right frontal (BA 47) has been noted in unconstrained hypothesis generation (Vartanian and Goel, 2005). As mentioned above, superior parietal lobe is involved in the interaction between language processing and the control of movement. The cerebellar vermis is involved in autonomic and motor responses to an emotional state (Strata et al., 2011). Activation in left fusiform has been involved in lexico-semantic processing (Tsapkini and Rapp, 2010; Thesen et al., 2012).

The interaction favoring the negative reasoning time window revealed activation in left caudate nucleus, left inferior frontal (opercularis, or BA 44), left occipital lobe, and right precentral gyrus, as well as relative deactivation in right middle temporal gyrus. In the literature, caudate nucleus has been shown to have a crucial role in reasoning (Melrose et al., 2007) unless insufficient processing time has been allotted for reasoning (Kalbfleisch et al., 2007). Activation in left inferior frontal (BA 44) is associated more with phonological than with semantic fluency (Katzev et al., 2013). Right precentral gyrus is implicated in the representation of coordinated hand–mouth movements (Desmurget et al., 2014) and the neural coding of oculomotor and somatomotor space (Iacoboni et al., 1997). Activation in right middle temporal lobe has been associated with verbal fluency (Krug et al., 2011) and with semantic priming (Laufer et al., 2011).

Goel and Dolan (2003b) had manipulated emotion using the content of the syllogism such that content was either emotionally provocative or neutral; they found that reasoning with negatively charged material was associated with activation in ventromedial prefrontal cortex, whereas reasoning with neutral material was associated with activation in left dorsolateral prefrontal cortex. We have extended their findings by manipulating emotion separately from the material itself. Our emotion manipulation provides an emotional context in which to reason about neutral material, rather than providing emotional content. Therefore, it is not surprising that our findings differ from those in Goel and Dolan (2003b). Reasoning in an emotional but unrelated context involves a different neural underpinning than does reasoning about emotional content.

The fact that we found neural level differences in reasoning, despite a lack of behavioral difference, suggests that the neural systems underlying reasoning are sensitive to neural systems previously recruited by emotional context, and can to some extent compensate for these effects of emotions. It is possible that the behavioral manifestations (that is, impairment of reasoning) emerge only when the system is stressed.

In summary, we had predicted that both positive and negative emotion would be detrimental to reasoning, and that the neural systems underlying reasoning under those two conditions would differ from that in the neutral condition. We found that, although neither positive nor negative emotional context significantly impaired reasoning performance, positive and negative context did have dissociable effects on the underlying neural mechanisms involved in reasoning.

ACKNOWLEDGMENTS

This study was funded by a Wellcome Trust Grant (ABH00 FA032YBH064) to Vinod Goel.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at http://www.frontiersin.org/Journal/10.3389/fnhum.2014.00736/abstract

REFERENCES

Ackerman, C. M., and Courtney, S. M. (2012). Spatial relations and spatial locations are dissociated within prefrontal and parietal cortex. *J. Neurophysiol.* 108, 2419–2429. doi:10.1152/jn.01024.2011

Alloy, L. B., and Abramson, L. Y. (1979). Judgment of contingency in depressed and nondepressed students: sadder but wiser? *J. Exp. Psychol. Gen.* 108, 441–485. doi:10.1037/0096-3445.108.4.441

Ashburner, J., and Friston, K. J. (1999). Nonlinear spatial normalization using basis functions. *Hum. Brain Mapp.* 7, 254–266. doi:10.1002/(SICI)1097-0193(1999)7:4<254::AID-HBM4>3.0.CO;2-G

Biss, R. K., and Hasher, L. (2011). Delighted and distracted: positive affect increases priming for irrelevant information. *Emotion* 11, 1474–1478. doi:10.1037/a0023855

Blanchette, I. (2006). The effect of emotion on interpretation and logic in a conditional reasoning task. *Mem. Cognit.* 34, 1112–1125. doi:10.3758/BF03193257

Blanchette, I., and Leese, J. (2011). The effect of negative emotion on deductive reasoning: examining the contribution of physiological arousal. *Exp. Psychol.* 58, 235–246. doi:10.1027/1618-3169/a000090

Blanchette, I., and Richards, A. (2004). Reasoning about emotional and neutral materials: is logic affected by emotion? *Psychol. Sci.* 15, 745–752. doi:10.1111/j.0956-7976.2004.00751.x

Blanchette, I., and Richards, A. (2010). The influence of affect on higher level cognition: a review of research on interpretation, judgement, decision making and reasoning. *Cogn. Emot.* 24, 561–595. doi:10.1080/02699930903132496

Blanchette, I., Richards, A., Melnyk, L., and Lavda, A. (2007). Reasoning about emotional contents following shocking terrorist attacks: a tale of three cities. *J. Exp. Psychol. Appl.* 13, 47–56. doi:10.1037/1076-898X.13.1.47

Bless, H., Mackie, D. M., and Schwarz, N. (1992). Mood effects on attitude judgments: independent effects of mood before and after message elaboration. *J. Pers. Soc. Psychol.* 63, 585–595. doi:10.1037/0022-3514.63.4.585

Bodenhausen, G. V., Kramer, G. P., and Susser, K. (1994). Happiness and stereotypic thinking in social judgment. *J. Pers. Soc. Psychol.* 66, 621–632. doi:10.1037/0022-3514.66.4.621

Bohner, G., Crow, K., Erb, H.-P., and Schwarz, N. (1992). Affect and persuasion: mood effects on the processing of message content and context cues and on subsequent behaviour. *Eur. J. Soc. Psychol.* 22, 511–530. doi:10.1002/ejsp.2420220602

Bugden, S., Price, G. R., McLean, D. A., and Ansari, D. (2012). The role of the left intraparietal sulcus in the relationship between symbolic number processing and children's arithmetic competence. *Dev. Cogn. Neurosci.* 2, 448–457. doi:10.1016/j.dcn.2012.04.001

Burianová, H., Ciaramelli, E., Grady, C. L., and Moscovitch, M. (2012). Top-down and bottom-up attention-to-memory: mapping functional connectivity in two distinct networks that underlie cued and uncued recognition memory. *Neuroimage* 63, 1343–1352. doi:10.1016/j.neuroimage.2012.07.057

Corbetta, M., Patel, G., and Shulman, G. L. (2008). The reorienting system of the human brain: from environment to theory of mind. *Neuron* 58, 306–324. doi:10.1016/j.neuron.2008.04.017

Dandan, T., Haixue, Z., Wenfu, L., Wenjing, Y., Jiang, Q., and Qinglin, Z. (2013a). Brain activity in using heuristic prototype to solve insightful problems. *Behav. Brain Res.* 253, 139–144. doi:10.1016/j.bbr.2013.07.017

Dandan, T., Wenfu, L., Tianen, D., Nusbaum, H. C., Jiang, Q., and Qinglin, Z. (2013b). Brain mechanisms of valuable scientific problem finding inspired by heuristic knowledge. *Exp. Brain Res.* 228, 437–443. doi:10.1007/s00221-013-3575-4

Davranche, K., Nazarian, B., Vidal, F., and Coull, J. (2011). Orienting attention in time activates left intraparietal sulcus for both perceptual and motor task goals. *J. Cogn. Neurosci.* 23, 3318–3330. doi:10.1162/jocn_a_00030

Desmurget, M., Richard, N., Harquel, S., Baraduc, P., Szathmari, A., Mottolese, C., et al. (2014). Neural representations of ethologically relevant hand/mouth synergies in the human precentral gyrus. *Proc. Natl. Acad. Sci. U.S.A.* 111, 5718–5722. doi:10.1073/pnas.1321909111

Dolcos, F., LaBar, K. S., and Cabeza, R. (2004). Dissociable effects of arousal and valence on prefrontal activity indexing emotional evaluation and subsequent memory: an event-related fMRI study. *Neuroimage* 23, 64–74. doi:10.1016/j.neuroimage.2004.05.015

Edwards, J. A., and Weary, G. (1993). Depression and the impression-formation continuum: piecemeal processing despite the availability of category information. *J. Pers. Soc. Psychol.* 64, 636–645. doi:10.1037/0022-3514.64.4.636

Eickhoff, S., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, Z., et al. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 25, 1325–1335. doi:10.1016/j.neuroimage.2004.12.034

Evans, A. C., Collins, D. L., Mills, S. R., Brown, E. D., Kelly, R. L., and Peters, T. M. (1993). "3D statistical neuroanatomical models from 305 MRI volumes," in *IEEE Conference Record, Nuclear Science Symposium and Medical Imaging Conference*, Vol. 3 (San Francisco, CA: IEEE), 1813–1817.

Evans, J. S. (2003). In two minds: dual-process accounts of reasoning. *Trends Cogn. Sci.* 7, 454–459. doi:10.1016/j.tics.2003.08.012

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., and Frackowiak, R. S. J. (1994). Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* 2, 189–210. doi:10.1002/hbm.460020402

Genovese, C. R., Lazar, N. A., and Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage* 15, 870–878. doi:10.1006/nimg.2001.1037

Gillebert, C. R., Caspari, N., Wagemans, J., Peeters, R., Dupont, P., and Vandenberghe, R. (2013). Spatial stimulus configuration and attentional selection: extrastriate and superior parietal interactions. *Cereb. Cortex* 23, 2840–2854. doi:10.1093/cercor/bhs263

Goel, V. (2009). "Fractionating the system of deductive reasoning," in *Neural Correlates of Thinking*, eds E. Kraft, B. Guylas, and E. Poppel (New York, NY: Springer Press), 203–218.

Goel, V., Buchel, C., Frith, C., and Dolan, R. (2000). Dissociation of mechanisms underlying syllogistic reasoning. *Neuroimage* 12, 504–514. doi:10.1006/nimg.2000.0636

Goel, V., and Dolan, R. J. (2003a). Explaining modulation of reasoning by belief. *Cognition* 87, B11–B22. doi:10.1016/S0010-0277(02)00185-3

Goel, V., and Dolan, R. J. (2003b). Reciprocal neural response within lateral and ventral medial prefrontal cortex during hot and cold reasoning. *Neuroimage* 20, 2314–2341. doi:10.1016/j.neuroimage.2003.07.027

Goel, V., and Vartanian, O. (2011). Negative emotions can attenuate the influence of beliefs on logical reasoning. *Cogn. Emot.* 25, 121–131. doi:10.1080/02699931003593942

Grabner, R. H., Ansari, D., Koschutnig, K., Reishofer, G., and Ebner, F. (2013). The function of the left angular gyrus in mental arithmetic: evidence from the associative confusion effect. *Hum. Brain Mapp.* 34, 1013–1024. doi:10.1002/hbm.21489

Hervé, P. Y., Razafimandimby, A., Vigneau, M., Mazoyer, B., and Tzourio-Mazoyer, N. (2012). Disentangling the brain networks supporting affective speech comprehension. *Neuroimage* 61, 1255–1267. doi:10.1016/j.neuroimage.2012.03.073

Hojo, M., Takahashi, T., Nagahara, A., Sasaki, H., Oguro, M., Asaoka, D., et al. (2012). Analysis of brain activity during visceral stimulation. *J. Gastroenterol. Hepatol.* 27(Suppl. 3), 49–52. doi:10.1111/j.1440-1746.2012.07072.x

Huang, J., Zhu, Z., Zhang, J. X., Wu, M., Chen, H. C., and Wang, S. (2012). The role of left inferior frontal gyrus in explicit and implicit semantic processing. *Brain Res.* 1440, 56–64. doi:10.1016/j.brainres.2011.11.060

Iacoboni, M., Woods, R. P., Lenzi, G. L., and Mazziotta, J. C. (1997). Merging of oculomotor and somatomotor space coding in the human right precentral gyrus. *Brain* 120, 1635–1645. doi:10.1093/brain/120.9.1635

Isen, A. M., and Daubman, K. A. (1984). The influence of affect on categorization. *J. Pers. Soc. Psychol.* 47, 1206–1217. doi:10.1037/0022-3514.47.6.1206

Isen, A. M., Daubman, K. A., and Nowicki, G. P. (1987). Positive affect facilitates creative problem solving. *J. Pers. Soc. Psychol.* 52, 1122–1131. doi:10.1037/0022-3514.52.6.1122

Jacobson, S. C., Blanchard, M., Connolly, C. C., Cannon, M., and Garavan, H. (2011). An fMRI investigation of a novel analogue to the trail-making test. *Brain Cogn.* 77, 60–70. doi:10.1016/j.bandc.2011.06.001

Kalbfleisch, M. L., Van Meter, J. W., and Zeffiro, T. A. (2007). The influences of task difficulty and response correctness on neural systems supporting fluid reasoning. *Cogn. Neurodyn.* 1, 71–84. doi:10.1007/s11571-006-9007-4

Kaplan, J. T., and Meyer, K. (2012). Multivariate pattern analysis reveals common neural patterns across individuals during touch observation. *Neuroimage* 60, 204–212. doi:10.1016/j.neuroimage.2011.12.059

Katzev, M., Tüscher, O., Hennig, J., Weiller, C., and Kaller, C. P. (2013). Revisiting the functional specialization of left inferior frontal gyrus in phonological and semantic fluency: the crucial role of task demands and individual ability. *J. Neurosci.* 33, 7837–7845. doi:10.1523/JNEUROSCI.3147-12.2013

Kristensen, L. B., Wang, L., Petersson, K. M., and Hagoort, P. (2013). The interface between language and attention: prosodic focus marking recruits a general attention network in spoken language comprehension. *Cereb. Cortex* 23, 1836–1848. doi:10.1093/cercor/bhs164

Krug, A., Markov, V., Krach, S., Jansen, A., Zerres, K., Eggermann, T., et al. (2011). Genetic variation in G72 correlates with brain activation in the right middle temporal gyrus in a verbal fluency task. *Hum. Brain Mapp.* 32, 118–126. doi:10.1002/hbm.21005

Kuo, W.-J., Sjöström, T., Chen, Y.-P., Wang, Y.-H., and Huang, C.-Y. (2009). Intuition and deliberation: two systems for strategizing in the brain. *Science* 324, 519–522. doi:10.1126/science.1165598

Lang, P. J., Bradley, M. M., and Cuthberg, B. N. (1997). *International Affective Picture System [Pictures]*. Gainesville, FL: NIMH Center for the Study of Emotion and Attention.

Laufer, I., Negishi, M., Lacadie, C. M., Papademetris, X., and Constable, R. T. (2011). Dissociation between the activity of the right middle frontal gyrus and the middle temporal gyrus in processing semantic priming. *PLoS ONE* 6:e22368. doi:10.1371/journal.pone.0022368

Lefford, A. (1946). The influence of emotonal subject matter on logical reasoning. *J. Gen. Psychol.* 34, 127–151. doi:10.1080/00221309.1946.10544530

Lerner, J. S., Gonzalez, R. M., Small, D. A., and Fischhoff, B. (2003). Effects of fear and anger on perceived risks of terrorism: a national field experiment. *Psychol. Sci.* 14, 144–150. doi:10.1111/1467-9280.01433

Lerner, J. S., and Keltner, D. (2001). Fear, anger, and risk. *J. Pers. Soc. Psychol.* 81, 146–159. doi:10.1037//O022-3514.81.1.146

Lieberman, M. D., and Cunningham, W. A. (2009). Type I and type II error concerns in fMRI research: re-balancing the scale. *Soc. Cogn. Affect. Neurosci.* 4, 423–428. doi:10.1093/scan/nsp052

Lupyan, G., Mirman, D., Hamilton, R., and Thompson-Schill, S. L. (2012). Categorization is modulated by transcranial direct current stimulation over left prefrontal cortex. *Cognition* 124, 36–49. doi:10.1016/j.cognition.2012.04.002

Martin, E. A., and Kerns, J. G. (2011). The influence of positive mood on different aspects of cognitive control. *Cogn. Emot.* 25, 265–279. doi:10.1080/02699931.2010.491652

Melrose, R. J., Poulin, R. M., and Stern, C. E. (2007). An fMRI investigation of the role of the basal ganglia in reasoning. *Brain Res.* 1142, 146–158. doi:10.1016/j.brainres.2007.01.060

Melton, R. J. (1995). The role of positive affect in syllogism performance. *Pers. Soc. Psychol. Bull.* 21, 788–794. doi:10.1177/0146167295218001

Meltzer, J. A., McArdle, J. J., Schafer, R. J., and Braun, A. R. (2010). Neural aspects of sentence comprehension: syntactic complexity, reversibility, and reanalysis. *Cereb. Cortex* 20, 1853–1864. doi:10.1093/cercor/bhp249

Murphy, F. C., Nimmo-Smith, I., and Lawrence, A. D. (2003). Functional neuroanatomy of emotions: a meta-analysis. *Cogn. Affect. Behav. Neurosci.* 3, 207–233. doi:10.3758/CABN.3.3.207

Oaksford, M., Morris, F., Grainger, B., and Williams, J. M. G. (1996). Mood, reasoning, and central executive processes. *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 476–492. doi:10.1016/B978-0-444-53702-7.00007-5

Padmala, S., and Pessoa, L. (2010). Interactions between cognition and motivation during response inhibition. *Neuropsychologia* 48, 558–565. doi:10.1016/j.neuropsychologia.2009.10.017

Philipp, A. M., Weidner, R., Koch, I., and Fink, G. R. (2013). Differential roles of inferior frontal and inferior parietal cortex in task switching: evidence from stimulus-categorization switching and response-modality switching. *Hum. Brain Mapp.* 34, 1910–1920. doi:10.1002/hbm.22036

Planetta, P. J., and Servos, P. (2012). The postcentral gyrus shows sustained fMRI activation during the tactile motion aftereffect. *Exp. Brain Res.* 216, 535–544. doi:10.1007/s00221-011-2957-8

Schmitz, T. W., De Rosa, E., and Anderson, A. K. (2009). Opposing influences of affective state valence on visual cortical encoding. *J. Neurosci.* 29, 7199–7207. doi:10.1523/JNEUROSCI.5387-08.2009

Schwarz, N., and Bless, H. (1991). "Happy and mindless, but sad and smart? The impact of affective states on analytic reasoning," in *Emotion and Social Judgments*, ed. J. P. Forgas (Oxford: Pergamon Press), 55–71.

Schwarz, N., and Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: informative and directive functions of affective states. *J. Pers. Soc. Psychol.* 45, 513–523. doi:10.1037/0022-3514.45.3.513

Segal, E., and Petrides, M. (2012). The anterior superior parietal lobule and its interactions with language and motor areas during writing. *Eur. J. Neurosci.* 35, 309–322. doi:10.1111/j.1460-9568.2011.07937.x

Seghier, M. L., Fagan, E., and Price, C. J. (2010). Functional subdivisions in the left angular gyrus where the semantic system meets and diverges from the default network. *J. Neurosci.* 30, 16809–16817. doi:10.1523/JNEUROSCI.3377-10.2010

Sharp, D. J., Awad, M., Warren, J. E., Wise, R. J., Vigliocco, G., and Scott, S. K. (2010). The neural response to changing semantic and perceptual complexity during language processing. *Hum. Brain Mapp.* 31, 365–377. doi:10.1002/hbm.20871

Strata, P., Scelfo, B., and Sacchetti, B. (2011). Involvement of cerebellum in emotional behavior. *Physiol. Res.* 60(Suppl. 1), S39–S48.

Taylor, S. F., Liberzon, I., and Koeppe, R. A. (2000). The effect of graded aversive stimuli on limbic and visual activation. *Neuropsychologia* 38, 1415–1425. doi:10.1016/S0028-3932(00)00032-4

Thesen, T., McDonald, C. R., Carlson, C., Doyle, W., Cash, S., Sherfey, J., et al. (2012). Sequential then interactive processing of letters and words in the left fusiform gyrus. *Nat. Commun.* 3, 1284. doi:10.1038/ncomms2220

Tsapkini, K., and Rapp, B. (2010). The orthography-specific functions of the left fusiform gyrus: evidence of modality and category specificity. *Cortex* 46, 185–205. doi:10.1016/j.cortex.2009.02.025

Tsujii, T., Sakatani, K., Masuda, S., Akiyama, T., and Watanabe, S. (2011). Evaluating the roles of the inferior frontal gyrus and superior parietal lobule in deductive reasoning: an rTMS study. *Neuroimage* 58, 640–646. doi:10.1016/j.neuroimage.2011.06.076

Vanlessen, N., Rossi, V., De Raedt, R., and Pourtois, G. (2013). Positive emotion broadens attention focus through decreased position-specific spatial encoding in early visual cortex: evidence from ERPs. *Cogn. Affect. Behav. Neurosci.* 13, 60–79. doi:10.3758/s13415-012-0130-x

Vartanian, O., and Goel, V. (2005). Task constraints modulate activation in right ventral lateral prefrontal cortex. *Neuroimage* 27, 927–933. doi:10.1016/j.neuroimage.2005.05.016

Wager, T. D., Barrett, L. F., Bliss-Moreau, E., Lindquist, K. A., Duncan, S., Kober, H., et al. (2008). "The neuroimaging of emotion," in *Handbook of Emotions*, eds M. Lewis, J. M. Haviland-Jones, and L. F. Barrett (New York, NY: Guilford Press), 249–271.

Wager, T. D., Phan, K. L., Liberzon, I., and Taylor, S. F. (2003). Valence, gender, and lateralization of functional brain anatomy in emotion: a meta-analysis of findings from neuroimaging. *Neuroimage* 19, 513–531. doi:10.1016/S1053-8119(03)00078-8

Watson, C. E., and Chatterjee, A. (2012). A bilateral frontoparietal network underlies visuospatial analogical reasoning. *Neuroimage* 59, 2831–2838. doi:10.1016/j.neuroimage.2011.09.030

Wenzlaff, H., Bauer, M., Maess, B., and Heekeren, H. R. (2011). Neural characterization of the speed-accuracy tradeoff in a perceptual decision-making task. *J. Neurosci.* 31, 1254–1266. doi:10.1523/JNEUROSCI.4000-10.2011

Worsley, K. J., and Friston, K. J. (1995). Analysis of fMRI time-series revisited – again. *Neuroimage* 2, 173–181. doi:10.1006/nimg.1995.1023

Ye, Z., and Zhou, X. (2009). Conflict control during sentence comprehension: fMRI evidence. *Neuroimage* 48, 280–290. doi:10.1016/j.neuroimage.2009.06.032

Yu, T., Lang, S., Birbaumer, N., and Kotchoubey, B. (2011). Listening to factually incorrect sentences activates classical language areas and thalamus. *Neuroreport* 22, 865–869. doi:10.1097/WNR.0b013e32834b6fc6

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

Received: 16 May 2014; accepted: 02 September 2014; published online: 23 September 2014.
Citation: Smith KW, Vartanian O and Goel V (2014) Dissociable neural systems underwrite logical reasoning in the context of induced emotions with positive and negative valence. Front. Hum. Neurosci. ***8****:736. doi: 10.3389/fnhum.2014.00736*
This article was submitted to the journal Frontiers in Human Neuroscience.

ORIGINAL RESEARCH ARTICLE
published: 10 February 2015
doi: 10.3389/fnhum.2015.00061

Analyzing the association between functional connectivity of the brain and intellectual performance

Gustavo S. P. Pamplona[1]*, Gérson S. Santos Neto[2], Sara R. E. Rosset[2], Baxter P. Rogers[3] and Carlos E. G. Salmon[1]

[1] *InBrain Lab, Department of Physics, Faculty of Philosophy, Sciences and Letters of Ribeirão Preto, University of São Paulo, São Paulo, Brazil*
[2] *Faculty of Medicine of Ribeirão Preto, University of São Paulo, São Paulo, Brazil*
[3] *Department of Radiology and Radiological Sciences, Department of Biomedical Engineering, Institute of Imaging Science, Vanderbilt University, Nashville, TN, USA*

Edited by:
Gorka Navarrete, Universidad Diego Portales, Chile

Reviewed by:
Lucas Sedeño, Institute of Cognitive Neurology (INECO), Argentina
Roberto Colom, Universidad Autonoma de Madrid, Spain

****Correspondence:***
Gustavo S. P. Pamplona, InBrain Lab, Department of Physics, Faculty of Philosophy, Sciences and Letters of Ribeirão Preto, University of São Paulo, Av. Bandeirantes 3900, Ribeirão Preto, São Paulo, 14040-900, Brazil
e-mail: gustavopamplona@usp.br

Measurements of functional connectivity support the hypothesis that the brain is composed of distinct networks with anatomically separated nodes but common functionality. A few studies have suggested that intellectual performance may be associated with greater functional connectivity in the fronto-parietal network and enhanced global efficiency. In this fMRI study, we performed an exploratory analysis of the relationship between the brain's functional connectivity and intelligence scores derived from the Portuguese language version of the Wechsler Adult Intelligence Scale (WAIS-III) in a sample of 29 people, born and raised in Brazil. We examined functional connectivity between 82 regions, including graph theoretic properties of the overall network. Some previous findings were extended to the Portuguese-speaking population, specifically the presence of small-world organization of the brain and relationships of intelligence with connectivity of frontal, pre-central, parietal, occipital, fusiform and supramarginal gyrus, and caudate nucleus. Verbal comprehension was associated with global network efficiency, a new finding.

Keywords: functional connectivity, fMRI, network parameters, intelligence, Wechsler intelligence scales, exploratory data analysis

INTRODUCTION

Functional connectivity is expressed as correlations between the blood oxygenation level dependent signals in different regions of the brain (Friston et al., 1993; Biswal et al., 1995; Van den Heuvel and Hulshoff Pol, 2010). Consistent spatial patterns of functional connectivity are found for individuals at rest and are presumed to reflect information processing networks (Lowe et al., 1998; Raichle et al., 2001; Beckmann et al., 2005; Damoiseaux et al., 2006). Recent advances in neuroimaging have provided new tools to measure and analyze interactions between brain regions, catalyzing the study of functional connectivity of the brain (Van den Heuvel and Hulshoff Pol, 2010). An important recent expansion of functional connectivity studies was the use of the principles of graph theory (Watts and Strogatz, 1998) to depict the brain as an efficient complex network, with brain regions as the nodes and functional connectivity as the edge weights (Sporns and Zwi, 2004; Bullmore and Sporns, 2009). The functional brain network shows a highly efficient small-world organization, with a high level of local clustering and short effective lengths between brain regions. This leads to high global efficiency of information flow in the network (Sporns and Zwi, 2004; Van den Heuvel et al., 2008).

An important tool to measure the intelligence in adults is the Wechsler Adult Intelligence Scale (WAIS), based on the "global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment" (Wechsler, 1939). Some studies have applied intelligence indices to anatomical and functional brain measurements (Gray et al., 2003; Haier et al., 2004; Song et al., 2008; Gläscher et al., 2009; Li et al., 2009). A previous study found that higher IQ scores are associated with greater functional connectivity within a fronto-parietal network, suggesting that the coordination of these regions is an important neural basis of individual intelligence (Song et al., 2008). A region-specific analysis of the lateral prefrontal cortex, part of the fronto-parietal network, found that its global connectivity predicted working memory performance and fluid intelligence (Cole et al., 2012). Two studies have reported an association between efficiency of global communication and intellectual performance, suggesting that individuals with higher intelligence have a more organized brain network overall (Van den Heuvel et al., 2008; Song et al., 2009).

However, the relationships between brain functional connectivity and psychological measures such as intelligence are not fully defined. In the present exploratory study, we pursued this line of research further by considering how the several indices of intelligence measured by the Wechsler Adult Intelligence Scale (WAIS-III) related to connection strengths and network properties in a brain network defined by a set of 82 a priori cortical and subcortical regions derived from an atlas (Tzourio-Mazoyer et al., 2002). The use of a smaller set of regions of interest preserves structural and physiological similarities, while simplifying the analysis and easing the interpretation of the findings relative to the commonly used voxel-wise approach. In contrast to some studies that considered a priori regions known to be related to intelligence (Song et al., 2008; Cole et al., 2012), the present study explored the brain as a whole, with no region-specific or network-specific hypotheses. This analysis could help to elucidate how the human brain supports particular intellectual processes, extending previous work and providing background to future studies.

MATERIALS AND METHODS

PARTICIPANTS

Thirty one healthy people were recruited from the academic community and the local population living in the state of São Paulo, Brazil. They were right-handed, had no history of neurological or psychological illnesses, and were native speakers of Brazilian Portuguese. People with a range of educational levels were recruited to provide a greater range of intelligence scores (**Table 1**). Thirty of these participants made up Dataset 1. Volunteers participated in this study after responding to the standard screening interview of the Hospital of Clinics in Ribeirão Preto, and providing written consent as approved by the Research Ethics Committee of University of São Paulo.

MEASURES OF INDIVIDUAL INTELLIGENCE

The level of intellectual performance was measured (Gérson S. Santos Neto and Sara R. E. Rosset) using the WAIS III test (Wechsler Adult Intelligence Scale) as modified for the Portuguese-speaking population of Brazil (Nascimento, 1998). WAIS-III is a widely used instrument that assesses several cognitive domains contributing to intelligence. It has high test-retest reliability and a large database for comparison and standardization (Gläscher et al., 2009). Measurements originating from the third version of the test are the four fundamental indices Verbal Comprehension Index, Perceptual Organization Index, Working Memory Index, and Processing Speed Index; and the overall score, Full-Scale IQ. The test took 1 h 30 min on average and was given at a separate time from the image acquisition (less than 2 months apart, except for one participant with a 3-month difference).

DATA ACQUISITION

Resting-state functional magnetic resonance images (eyes open, no fixation) from each participant were acquired in a Phillips 3 Tesla scanner with a Quasar Dual gradient system (80 mT/m, 200 mT/m/ms), using an eight channel head coil and SENSE encoding. An EPI sequence was performed with the following parameters: 2000 ms repetition time, 30 ms echo time, 240 × 240 mm field of view, 3 × 3 mm in-plane voxel size, 4.0 mm slice thickness, 0.5 mm slice gap, 32 slices, 80° flip angle, 200 volumes, 25.2 Hz bandwidth per pixel. Overall functional acquisition time was 6:48, including four initial volumes that were discarded prior to analysis.

High-resolution anatomical images were also acquired using a 3D T1 weighted turbo-field-echo gradient sequence with the following parameters: 2500 ms repetition time, 3.2 ms echo time, 7.0 ms time echo spacing, 900 ms inversion time, 1 mm isotropic voxel size, 8° flip angle, 240 × 240 × 160 mm^3 field of view, and overall time 5:19. Diffusion and other functional images were also acquired, but not used in the present analysis.

A separate set of resting-state functional magnetic resonance images (open eyes, with fixation) from 30 subjects (13M/17F, age: 26.5 ± 5.5, age range: 20–42, right-handed) was included in the analysis to provide a baseline for the small-worldness measurement, and classified as Dataset 2. These images were from the 1000 Functional Connectomes Project (Biswal et al., 2010), specifically the data acquired in Leipzig, Germany, in a 3 Tesla scanner with the following parameters: 2300 ms repetition time, 34 slices, 195 volumes.

PRE-PROCESSING

Functional MRI data were processed using the SPM8 software (http://www.fil.ion.ucl.ac.uk/spm/software/spm8) and the CONN functional connectivity toolbox (14), both implemented in MatLab (R2013a, The MathWorks, Natick, MA, USA). For each individual's functional images, rigid body movement was measured and corrected using a two-step procedure in which the first of the specified functional images was used as a reference to which all subsequent images were realigned, then the functional images were re-registered to the mean image. Participants who moved more than 2 mm in translation or 1 degree in rotation were excluded from analysis. Functional images were then spatially smoothed using a Gaussian filter of 5 mm full width at half maximum.

Anatomical images from each volunteer were registered to the mean functional image created in the previous step. The anatomical volumes were segmented into gray matter, white matter and cerebrospinal fluid compartments and non-linearly registered to the MNI standard space. The resulting masks were eroded once at an isotropic voxel size of 2 mm to minimize partial volume effects. This step produced spatial normalization parameters that were used to apply the transformations to the functional images.

Voxel time series were additionally processed to reduce noise. Signals from the white matter and CSF compartments (5 principal components each) and the estimated head motion time series and first differences were removed by regression. A temporal band-pass filter was applied to remove signals outside the range 0.008–0.09 Hz (Whitfield-Gabrieli and Nieto-Castanon, 2012).

Average signals were extracted from a set of 116 regions defined by the Automated Anatomical Labeling (AAL) atlas, which is a macroanatomical parcellation of the single subject MNI-space template brain (Tzourio-Mazoyer et al., 2002). Eight of the AAL regions were excluded from the analysis due to their small size (less than 300 voxels), which increased the likelihood that partial volume effects would contaminate signals from those regions. Cerebellum and cerebellar vermis regions were also excluded because they were not fully covered by the fMRI. Therefore, 82 cortical and subcortical regions were included in total, all of them shown in the Supplemental Material (Table S1) with their AAL abbreviations and the locations of their centers, in x, y, and z.

ANALYSIS OF FUNCTIONAL CONNECTIVITY AND INTELLIGENCE

Weighted association matrices were created (**Figure 1**) using the Pearson correlations between the time series of each pair of brain regions. Functional connectivity of each path was compared with the four fundamental intelligence indices and the Full-Scale IQ using the Pearson correlation coefficient (**Table 3**, **Figure 2**). Negative values of the matrices were included to consider also the functional anticorrelations. Functional connectivity values were

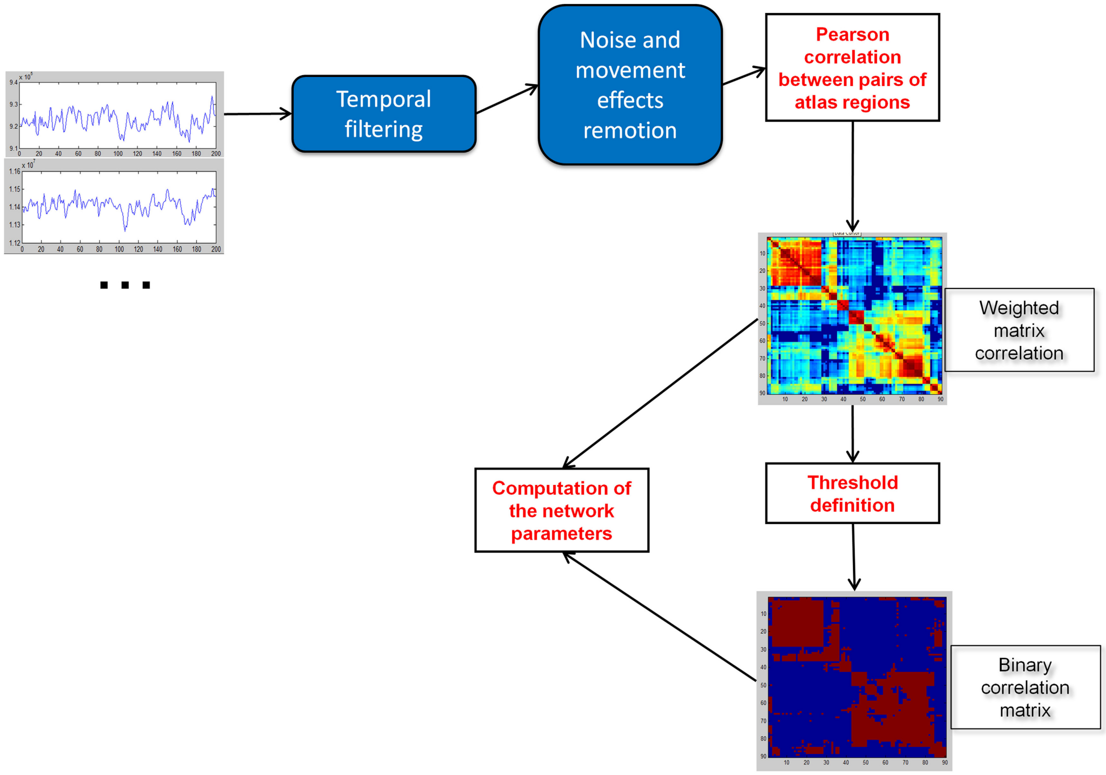

Temporal filtering
Noise and movement effects remotion
Pearson correlation between pairs of atlas regions
Weighted matrix correlation
Threshold definition
Binary correlation matrix
Computation of the network parameters
...

FIGURE 1 | Construction of weighted and binary correlation matrices of the brain.

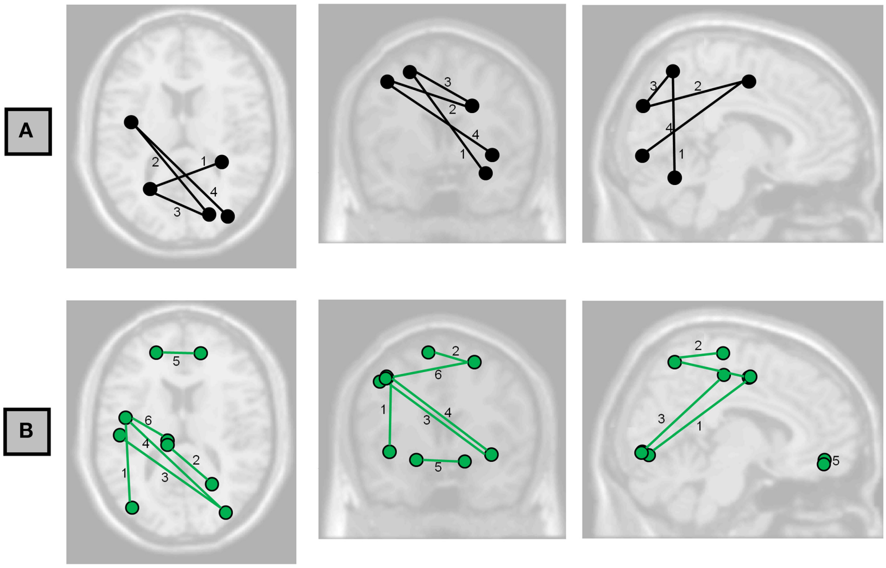


FIGURE 2 | Axial, coronal, and sagittal projections of the brain showing the functional connections having associations with (A) Full-Scale IQ and (B) Perceptual Organization Index at FDR < 0.05. Numbers correspond to the labels in **Table 3**.

the Fisher Z scores computed between the time series of each pair of regions. Each list of 3321 *p*-values (all pairs of 82 regions) was adjusted to maintain a false discovery rate of 0.05, separately for each IQ index.

GRAPH ANALYSIS

We examined small-worldness, characteristic path length, clustering coefficient, and global and local efficiency. Characteristic path length is the shortest path length between all pairs of nodes.

Clustering coefficient is the number of connections in the neighborhood of a certain node divided by the maximum number of possible connections between the neighbors of this node. Global efficiency is inversely related to the characteristic path length and measures how efficiently information is communicated between nodes. Local efficiency of a given node is the inverse of the average shortest path connecting all neighbors of that node and evaluates the influence of different paths based on the connection weights of the node's neighbors, i.e., a path made of strong connections contributes to the local efficiency more than a path made of weak connections. Therefore, local efficiency of a node is related to its clustering coefficient, since more connections or stronger ones between neighbors directly affect both measures.

All the network parameters were computed using the Brain Connectivity Toolbox (BCT) (Rubinov and Sporns, 2010). Negative correlations in association matrices were not included in any analysis of network measures, since they need to be removed prior to BCT computations (Rubinov and Sporns, 2010, 2011). Different network measures require different pre-processing of the association matrix.

Small-worldness analysis

Characteristic path length (L) and clustering coefficient (C) were computed to study the small-worldness of our data (Dataset 1, **Figure 3**) and of an independent set of resting-state fMRI (Dataset 2, **Figure 4**) to verify the small-worldness of the network in our sample and to provide a baseline for our measurements. These calculations used binary matrices obtained by thresholding the correlation matrices (**Figure 1**) at a range of values. The same analysis was applied to 20 random matrices with the same number of connections and similar distribution of connections (Sporns and Zwi, 2004), to obtain a random-matrix characteristic path length (L_{random}) and clustering coefficient (C_{random}). The networks are said to have small-world organization for correlation thresholds in which $L = L_{random}$ and $C > C_{random}$; this was calculated using a 2-sample t-test for $p \leq 0.01$.

Analysis of global network properties and intelligence

Global network parameters (characteristic path length, clustering coefficient, and efficiency), obtained using weighted networks, were related to the intelligence indices using the Pearson correlation coefficient (**Table 4**). The Z-transformed correlation matrix was used for the association matrix, except for global efficiency, which used the Pearson correlations due to the need to restrict the range to [0,1]. Negative values were set to zero. Some form of normalization is necessary to obtain measures that are independent of the network size, dividing parameters obtained from brain networks by those obtained from random networks. For normalization of weighted networks, a recently approach purposes to compute the average value from an ensemble of surrogate graphs (Stam et al., 2009). In our study, 100 surrogate random weighted networks were constructed, derived from the original networks by randomly permuting the edge weights. The parameters of these random weighted networks were averaged

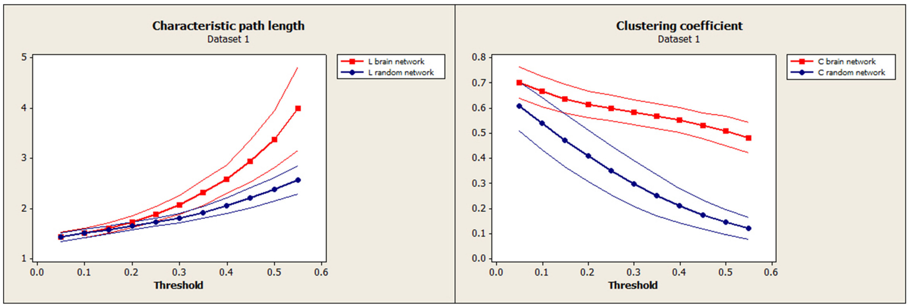


FIGURE 3 | Our data (Dataset 1): Mean characteristic path length for brain (red) and random (blue) networks are shown on the left as a function of threshold. Mean clustering coefficient for brain (red) and random (blue) networks are shown on the right. Confidence bands represent ±1 standard deviation.

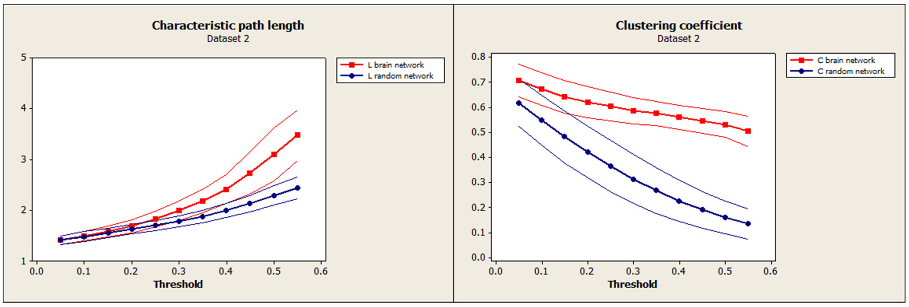


FIGURE 4 | Independent test data (Dataset 2): Mean characteristic path length for brain (red) and random (blue) networks are shown on the left. Mean clustering coefficient for brain (red) and random (blue) networks are shown on the right. Confidence bands represent ±1 standard deviation.

and used in normalization. For this analysis, p-values were not adjusted.

An additional analysis of global characteristic path length and global clustering coefficient associated to intelligence indices was performed using a binarized association matrix (thresholded at $r = 0.45$) to facilitate comparisons with Van den Heuvel et al. (2009) (**Figure 5**). Both metrics were normalized using the same 20 equivalent random binary matrices, specified in Section Small-Worldness Analysis, averaged for each brain network. Pearson correlations were also transformed using the Fisher Z in this analysis.

Analysis of local network properties and intelligence

Finally, local efficiency, which is related to clustering coefficient, was related to the intelligence indices using the Pearson correlation coefficient (**Table 5**, **Figure 6**). Local efficiency calculations used the untransformed Pearson correlation matrix for the association matrix, except that negative weights were replaced with 0. For this analysis, false discovery rates were computed per node (over the list of the 81 other regions).

RESULTS

Of the 31 volunteers, one did not perform the intelligence test and exhibited excessive movement during imaging acquisition; thus 30 participants (Dataset 1) were included in the small-world organization study (ages: mean 27 years, standard deviation 6, range: 19–38; 15 women) and 29 participants were included in the intellectual performance study (ages: mean 27 years, standard deviation 6, range: 19–38; 14 women). Demographic data for the intellectual performance study (29 participants) are in **Table 1**.

We have included a table of correlations between the intelligence indices in our sample (**Table 2**). Verbal IQ (VIQ) was strongly correlated with Verbal Comprehension Index (VCI) and Working Memory Index (WMI). Performance IQ (PIQ) was correlated strongly with Perceptual Organization Index (POI) and moderately with Processing Speed Index (PSI). This was expected because VIQ and PIQ are derived from the fundamental indices, and so these indices were not used in the analysis of this study. Full scale IQ (FSIQ) was strongly correlated with Perceptual Organization and Working Memory indices and moderately correlated with Verbal Comprehension and Processing Speed Indices, also expected.

ASSOCIATIONS BETWEEN FUNCTIONAL CONNECTIVITY AND INTELLIGENCE

Possible correlations of functional connectivity with FSIQ and perceptual organization are shown in **Table 3** and **Figure 2**. **Table 3** shows all correlations with FDR<0.05; Tables S2–S6 in the Supplemental Material show complete results for the 15 most significant associations for each IQ index. The most prevalent regions were pre-central, parietal, and occipital.

SMALL-WORLDNESS ANALYSIS

To establish the baseline validity of the network analysis, we computed small-worldness for our data and compared the results to an independent data set. Brain networks showed a clear small-world organization over a range of thresholds. **Figure 3** (left) and **Figure 4** (left) show normalized characteristic path length from binary networks as a function of threshold for participants for Dataset 1 and Dataset 2, respectively. Mean values for 20 matched random networks are also shown for comparison. **Figure 3** (right) and **Figure 4** (right) shows the same for the normalized clustering coefficient. In both datasets, networks showed a clear small-world organization for correlation thresholds between 0.05 and 0.20, characterized by $L \approx L_{\text{random}}$ for thresholds lower than 0.20 and $C \gg C_{\text{random}}$ for thresholds higher than 0.05 (2-sample t-test, all $p < \alpha = 0.01$, Bonferroni corrected for multiple thresholds).

ASSOCIATIONS BETWEEN GLOBAL NETWORK PROPERTIES AND INTELLIGENCE

We observed a negative, though statistically weak ($p = 0.14$), correlation between FSIQ and normalized characteristic path length (lambda) (**Figure 5**, left). This was computed using correlation matrices binarized at a threshold of 0.45, the same threshold applied by Van den Heuvel et al. (2009), for the purpose of direct comparison.

Verbal comprehension was associated with normalized global efficiency ($r = 0.43$, $p = 0.02$, uncorrected p-value). Also, global efficiency was weakly correlated with FSIQ ($r = 0.24$, $p = 0.22$, uncorrected p-value). These results along with a complete list of correlations between intelligence scores and global network parameters are shown in **Table 4**.

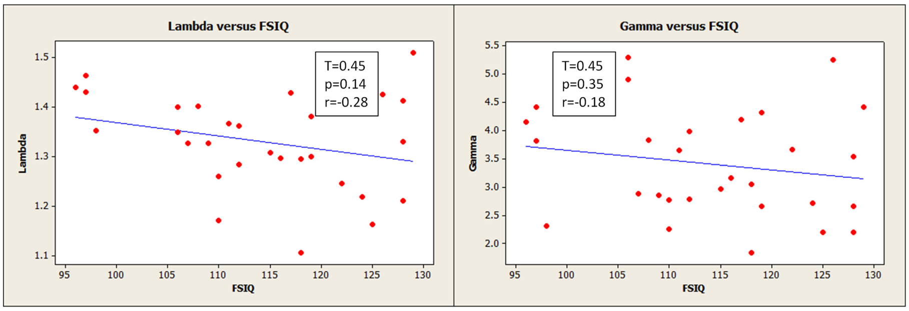


FIGURE 5 | Normalized characteristic path length (lambda) (left) and normalized clustering coefficient (gamma) (right) had slight negative relationships with Full Scale IQ, though these were not statistically robust. The network path strengths were based on binarized correlation matrices thresholded at 0.45 for this analysis. (♦) corresponds to measurements for an individual participant.

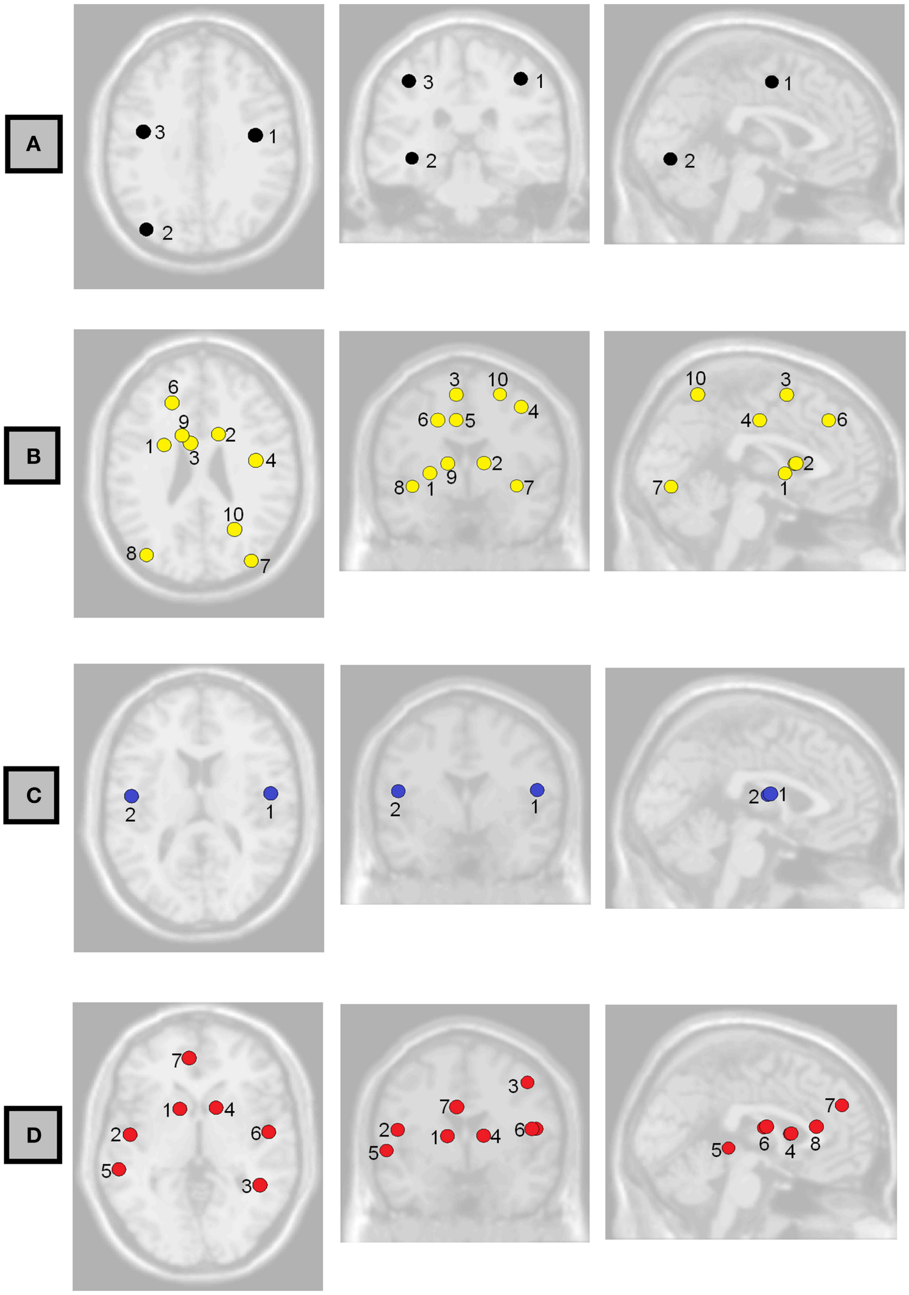


FIGURE 6 | Axial, coronal, and sagittal views of the brain showing the non-normalized weighted-network local efficiency in the regions where it had the strongest association with (A) Full-Scale IQ, (B) Verbal Comprehension Index, (C) Working Memory Index, and (D) Processing Speed Index. Labels correspond to those shown in **Table 5**.

ASSOCATIONS BETWEEN LOCAL NETWORK PROPERTIES AND INTELLIGENCE

We observed also possible relationships between local efficiency and measures of intelligence (**Table 5**, **Figure 6**). Prominent regions were pre-central gyrus, associated with FSIQ; caudate nucleus, associated with verbal comprehension and processing speed; bilateral inferior occipital gyrus, associated with verbal comprehension; and bilateral rolandic operculum, associated with working memory and processing speed. However, in all cases the false discovery rate was >0.05; uncorrected *p*-values are reported here.

DISCUSSION

We have extended a number of previous observations concerning brain functional connectivity and intelligence to the Portuguese-speaking population. These include the presence of small-world organization and correlations of intelligence with global and local characteristics of the brain's functional networks. Additionally, some novel findings in this exploratory study suggest hypotheses for future research.

The global functional brain network exhibited small-world organization at correlation thresholds between 0.05 and 0.20, α = 0.01, Bonferroni corrected for multiple comparisons of thresholds, and this closely matched the small-world organization that was apparent in the confirmation data set (**Figures 3, 4**). This suggests a high level of local clustering combined with a relatively small number of long-distance connections (Watts and Strogatz, 1998). This threshold range is smaller than the thresholds of 0.3–0.5 reported in previous observations of small-worldness in whole-brain networks (Van den Heuvel et al., 2008, 2009). However, node definitions differed substantially between the studies as well. Small-world networks are an attractive model for the connected human brain, because of their ability to transfer information with high efficiency for low wiring cost (Watts and Strogatz, 1998), and seem ubiquitous in the organization of anatomical connectivity, affected in a variety of diseases (Bassett and Bullmore, 2009). Moreover, Sporns and Zwi, in 2004, stated that information integration and even mental awareness depend on the small-world structure. Our replication of this effect supports the validity and the reliability of the network measures in this sample.

Table 1 | Demographic data and estimated intelligence scores.

Category	Data
Gender (M/F)	15/14
Age (years-old)	26.8 ± 5.8
Verbal IQ	111.7 ± 10.8
Performance IQ	116.0 ± 11.4
Full-scale IQ	114.2 ± 10.0
Verbal comprehension index	111.9 ± 11.0
Perceptual organization index	115.3 ± 11.9
Working memory index	111.4 ± 12.3
Processing speed index	116.1 ± 12.0

Age and intelligence scores are shown as mean ± standard deviation.

Table 2 | Correlations between intelligence scores.

	VIQ	PIQ	FSIQ	VCI	POI	WMI
PIQ	**0. 54**					
FSIQ	**0. 90**	**0. 85**				
VCI	**0. 84**	0. 29	**0. 67**			
POI	**0. 52**	**0. 95**	**0. 81**	0. 23		
WMI	**0. 73**	**0. 56**	**0. 74**	**0. 49**	**0. 52**	
PSI	0. 45	**0. 55**	**0. 55**	0. 38	0. 35	**0. 53**

Bold numbers represent significant values for α < 0.01.

Globally, FSIQ showed a weak negative correlation with characteristic path length (**Figure 5**, left; $r = -0.28$, 95% CI = −0.59, 0.10), although with no statistical significance. Additionally, global efficiency (inversely correlated with path length) showed a weak positive correlation with FSIQ (**Table 4**; $r = 0.24$ 95% CI = −0.14, 0.56), not statistically significant also. These same correlations were weaker when the full (weighted) association matrix was used (**Table 4**) instead of a binarized matrix (**Figure 5**). It is not known whether the thresholding step increases or decreases the reliability of the resulting measurements; however, possibly of note, correlations were observed to be the same sign in our results and in previous literature regardless of method or statistical significance. The consistent finding of a negative correlation between characteristic path length and FSIQ could be an extension to Portuguese speakers of the previous finding in Dutch speakers (Van den Heuvel et al., 2009): for characteristic path length, $r = -0.54$, 95% CI −0.80,−0.11. The negative correlation is consistent with the previously proposed idea that human intelligence is related to how efficiently different brain regions are organized and integrated (Van den Heuvel et al., 2009). It also suggests that functional brain networks are optimized in computational efficiency to promote higher

Table 3 | Associations between functional connectivity and intelligence indices (Full-Scale IQ—FSIQ, Perceptual Organization Index—POI) for specific nodes (center coordinates in x, y, and z) in the overall network, and with (uncorrected) 95% confidence intervals.

Index	Label	Functional connectivity between (AAL label)				Correlation	FDR
		MNI Region A	Center A (mm)	MNI Region B	Center B (mm)		
FSIQ	1	Fusiform R	(33.7, −40.2, −21.5)	Parietal Sup L	(−23.7, −60.8, 57.7)	0.62 (0.36, 0.80)	0.003
	2	Pre-central L	(−39.0, −7.0, 49.6)	Occipital Sup R	(24.0, −82.2, 29.3)	0.60 (0.30, 0.79)	0.05
	3	Occipital Sup R	(24.0, −82.2, 29.3)	Parietal Sup L	(−23.7, −60.8, 57.7)	0.59 (0.29, 0.79)	0.03
	4	Pre-central L	(−39.0, −7.0, 49.6)	Occipital Inf R	(37.9, −83.2, −90)	0.57 (0.26, 0.78)	0.05
POI	1	Pre-central L	(−39.0, −7.0, 49.6)	Occipital Inf L	(−36.5, −79.6, −9.2)	0.67 (0.40, 0.83)	0.006
	2	Parietal Sup R	(25.8, −60.4, 60.7)	Paracentral Lobule L	(−8.0, −26.7, 68.7)	0.66 (0.38, 0.82)	0.009
	3	Occipital Inf R	(37.9, −83.2, −90)	Post-central L	(−42.9, −23.8, 47.5)	0.63 (0.35, 0.81)	0.015
	4	Pre-central L	(−39.0, −7.0, 49.6)	Occipital Inf R	(37.9, −83.2,−90)	0.62 (0.32, 0.80)	0.015
	5	Frontal Sup Orb L	(−5.4, 52.5, −8.9)	Frontal Sup Orb R	(7.8, 50.4, −8.5)	0.61 (0.31, 0.80)	0.04
	6	Pre-central L	(−39.0, −7.0, 49.6)	Parietal Sup R	(25.8, −60.4, 60.7)	0.59 (0.29, 0.79)	0.020

Functional connectivity was measured as the Fisher transformed correlation between the two regions' time series. Only region pairs whose connectivity was correlated with IQ index at FDR < 0.05 are shown. Tables S2–S6 in the Supplemental Material show further results.

processing speed (Van den Heuvel et al., 2009) with minimal wiring cost (Chklovskii et al., 2002).

The network parameters studied here were measurements of functional segregation (clustering coefficient and local efficiency), that describe the processing occurring within densely interconnected networks of brain regions; and functional integration (characteristic path length, and its inverse, global efficiency), that is related to how information from distributed brain regions is combined (Rubinov and Sporns, 2010). Global efficiency was associated with verbal comprehension ($r = 0.43$; 95% CI = 0.08, 0.69) (**Table 4**), a novel suggestive finding worthy of further study. This finding, combined with associations between VCI and local efficiency found in several brain regions (**Table 5**, **Figure 6B** and further discussed below) suggests that linguistic and verbal abilities are linked with a higher brain efficiency, at both global and local levels.

No other associations were found between global network parameters and intellectual performance (**Table 4**). Because of the relatively small sample size of this study, we are not able to make strong conclusions from this and it does not necessarily conflict with prior findings, as our estimated 95% confidence intervals included the statistically significant correlation values found by others (Song et al., 2009; Van den Heuvel et al., 2009). However, it is possible that relationships between functional connectivity and intelligence could be limited to sub-networks of the brain, rather than being present at a global level, so we proceeded to examine network characteristics at a regional level also.

Local efficiency in the caudate nuclei was associated with VCI (**Table 5**). Some studies show that this region is important for language and verbal abilities, revealing that a smaller shortest path between the caudate and neighbor regions would be related to a higher verbal intelligence. This was not the only feature involving the caudate that was related with verbal abilities. Caudate function has also been related to verbal fluency during a working memory task (Gruber and von Cramon, 2003), and has shown activity during speech contrasted with a non-speech rest baseline condition (Simmonds et al., 2011). Significant associations with verbal fluency performance have also been found for caudate nuclei volume, suggesting that this region is implicated in the circuitry mediating this ability (Hannan et al., 2010). Left caudate plays an important role in language selection in both monolingual and multilingual people (Crinion et al., 2006), and some studies propose that the caudate would act to fine-tune interactions between automatic and more complex language processing (Friederici, 2006) or in the resolution of word ambiguity (Ketteler et al., 2008).

Local efficiency in the parietal gyrus was correlated with Verbal Comprehension and Processing Speed indices (**Table 5**), and connection strengths to the parietal lobe correlated with Perceptual

Table 5 | Associations between non-normalized weighted-network local efficiency and intelligence indices (Full-Scale IQ—FSIQ, Verbal Comprehension Index—VCI, Working Memory Index—WMI, Processing Speed Index—PSI) for specific nodes in the overall network, with 95% confidence intervals and *p*-values (uncorrected for multiple comparisons).

Intelligence index	Label	AAL atlas region	Correlation with local efficiency
FSIQ	1	Pre-central R	0.48 (0.14, 0.72) $p = 0.009$
	2	Occipital Inf L	0.45 (0.11, 0.70) $p = 0.013$
	3	Pre-central L	0.37 (0.010, 0.65) $p = 0.05$
VCI	1	Putamen L	0.50 (0.16, 0.73) $p = 0.006$
	2	Caudate R	0.48 (0.13, 0.72) $p = 0.009$
	3	Supp Motor Area L	0.42 (0.07, 0.68) $p = 0.022$
	4	Pre-central R	0.42 (0.06, 0.68) $p = 0.024$
	5	Cingulum Mid L	0.42 (0.06, 0.68) $p = 0.024$
	6	Frontal Sup L	0.41 (0.06, 0.68) $p = 0.026$
	7	Occipital Inf R	0.41 (0.05, 0.67) $p = 0.028$
	8	Occipital Inf L	0.40 (0.04, 0.67) $p = 0.03$
	9	Caudate L	0.38 (0.016, 0.66) $p = 0.04$
	10	Parietal Sup R	0.37 (0.010, 0.65) $p = 0.05$
WMI	1	Rolandic Oper R	0.52 (0.19, 0.74) $p = 0.004$
	2	Rolandic Oper L	0.42 (0.07, 0.68) $p = 0.022$
PSI	1	Caudate L	0.46 (0.11, 0.70) $p = 0.013$
	2	Rolandic Oper L	0.45 (0.10, 0.70) $p = 0.014$
	3	Parietal Inf R	0.41 (0.05, 0.67) $p = 0.028$
	4	Caudate R	0.41 (0.05, 0.67) $p = 0.029$
	5	Temporal Mid L	0.39 (0.03, 0.66) $p = 0.03$
	6	Rolandic Oper R	0.39 (0.03, 0.66) $p = 0.04$
	7	Frontal Sup Medial L	0.39 (0.03, 0.66) $p = 0.04$
	8	Frontal Inf Tri R	0.38 (0.013, 0.65) $p = 0.04$

Only the subset with correlations at $p < 0.05$ are shown (uncorrected for multiple comparisons).

Table 4 | Pearson correlations between normalized weighted-network global parameters (characteristic path length, global efficiency, and global clustering coefficient) and intelligence indices (Full-Scale IQ—FSIQ, Verbal Comprehension Index—VCI, Perceptual Organization Index—POI, Working Memory Index—WMI, Processing Speed Index—PSI) with 95% confidence intervals and *p*-values (uncorrected for multiple comparisons).

	Normalized characteristic path length	Normalized global efficiency	Normalized global clustering coefficient
FSIQ	−0.15 (−0.49, 0.22) $p = 0.42$	0.24 (−0.14, 0.56) $p = 0.22$	−0.07 (−0.42, 0.31) $p = 0.74$
VCI	−0.27 (−0.58, 0.11) $p = 0.16$	0.43 (0.08, 0.69) $p = 0.02$	−0.25 (−0.57, 0.12) $p = 0.18$
POI	−0.09 (−0.44, 0.29) $p = 0.64$	0.08 (−0.29, 0.44) $p = 0.67$	0.02 (−0.34, 0.39) $p = 0.90$
WMI	−0.08 (−0.43, 0.30) $p = 0.68$	0.17 (−0.20, 0.51) $p = 0.36$	−0.03 (−0.39, 0.34) $p = 0.88$
PSI	−0.04 (−0.40, 0.33) $p = 0.82$	0.12 (−0.26, 0.46) $p = 0.55$	0.15 (−0.23, 0.49) $p = 0.43$

Organization and Working Memory indices (Tables S4, S5 in Supplemental Material).

Local efficiency and connection strength in occipital lobe regions were associated with higher general intelligence scores and other indices (**Tables 3**, **5**). This suggests an impact of early perceptual processing on WAIS scores, especially Perceptual Organization. Although we did not observe correlations between the POI and segregational network properties (**Table 5**), there were some correlations with individual connections (**Table 3**). This may mean that this index is more related to individual connections than to network organization, possibly because of the necessity of rapid transfer of information of this region to others. It may reflect the same phenomenon observed in a recent study where higher IQ was correlated with shorter inspection time measured by EEG (which tells how fast the system extracts information from a given stimulus) because recurrent signals—those that are transmitted from a higher-tier sensory region to a lower one and that cognitive functions rely on—reach visual areas faster (Jolij et al., 2007).

Local efficiency of bilateral rolandic operculum correlated with WMI (**Table 5**, **Figure 6C**). This region encompasses part of the pre-central gyrus. This is consistent with a number of other findings relating pre-central areas to working memory, in terms of both activity (Gruber and von Cramon, 2003; Colom et al., 2010) and functional connectivity (Newton et al., 2011; Cole et al., 2012). We also observed a correlation between left pre-central regions and occipital ones with measures of general and fluid intelligence (**Table 3**, **Figure 2A**). Although other findings reported that pre-central activity and connectivity properties are related to fluid intelligence (Cole et al., 2012) as well as general intelligence (Gray et al., 2003), the specific role of the pre-central-occipital connection to the general intelligence is not known. Since these relationships are not described yet in the literature, this study may be a starting point for this question.

At the level of single paths, the strongest correlations we observed between FSIQ and functional connectivity (**Table 3**, **Figure 2A**) are consistent with the parieto-frontal integration theory (P-FIT) of Jung and Haier (2007), which was based on an extensive review of the literature relating measures of intelligence to brain structure and function. Individual differences of the described connections in this model are predicted to correlate with differences in intellectual performance. That is what we have partially observed in the patterns of functional connectivity, with higher functional connectivity predicting greater FSIQ and perceptual organization capacity. The model proposes information flow from basic sensory/perceptual processing regions to areas where structural abstraction and elaboration are involved. This is represented in our results by the connection between fusiform gyrus—a region involved in recognition of visual input and visual imagery—and parietal gyrus; and the connection between occipital and parietal cortex (**Table 3**). Then, a parieto-frontal network is responsible for information processing and abstraction, and finally the anterior cingulate selects the response (Jung and Haier, 2007), although no associations could be detected in our study to corroborate these two parts of the model. Nevertheless, direct connections between occipital regions and pre-central ones were associated with FSIQ (**Table 3**, Table S2 in Supplementary Material), which is not in accordance with the P-FIT and thus suggests a need for further study. Of note, as not all of the relationships predicted by this model were present, more experiments would be needed to robustly confirm or reject all aspects of the model.

Our selection of 82 pre-defined atlas regions as network nodes offers reduced complexity of the networks and higher data processing speed compared to a voxel-wise approach, and possibly easier interpretability of the findings in terms of known properties of the relatively large regions. The finding of small-world organization bolsters the comparability of our results to those of other studies that used different node definitions. However, it is also true that results of this study are partially dependent on the node definitions, and the node definitions used here may not coincide with others. Example of correspondences include an association between local efficiency in the left pre-central gyrus and the Full-Scale IQ for a weighted anatomical network made of 90 AAL atlas regions (Li et al., 2009) ($r = 0.25$; 0.03, 0.45), endorsing our result in **Table 5** ($r = 0.37$; 0.010, 0.65). In addition, we observed a weak correlation ($r = 0.24$; $p = 0.22$) between global efficiency and Full-Scale IQ (**Table 4**), just as Song et al. (2009) did for the default mode network ($r = 0.24$; $p = 0.072$). Findings we did not observe include those involving local efficiency of a number of cortical and subcortical regions (Li et al., 2009) and the associations between intelligence and functional connectivity reported by Song et al. (2008, 2009). Direct comparisons are reported in the Supplement Material (Tables S7, S8).

In an exploratory study such as this one, the possibility of chance findings must be clearly communicated. Failing to acknowledge multiple tests would lead to many false positive associations. On the other hand, strictly controlling type I error is likely to eliminate interesting leads in a sample of this size. Therefore, in associations between path connectivity values and intelligence scores, we compromised by controlling the false discovery rate (estimated fraction of positive findings that were false) at 5% for each path (3321 values). As the associations with global and local network parameters showed high p-values, FDR control was not performed in these cases to conserve a few of the most relevant associations. Our findings that certain regions were important in more than one context, and that some regions showed symmetric bilateral effects, do lend some apparent validity to the results. We have provided complete information about the statistical reliability of all findings to facilitate hypothesis development and comparisons with other studies.

Further study of the relationships between brain network organization and intelligence would be necessary to complement and extend the findings shown here. This study considered a Portuguese-speaking population, but further data from different populations should be analyzed to allow the results to be generalized, in particular the relationship between global efficiency and verbal intelligence that was strongly apparent in our work. More detailed templates could be used in the definition of the network nodes for a finer-grained investigation of the brain's connectivity. It is also noteworthy that we considered only positive correlations between nodes; anticorrelations may provide complementary data once methods to quantify them arise (Rubinov and Sporns, 2010).

The findings shown here replicate and extend the negative association between characteristic path length of the functional brain network and cognitive general intelligence for a Portuguese-speaking population. The small-world organization model was verified as a feature of brain networks, suggesting an ability to transfer information with high efficiency and low wiring cost. Global efficiency was weakly associated with general intelligence but strongly associated with VCI, a novel finding. Combined with the observed relationship between verbal comprehension and local efficiency in several regions, this suggests that a possible link between language ability and organizational and integrational properties of the brain network warrants further study. Additionally, an exploratory analysis suggested associations between intelligence and network properties of frontal, parietal, and occipital cortices; and fusiform, supramarginal, pre-central gyrus, and caudate nuclei.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fnhum.2015.00061/abstract

REFERENCES

Bassett, D. S., and Bullmore, E. T. (2009). Human brain networks in health and disease. *Curr. Opin. Neurol.* 22, 340–347. doi: 10.1097/WCO.0b013e32832d93dd

Beckmann, C. F., DeLuca, M., Devlin, J. T., and Smith, S. M. (2005). Investigations into resting-state connectivity using independent component analysis. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 360, 1001–1013. doi: 10.1098/rstb.2005.1634

Biswal, B. B., Mennes, M., Zuo, X.-N., Gohel, S., Kelly, C., Smith, S. M., et al. (2010). Toward discovery science of human brain function. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4734–4739. doi: 10.1073/pnas.0911855107

Biswal, B., Yetkin, F. Z., Haughton, V. M., and Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.* 34, 537–541.

Bullmore, E., and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10, 186–198. doi: 10.1038/nrn2575

Chklovskii, D. B., Schikorski, T., and Stevens, C. F. (2002). Wiring optimization in cortical circuits. *Neuron* 34, 341–347. doi: 10.1016/S0896-6273(02)00679-7

Cole, M. W., Yarkoni, T., Repovs, G., Anticevic, A., and Braver, T. S. (2012). Global connectivity of prefrontal cortex predicts cognitive control and intelligence. *J. Neurosci.* 32, 8988–8999. doi: 10.1523/JNEUROSCI.0536-12.2012

Colom, R., Karama, S., Jung, R. E., and Haier, R. J. (2010). Human intelligence and brain networks. *Dialogues Clin. Neurosci.* 12, 489–501.

Crinion, J., Turner, R., Grogan, A., Hanakawa, T., Noppeney, U., Devlin, J. T., et al. (2006). Language control in the bilingual brain. *Science* 312, 1537–1540. doi: 10.1126/science.1127761

Damoiseaux, J. S., Rombouts, S. A. R. B., Barkhof, F., Scheltens, P., Stam, C. J., Smith, S. M., et al. (2006). Consistent resting-state networks across healthy subjects. *Proc. Natl. Acad. Sci. U.S.A.* 103, 13848–13853. doi: 10.1073/pnas.0601417103

Friederici, A. D. (2006). What's in control of language? *Nat. Neurosci.* 9, 991–992. doi: 10.1038/nn0806-991

Friston, K. J., Frith, C. D., Liddle, P. F., and Frackowiak, R. S. J. (1993). Functional connectivity: the principal-component analysis of large (PET) data sets. *J. Cereb. Blood Flow Metab.* 13, 5–14. doi: 10.1038/jcbfm.1993.4

Gläscher, J., Tranel, D., Paul, L. K., Rudrauf, D., Rorden, C., Hornaday, A., et al. (2009). Lesion mapping of cognitive abilities linked to intelligence. *Neuron* 61, 681–691. doi: 10.1016/j.neuron.2009.01.026

Gray, J. R., Chabris, C. F., and Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nat. Neurosci.* 6, 316–322. doi: 10.1038/nn1014

Gruber, O., and von Cramon, D. Y. (2003). The functional neuroanatomy of human working memory revisited: evidence from 3-T fMRI studies using classical domain-specific interference tasks. *Neuroimage* 19, 797–809. doi: 10.1016/S1053-8119(03)00089-2

Haier, R. J., Jung, R. E., Yeo, R. A., Head, K., and Alkire, M. T. (2004). Structural brain variation and general intelligence. *Neuroimage* 23, 425–433. doi: 10.1016/j.neuroimage.2004.04.025

Hannan, K. L., Wood, S. J., Yung, A. R., Velakoulis, D., Phillips, L. J., Soulsby, B., et al. (2010). Caudate nucleus volume in individuals at ultra-high risk of psychosis: a cross-sectional magnetic resonance imaging study. *Psychiatry Res.* 182, 223–230. doi: 10.1016/j.pscychresns.2010.02.006

Jolij, J., Huisman, D., Scholte, S., Hamel, R., Kemner, C., and Lamme, V. A. F. (2007). Processing speed in recurrent visual networks correlates with general intelligence. *Neuroreport* 18, 39–43. doi: 10.1097/01.wnr.0000236863.46952.a6

Jung, R. E., and Haier, R. J. (2007). The Parieto-Frontal Integration Theory (P-FIT) of intelligence: converging neuroimaging evidence. *Behav. Brain Sci.* 30, 135–187. doi: 10.1017/S0140525X07001185

Ketteler, D., Kastrau, F., Vohn, R., and Huber, W. (2008). The subcortical role of language processing. High level linguistic features such as ambiguity-resolution and the human brain; an fMRI study. *Neuroimage* 39, 2002–2009. doi: 10.1016/j.neuroimage.2007.10.023

Li, Y., Liu, Y., Li, J., Qin, W., Li, K., Yu, C., et al. (2009). Brain anatomical network and intelligence. *PLoS Comput. Biol.* 5,1–17. doi: 10.1371/journal.pcbi.1000395

Lowe, M. J., Mock, B. J., and Sorenson, J. A. (1998). Functional connectivity in single and multislice echoplanar imaging using resting-state fluctuations. *Neuroimage* 7, 119–132. doi: 10.1006/nimg.1997.0315

Nascimento, E. (1998). Adaptação da terceira edição da escala Wechsler de inteligência para adultos (WAIS-III) para uso no contexto brasileiro. *Temas Psicol.* 6, 263–270. Available online at: http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1413-389X1998000300009&lng=pt&tlng=pt

Newton, A. T., Morgan, V. L., Rogers, B. P., and Gore, J. C. (2011). Modulation of steady state functional connectivity in the default mode and working memory networks by cognitive load. *Hum. Brain Mapp.* 32, 1649–1659. doi: 10.1002/hbm.21138

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., and Shulman, G. L. (2001). A default mode of brain function. *Proc. Natl. Acad. Sci. U.S.A.* 98, 676–682. doi: 10.1073/pnas.98.2.676

Rubinov, M., and Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* 52, 1059–1069. doi: 10.1016/j.neuroimage.2009.10.003

Rubinov, M., and Sporns, O. (2011). Weight-conserving characterization of complex functional brain networks. *Neuroimage* 56, 2068–2079. doi: 10.1016/j.neuroimage.2011.03.069

Simmonds, A. J., Wise, R. J. S., Dhanjal, N. S., and Leech, R. (2011). A comparison of sensory-motor activity during speech in first and second languages. *J. Neurophysiol.* 106, 470–478. doi: 10.1152/jn.00343.2011

Song, M., Liu, Y., Zhou, Y., Wang, K., Yu, C., and Jiang, T. (2009). Default network and intelligence difference. *IEEE Trans. Auton. Ment. Dev.* 1, 101–109. doi: 10.1109/IEMBS.2009.5334874

Song, M., Zhou, Y., Li, J., Liu, Y., Tian, L., Yu, C., et al. (2008). Brain spontaneous functional connectivity and intelligence. *Neuroimage* 41, 1168–1176. doi: 10.1016/j.neuroimage.2008.02.036

Sporns, O., and Zwi, J. D. (2004). The small world of the cerebral cortex. *Neuroinformatics* 2, 145–162. doi: 10.1385/NI:2:2:145

Stam, C. J., de Haan, W., Daffertshofer, A., Jones, B. F., Manshanden, I., van Cappellen van Walsum, A. M., et al. (2009). Graph theoretical analysis of magnetoencephalographic functional connectivity in Alzheimer's disease. *Brain* 132, 213–224. doi: 10.1093/brain/awn262

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289. doi: 10.1006/nimg.2001.0978

Van den Heuvel, M. P., and Hulshoff Pol, H. E. (2010). Exploring the brain network: a review on resting-state fMRI functional connectivity. *Eur. Neuropsychopharmacol.* 20, 519–534. doi: 10.1016/j.euroneuro.2010.03.008

Van den Heuvel, M. P., Stam, C. J., Boersma, M., and Hulshoff Pol, H. E. (2008). Small-world and scale-free organization of voxel-based resting-state functional connectivity in the human brain. *Neuroimage* 43, 528–539. doi: 10.1016/j.neuroimage.2008.08.010

Van den Heuvel, M. P., Stam, C. J., Kahn, R. S., and Hulshoff Pol, H. E. (2009). Efficiency of functional brain networks and intellectual performance. *J. Neurosci.* 29, 7619–7624. doi: 10.1523/JNEUROSCI.1443-09.2009

Watts, D. J., and Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. *Nature* 393, 440–442. doi: 10.1038/30918

Wechsler, D. (1939). *The Measurement of Adult Intelligence.* Baltimore, MD: Williams & Wilkins Co. doi: 10.1037/10020-000

Whitfield-Gabrieli, S., and Nieto-Castanon, A. (2012). Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain Connect.* 2, 125–141. doi: 10.1089/brain.2012.0073

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: 10 October 2014; accepted: 23 January 2015; published online: 10 February 2015.

*Citation: Pamplona GSP, Santos Neto GS, Rosset SRE, Rogers BP and Salmon CEG (2015) Analyzing the association between functional connectivity of the brain and intellectual performance. Front. Hum. Neurosci. **9**:61. doi: 10.3389/fnhum.2015.00061*

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

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 23 December 2014
doi: 10.3389/fnhum.2014.01032

Planning following stroke: a relational complexity approach using the Tower of London

Glenda Andrews[1]*, Graeme S. Halford[2], Mark Chappell[2], Annick Maujean[3] and David H. K. Shum[2]

[1] Behavioural Basis of Health Program, Griffith Health Institute, School of Applied Psychology, Griffith University, Gold Coast, QLD, Australia
[2] Behavioural Basis of Health Program, Griffith Health Institute, School of Applied Psychology, Griffith University, Brisbane, QLD, Australia
[3] Centre for National Research on Disability and Rehabilitation Medicine (CONROD), Griffith Health Institute, Griffith University, Meadowbrook, QLD, Australia

Edited by:
Vinod Goel, York University, Canada

Reviewed by:
Hasan Ayaz, Drexel University, USA
Sharlene D. Newman, Indiana University, USA
Sashank Varma, University of Minnesota, USA

***Correspondence:**
Glenda Andrews, School of Applied Psychology, Griffith University, Parklands Drive, Southport, Gold Coast, QLD 4222, Australia
e-mail: g.andrews@griffith.edu.au

Planning on the 4-disk version of the Tower of London (TOL4) was examined in stroke patients and unimpaired controls. Overall TOL4 solution scores indicated impaired planning in the frontal stroke but not non-frontal stroke patients. Consistent with the claim that processing the relations between current states, intermediate states, and goal states is a key process in planning, the domain-general relational complexity metric was a good indicator of the experienced difficulty of TOL4 problems. The relational complexity metric shared variance with task-specific metrics of moves to solution and search depth. Frontal stroke patients showed impaired planning compared to controls on problems at all three complexity levels, but at only two of the three levels of moves to solution, search depth and goal ambiguity. Non-frontal stroke patients showed impaired planning only on the most difficult quaternary-relational and high search depth problems. An independent measure of relational processing (viz., Latin square task) predicted TOL4 solution scores after controlling for stroke status and location, and executive processing (Trail Making Test). The findings suggest that planning involves a domain-general capacity for relational processing that depends on the frontal brain regions.

Keywords: Tower of London, planning, moves to solution, search depth, goal ambiguity, relational complexity, stroke, frontal lobes

INTRODUCTION

Planning is important in many areas of life and impairments in this capacity have adverse implications for independent living (Jefferson et al., 2006). Planning involves cognitive processes that depend on frontal regions of the brain (Shum et al., 2000, 2009; Unterrainer and Owen, 2006). In the current research, we examined the extent to which planning assessed using a 4-disk version of the Tower of London (TOL) is impaired in people who have suffered a stroke. A further issue relates to the nature of the cognitive processes that planning involves. More specifically, the research investigated the claim that processing the relations between current states, intermediate states, and goal states is a key process in planning (Halford et al., 1998) and that the complexity of these relations is a good indicator of the experienced difficulty of the TOL problems.

Planning in tower tasks such as the Tower of Hanoi and the TOL involves devising a sequence of moves in order to transform an initial state into a specified goal state. In the original 3-disk version of the TOL (viz., TOL3) developed by Shallice (1982), three colored disks are presented on three poles that differ in height. Respondents are required to rearrange the disks to match a target configuration (goal state) and to do so in a specified number of moves.

The results of several studies that employed the TOL3 to assess planning following traumatic brain injury (e.g., Cockburn, 1995; Rasmussen et al., 2006), suggested the need to increase the sensitivity of the TOL3 by including more difficult items. To address this issue, Tunstall (1999) developed the 4-disk version (TOL4) that includes ten items that require as many as nine moves. Shum et al. (2009) used the TOL4 to examine impairments in planning following traumatic brain injury. The patients performed more poorly than matched controls, but the impairment was specific to patients with frontal damage and to the items that required a greater number (i.e., six to nine) of moves. No planning impairment was observed on items that required fewer (i.e., two to five) moves. Planning performance in patients with no frontal damage was comparable to matched controls. The findings of Shum et al. (2009) demonstrated the importance of employing sensitive measures of planning. In that study, sensitivity was achieved by including simpler as well as more difficult problems that required fewer moves or more moves, respectively.

Moves to solution is widely used as a metric of TOL problem difficulty that has been employed in brain imaging studies and computational approaches to planning and problem solving in the TOL (e.g., Dehaene and Changeux, 1997; Newman et al., 2003). However, the number of moves to solution has been criticized as a complexity metric on the grounds that it does not sufficiently capture the cognitive processes underlying performance. Such criticisms have prompted researchers to consider alternate complexity metrics that tap different structural parameters of the tower tasks (Ward and Allport, 1997; Kaller et al., 2011, 2012; Köstering et al., 2014).

Köstering et al. (2014) examined two such factors (search depth and goal hierarchy) in the 3-disk TOL. Search depth refers to the number of intermediate moves that must be considered before the first goal move is made. When search depth is higher a longer series of intermediate moves and their interdependencies must be considered. Goal hierarchy (goal ambiguity) refers to the extent to which the correct sequential ordering of the goal moves is obvious from the specified goal state. When the goal state is vertical (i.e., all disks on the same pole), it is clear that the disk in the lowest position on the pole has to be placed before the disks in higher positions, so the sequential ordering of the moves is relatively unambiguous. When the goal state is flat (i.e., a disk on each of three poles), the sequential ordering of the moves is more ambiguous. Köstering et al. (2014) examined the effects of these two factors in a sample of normally aging adults. Adults aged from 60 to 76 years performed comparably on problems with low search depth, but performance declined significantly from 60 to 76 years on problems with high search depth. Adults over 76 years performed poorly irrespective of search depth. The effect of goal ambiguity was significant in that problems with less ambiguous goals were performed better than those with goals that were more ambiguous. However, this effect did not vary with age. The findings were interpreted as consistent with the frontal lobe theory of cognitive aging. Greater search depth imposes a higher demand on working memory, which is subserved by frontal regions, whereas increased goal ambiguity is thought to involve the striatum.

The search depth metric used by Köstering et al. (2014) to estimate the complexity of items on the 3-disk TOL is similar in some respects to the metric proposed in relational complexity theory (Halford et al., 1998). In this theory, complexity is defined in a domain-general way. It corresponds to the number of variables that are related in a cognitive representation, or the number of slots that must be filled. The simplest (unary) relations have a single slot. An example is class membership. The fact that Fido is a dog can be expressed as *dog*(Fido). Binary relations have two slots. An example is *larger-than*(elephant, mouse). Ternary relations have three slots as in *arithmetic addition*(2,3,5). Quaternary relations have four slots, as in *proportion*(2,3,6,9). More complex relations are predicted to impose higher processing loads than less complex relations. Thus, ternary relations impose a higher load than binary relations, and quaternary relations impose a higher load than ternary relations. On average, young adults can process four interacting variables in the same decision (Halford et al., 2005) consistent with a quaternary-relational limit.

The Method for Analysis of Relational Complexity (MARC) incorporates a set of principles for estimating the complexity of cognitive tasks (in terms of the metric) and the processing loads they impose (Halford et al., 2007b, 2010; Andrews and Halford, 2011). The estimates must be based on sound knowledge of how people perform the task and opportunities to reduce complexity and processing load through the use of segmentation and chunking must be taken into account. Segmentation involves decomposing (segmenting) complex tasks into less complex components that do not overload capacity and that can be processed in succession. Conceptual chunking involves recoding concepts into fewer variables. For example, the ternary-relational concept velocity, defined as *velocity* = *distance/time*, can be recoded into a unary-relational concept as when speed is indicated by the position of a pointer on a dial. However, the reduction in processing load occasioned by conceptual chunking comes at the cost of temporary loss of access to the relationships that make up the concept. For example, a unary-relational representation of velocity would not be sufficient to determine how velocity changes as a function of time or of distance, but it would be adequate if current velocity is the only variable of interest. By the principle of cognitive economy, humans will employ the least complex representation available to complete the task. More complex representations will be constructed only when less complex representations prove inadequate.

When tasks have multiple steps, task complexity corresponds to the most complex step. The processing load imposed will depend on the number of interacting variables that must be represented in parallel to perform the most complex step of the task, using the least demanding strategy available. Thus, demand corresponds to the peak load imposed during performance of the task, rather than to the total amount of processing involved. Complexity and number of steps can be manipulated independently as shown by Birney et al. (2006).

The relational complexity metric has been applied to tasks in many different content domains including transitive inference (Halford, 1984; Andrews and Halford, 1998; Andrews, 2010; Andrews and Mihelic, 2014), suppositional reasoning (Birney and Halford, 2002), categorical syllogisms (Zielinski et al., 2010), conditional reasoning (Cocchi et al., 2014), class inclusion (Halford and Leitch, 1989), inferences based on classification hierarchies (Halford et al., 2002b), card sorting (Halford et al., 2007a), balance scale reasoning (Halford et al., 2002a; Andrews et al., 2009), numerical reasoning (English and Halford, 1995; Andrews and Halford, 2002; Knox et al., 2010), and theory of mind (Andrews et al., 2003; Halford and Andrews, 2014), as well as decision making in gambling tasks (Bunch et al., 2007; Andrews et al., 2008), delay of gratification (Bunch and Andrews, 2012), reversal learning and conditional discrimination (Andrews et al., 2012), and comprehension of relative clause sentence (Andrews et al., 2006). The breadth with which the relational complexity metric has been (can be) applied contrasts with other metrics that apply to specific content domains or tasks with a specific structure.

Studies such as those cited above show that the complexity of relations that humans can process increases with age during childhood (Andrews and Halford, 2002, 2011; Bunch and Andrews, 2012), reaching quaternary relations in adulthood (Halford et al., 2005) before declining in later adulthood (Viskontas et al., 2005; Andrews and Todd, 2008).

In the current research, we tested the hypothesis that the difficulty of TOL4 problems stems from their complexity. A relational complexity analysis of the 10 TOL4 items was conducted. The complexity analysis of three of the problems will be illustrated. The initial configuration of disks on poles was the same for all problems and it is shown in **Figure 1A**. The yellow (Y) and white (W) disks were on the leftmost pole (1), the blue (Bu) and black (Bk) disks were on the rightmost pole (3), while the middle pole (2) was unoccupied.

A move is coded as the binary relation, shift(color, pole). In the first problem, the goal is to transform the initial configuration (**Figure 1A**) into the target configuration (**Figure 1B**) in which yellow and white are on pole 1 and black and blue disks are on

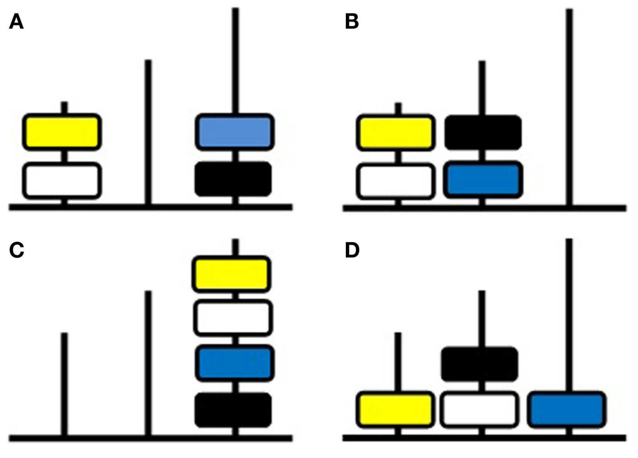


FIGURE 1 | Four-disk Tower of London (TOL4) test. **(A)** Initial configuration used in all problems; **(B)** target configuration for the binary-relational problem described in the text; **(C)** target configuration for the ternary-relational problem described in the text; **(D)** target configuration for the quaternary-relational problem described in the text.

pole 2. This requires two moves. First, blue must be moved to pole 2. This is expressed as shift(Bu, 2). Second, black must be moved to pole 2. This can be expressed as shift(Bk, 2). Each move can be performed without taking any other move into account so complexity depends solely on two slots, the disk to be moved and the location to which it is moved. Therefore both moves are binary-relational, so the maximum complexity during this problem is binary-relational.

In a more complex problem, the goal is to transform the initial configuration (**Figure 1A**) into the target configuration (**Figure 1C**) in which all four disks are on pole 3 in the top-down order yellow, white, blue, and black. This problem involves nested moves. Before white can be moved to pole 3, yellow must be moved to pole 2. Nested moves such as this are coded as the higher-order relation:

prior(shift(color, pole), shift(color, pole)).

For the problem described, this sequence can be expressed as:

prior(shift(W, 3), shift(Y, 2)).

Here, there are four slots to be filled, so *prima facie* a relation between four variables is being represented. However, conceptual chunking can be employed to reduce the task to ternary-relational. In the preceding example, Y, 2 can be chunked as a single entity corresponding to "obstructing disk" (Y2) that has to be removed to enable shift(W, 3). Thus the operative variables are: disk to be shifted (W), the goal for that disk (3), and the goal for the obstructing disk (2). The principle is that the color of the obstructing disk (Y) does not need to be processed independently of the need to find a pole to shift it to, so as to remove the obstruction of shifting white to pole 3. Planning these nested moves involves ternary-relational processing. The final move involves shifting the yellow disk to pole 3, shift(Y, 3), which is binary-relational, as in the previous example. Thus the maximum complexity during this problem is ternary-relational.

In an even more complex problem, the goal is to transform the initial configuration (**Figure 1A**) into the target configuration (**Figure 1D**) in which yellow is on pole 1, black is above white on pole 2, and blue is on pole 3. This problem involves multiple nestings and conceptual chunking. Before yellow can be placed at the base of pole 1, yellow must first be moved to pole 3 so that white can be moved to pole 2. Such situations can be expressed as the higher-order relation,

prior(shift(colour, pole), prior(shift(colour, pole)),
shift(colour, pole)).

These expressions can be read most easily starting at the rightmost move. Thus, in the example immediately below, Y, 3 is moved first, followed by W, 2, followed by Y, 1. For the problem described (**Figure 1D**), this move can be expressed as:

prior(shift(Y, 1), prior(shift(W, 2), shift(Y, 3))).

This can be chunked to quaternary-relational representation as;

prior(shift(Y, 1), prior(shift(W/Y, 2/3)))

The chunked portion can then be unpacked as;

prior(shift(W, 2), shift(Y, 3))

This yields the move to shift Y to 3 before shifting W to 2, then Y can be shifted to 1. The goal of the next move is to have blue on pole 3 and black on pole 2. To achieve this goal, blue must be first be moved to pole 1 so that black can be moved to pole 2 before blue is moved back to pole 3. This move can be expressed as,

prior(shift(Bu, 3), prior(shift(Bk, 2), shift(Bu, 1))).

As with the previous problem, chunks Bu/Bk and 2/1 can be formed, reducing the move to quaternary-relational complexity. The chunked representation can be unpacked yielding Bk on 2 and Bu on 1. Finally, Bu can be moved to 3. As in the ternary-relational problem described above, some chunking is possible. However, planning the sequence of moves will be more demanding in problems with multiple nestings because each nesting adds a new variable. By applying chunking according to the MARC principles the task can be performed with representations no more complex than quaternary-relational.

Our complexity analysis showed that the 10-item TOL4 (Shum et al., 2000, 2009) consists of two binary-relational, five ternary-relational, and three quaternary-relational problems. To ensure there were sufficient items at each complexity level, five additional items were generated, resulting in a 15-item test with three, six, and six problems at the binary-, ternary-, and quaternary-relational levels of complexity, for use in the current study.

We predicted that problems with lower estimated complexity would be easier than those with higher estimated complexity.

Based on previous research demonstrating a quaternary-relational limit in young to middle adulthood (Halford et al., 2005) and age-related declines in relational processing in later adulthood (Viskontas et al., 2005; Andrews and Todd, 2008), we expected that quaternary-relational problems would be very difficult for our participants whose mean age was 66.3 years. Problem difficulty was also examined in relation to three metrics that are specific to tower tasks; namely moves to solution, goal ambiguity, and search depth.

We predicted that frontal lobe lesions would particularly impair TOL4 performance. This prediction is based on two lines of evidence. First, planning as assessed by the TOL3 has been shown to depend on the frontal regions (Newman et al., 2003; Unterrainer and Owen, 2006; Köstering et al., 2014). Second, evidence from lesion (Waltz et al., 1999, 2004; Andrews et al., 2013) and imaging studies (Kroger et al., 2002; Crone et al., 2009) has demonstrated an important role for the frontal lobes in relational processing. Therefore, if participants who have suffered a stroke affecting the frontal brain regions should show greater impairment on the TOL4 problems than those who have suffered a stroke affecting non-frontal regions or those who have not suffered a stroke, this would be consistent with the relational processing interpretation. Group differences will be examined on TOL4 problems at each level of relational complexity and at each level of moves to solution, goal ambiguity, and search depth.

A further prediction based on relational complexity theory was that an independent measure of relational processing [viz., Latin square task (LST)] would predict TOL4 solution scores after controlling for stroke status and location. This prediction was based on research demonstrating the domain-general nature of capacity to process complex relations (Halford et al., 2002a,b; Andrews et al., 2006, 2013; Birney et al., 2006, 2012; Bunch and Andrews, 2012). The predictive ability of the LST which includes items at binary, ternary, and quaternary levels of complexity was compared to the Trail Making Test (TMT), which is widely used to assess executive processes and frontal functioning. TMT was expected to account for variance in TOL4 due to the tasks' common reliance on frontal regions (Müller et al., 2014). If the LST accounts for variance in TOL4 performance over and above the TMT this would further support the view that TOL4 involves complex relational processing.

MATERIALS AND METHODS

PARTICIPANTS

The sample consisted of 83 individuals who were all native speakers of English and who were living independently in the community. Forty-three participants had brain lesions due to stroke and 40 had no known brain injury. The unimpaired individuals were recruited through sporting and social clubs. The stroke sufferers were recruited through stroke support groups in the Brisbane and Gold Coast areas in QLD, Australia. They were assigned to a frontal stroke group ($n = 14$) or a non-frontal stroke group ($n = 29$) based on neurologists' reports and MRI/CT scan findings. Demographic details for the three groups are reported in **Table 1**.

The three groups did not differ significantly in terms of gender balance, χ^2 (2, $N = 83$) = 0.95, $p = 0.963$, age, F (2, 80) = 0.94, $p = 0.394$, nor years of education, F (2, 80) = 0.04, $p = 0.96$. Time since stroke was significantly longer for the frontal stroke group than for the non-frontal stroke group, t (41) = 2.59, $p = 0.013$. To the extent that there is some recovery of function over time, this longer time since stroke would advantage the frontal stroke group over the non-frontal stroke group, thus providing a counter-confound to predicted differences between this and the other groups.

Table 1 | Demographic details for participants in the unimpaired, non-frontal stroke, and frontal stroke groups.

Variable		Group		
		Unimpaired	Non-frontal stroke	Frontal stroke
Age (years)	*M*	68.28	64.79	64.29
	SD	12.16	13.41	8.82
Education (years)	*M*	11.75	11.76	11.50
	SD	3.12	3.30	3.23
Gender	Males	24	19	7
	Females	16	10	7
Time since stroke (years)	*M*	–	6.05	10.08
	SE	–	0.89	1.28
MMSE	*M*	28.80	27.03	26.14
	SE	0.40	0.47	0.67

N = 83.

Table 2 | Lesion location in the non-frontal and frontal stroke groups.

		Stroke group	
		Non-frontal $n = 29$	Frontal $n = 14$
Hemisphere of damage	Left	11	7
	Right	16	4
	Both	2	3
Regions of damage	Temporal	8	5
	Occipital	3	1
	Sub-cortical	19	5
	Parietal	6	7
	Frontal	0	14

Entries are frequencies.

Table 2 summarizes lesion location as a function of stroke group. There was no significant association between stroke group and damage to left, right, or both hemispheres, χ^2 (1, $N = 43$) = 3.48, $p = 0.09$, damage to temporal lobes, χ^2 (1, $N = 43$) = 0.30, $p = 0.73$, occipital lobes, χ^2 (1, $N = 43$) = 0.12, $p = 0.74$, sub-cortical regions, χ^2 (1, $N = 43$) = 3.40, $p = 0.10$, nor parietal regions, χ^2 (1, $N = 43$) = 3.85, $p = 0.08$ (exact tests).

The Mini-Mental State Examination (MMSE; Folstein et al., 1975) was administered to all participants in the standard manner. The test consists of items assessing orientation to time and place, concentration, language, constructional ability, and immediate and delayed recall. The score was the number of correct responses (max. = 30). Mean MMSE scores are shown in **Table 1**. Analysis of variance (ANOVA) revealed a significant effect of group, F (2, 80) = 7.59, $p = 0.001$, partial $\eta^2 = 0.159$. *Post hoc* Scheffe tests showed that the unimpaired group had significantly higher MMSE scores than the non-frontal stroke group ($p = 0.019$) and the frontal stroke group ($p = 0.004$). MMSE was therefore used as a covariate in all analyses that compared the groups.

MEASURES AND PROCEDURES

Ethical approval for the research was granted by the Griffith University Human Research Ethics Committee (GU Ref No: APY/82/04/HREC). Participants were tested individually at their residences by two female research assistants with postgraduate training in psychology and experience working with brain-injured individuals. The tests described below were administered as part of a larger battery. Testing was spread over two to four sessions, each 1–2 h in duration. Breaks were offered between tasks. Instructions were repeated or elaborated as required to ensure that participants understood the task requirements.

Tower of London

The task was an expanded 15-item version of the 4-disk TOL task of Shum et al. (2000, 2009). The apparatus consisted of four colored disks and a base with three vertical poles that differed in height and accommodated a maximum of two, three, or four disks. On all problems the apparatus was presented with the disks in the same initial configuration, which is shown in **Figure 1A** and **Table 3**. The goal states for the 15 problems are also shown in **Table 3** as are the moves to solution, estimated search depth, goal ambiguity, and relational complexity for each problem.

Participants were instructed to rearrange the disks into the target configuration (shown pictorially), and to do so in a specified number of moves. Only one disk could be moved at a time. Scores of three, two, or one were awarded for correct solutions on the first-, second-, and third-attempts, respectively, and zero for no solution after three attempts. All participants received the problems in the order shown in **Table 3** in which the problems with higher expected difficulty were concentrated later in the sequence. A stopping rule was implemented such that if participants failed to solve two consecutive problems after three attempts at each problem, no further problems were presented. The maximum score was 45 (based on 15 items). The mean number of TOL problems presented was 13.53 (SD = 2.11, range 5–15). Planning times were measured for the first attempt of each problem. Timing began at the commencement of each trial and ended when the first disk was moved. Instances of rule breaking (e.g., placing more than the allowed number of disks on a pole, moving two disks at a time) were also recorded. Rule breaks were not immediately corrected because doing so might have unduly influenced participants' subsequent attempts on the problem.

Latin square task

On each problem on the LST task, a 4 × 4 matrix was presented on the left side of the computer screen (Birney et al., 2006, 2012; Perret et al., 2011; Andrews and Maurer, 2012). Colored geometric objects filled some cells, while other cells were empty, as shown in **Figure 2**. The participants' task was to select one of four objects to fill a target cell (indicated by "?"). The response options were shown to the right of the matrix. The rule was that each of the four objects could occur only once in each row and column of the matrix. Consistent with the principles described previously, the complexity estimates reflect the most complex step within each problem.

For binary-relational problems, the most complex step required consideration of information from a single row or column. For example, the first step of the binary-relational problem shown in **Figure 2A**, involves working out that the empty cell in column 2 must be filled with a green square. This can be accomplished by considering the contents of a single column, column 2 in this example. On the next step, the object to be placed in the target cell can be identified by considering the contents of a single row, row 1 in this example. Row 1 now includes blue diamond, green square, and red circle, so it is clear that the pink cross must be placed in target cell. According to the analysis of Birney et al. (2006, 2012) considering the contents of a single row or a single column is binary-relational.

For ternary-relational problems, the most complex step required integration of information from a row and column. These two sources of variation must be integrated to determine the cell content. For the problem in **Figure 2B**, the first step is to identify the object to be placed in the cell at the intersection of column 3 and row 3 (blue square) by considering the objects already present in row 3 and column 3. Once this object is identified, the content of the target cell (pink cross) can be determined by considering the contents of row 3. The first (most complex) step is ternary-relational, whereas the second step is binary-relational.

For quaternary-relational problems, the most complex step required integration of information across multiple rows and columns. For the problem in **Figure 2C**, the first step is to identify the object to be placed in the cell at the intersection of column 1 and row 3 (light blue diamond) by considering the objects already present in this row and column. This step is ternary-relational. The next step requires consideration of the information in three columns (1, 2, and 4) to determine that light blue diamond should be placed in the target cell. According to the analysis provided by Birney et al. (2006, 2012) the second step is quaternary-relational.

There were four problems at each complexity level. Participants worked through the problems as quickly as possible doing all working in their heads. The score was number correct (max = 12).

Trail making test

In TMT Part A, numbers (1–25) were arranged randomly on a page. Participants drew lines connecting the numbers in ascending order as quickly as possible (Reitan and Wolfson, 1995). In TMT Part B, the stimuli were numbers (1–13) and letters (A–L). Participants drew lines connecting the numbers and letters in alternating order (1, A, 2, B, . . .). Part B required integration of two sequences (one numerical and one alphabetic) into a single alternating sequence. The two dependent measures corresponded to the times taken to complete Part A and Part B.

RESULTS

DIFFICULTY OF TOL PROBLEMS

Item-based correlations were computed to examine the extent of overlap among the four metrics and the extent to which each

Table 3 | Initial state[a], goal states, moves to solution, relational complexity, goal ambiguity, and search depth for the 15 Tower of London problems.

	Peg 1	Peg 2	Peg 3	Metric			
		Initial state					
	Yellow White	–[b]	Blue Black				
		Goal state		Moves to solution	Search depth	Goal ambiguity	Relational complexity
1	Yellow White	Black Blue	–	2	0	Moderate	Binary
2	–	–	Yellow White Blue Black	3	1	Low	Ternary
3	–	Yellow Blue	White Black	3	0	Moderate	Binary
4	Black	Blue White Yellow	–	4	0	Moderate	Binary
5	–	Blue	Yellow White Black	3	0	Moderate	Ternary
6	Black	Yellow Blue	White	5	0	High	Ternary
7	White Blue	–	Yellow Black	5	2	Moderate	Ternary
8	–	–	Blue Yellow White Black	5	2	Low	Quaternary
9	Blue Black	Yellow White	–	6	1	Moderate	Ternary
10	Yellow	Black White	Blue	6	1	High	Quaternary
11	White	Black	Yellow Blue	6	3	High	Quaternary
12	Blue	Yellow Black White	–	7	1	Moderate	Ternary
13	Yellow White	Black	Blue	7	3	High	Quaternary
14	–	Blue Yellow Black	White	9	5	Moderate	Quaternary
15	Yellow	Blue White Black	–	9	5	Moderate	Quaternary

[a] *The initial state was the same in all problems.*

[b] *Indicates an empty peg.*

metric was associated with performance on the fifteen TOL problems. As shown in **Table 4**, moves to solution, search depth and relational complexity were significantly and positively intercorrelated, but the correlations with goal ambiguity did not reach significance.

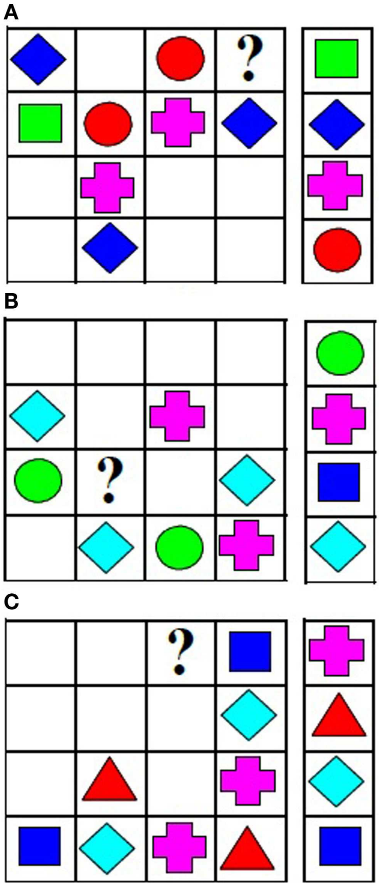


FIGURE 2 | Latin square problems at (A) binary-relational, (B) ternary-relational, and (C) quaternary-relational levels of complexity.

Moves to solution, search depth and relational complexity were significantly negatively correlated with solution accuracy on the TOL problems. Solution accuracy was lower for problems that required more moves, had greater search depth and higher relational complexity. Moves to solution, search depth and relational complexity were significantly positively correlated with planning times on problems correctly solved on the first attempt. Planning times were longer for problems that required more moves, had greater search depth and higher relational complexity. Goal ambiguity was not significantly associated with solution accuracy or planning times, therefore it was not included in subsequent regression analyses.

Item-based multiple regression analyses were conducted to determine which of three metrics accounted for independent variance in solution accuracy and planning times. Given the small sample size ($N = 15$) the findings should be interpreted with caution. In the first analysis, moves to solution, search depth and relational complexity together accounted for 88% variance in solution accuracy, F (3, 11) = 26.85, $p < 0.001$. Moves to solution (8.29%, $p = 0.019$) and search depth (6.6%, $p = 0.032$) each accounted for unique variance. The remaining variance (73%) was shared by the predictors. In the second analysis, moves to solution, search depth, and relational complexity together accounted for 76.3% variance in planning times, F (3, 11) = 11.79, $p = 0.001$. Search depth accounted for unique variance (10.96%, $p = 0.046$). The remaining variance (65%) was shared by the predictors.

TOL4 SOLUTION ACCURACY IN STROKE GROUPS

Mini-mental state examination was included as a covariate in all analyses examining group differences. The means reported for the group based analyses have been adjusted for the covariate.

A preliminary analysis of covariance (ANCOVA) was conducted with group (unimpaired, non-frontal stroke, and frontal stroke) as the between subjects variable, and MMSE as the covariate. The dependent variable was the total score (max = 45) for the 15 TOL4 problems. The analysis yielded a significant effect of Group, F (2, 79) = 5.12, $p = 0.008$, partial $\eta^2 = 0.115$. Contrast analyses showed that the difference between unimpaired group ($M = 32.29$; SE = 0.99) and the non-frontal stroke groups ($M = 30.72$; SE = 1.13) was not significant ($p = 0.31$). However,

Table 4 | Item-based correlations among moves to solution, relational complexity, goal ambiguity, search depth and solution accuracy, and planning times on the first attempt for correctly solved Tower of London problems ($N = 15$).

	Moves	Search depth	Goal ambiguity	Relational complexity	Solution accuracy	Planning times
Moves to solution						
Search depth	0.83**					
Goal ambiguity	0.28	0.05				
Relational complexity	0.75**	0.76**	0.23			
Solution accuracy	−0.90**	−0.89**	−0.32	−0.71**		
Planning times	0.81**	0.84**	0.27	0.61*	−0.90**	
Mean	5.40	1.60	1.13	3.20	2.04	22.59
SD	2.06	1.72	0.64	0.78	0.99	22.51

***p < 0.01; *p < 0.05.*

the frontal stroke group ($M = 25.91$; SE = 1.66) had significantly lower scores than the non-frontal stroke group ($p = 0.017$) and the unimpaired control group ($p = 0.002$). An analysis based on the original ten TOL4 problems yielded the same pattern of group differences.

SENSITIVITY OF THE DIFFICULTY METRICS TO STROKE DAMAGE

Four mixed ANCOVAs were conducted to examine group differences as a function of problem difficulty operationalized as moves, goal ambiguity, search depth, and relational complexity.

For the first analysis, the problems were categorized according to number of moves. The five low move problems required 2, 3, or 4 moves to solution, the six moderate move problems required 5 or 6 moves, and the four high move problems required 7 or 9 moves. Solution accuracy scores were converted to percentages and subjected to a mixed 3 × 3 ANCOVA in which Moves (low, moderate, and high) was a within-subject variable, Group was a between groups variable, and MMSE was the covariate. Consistent with the preceding ANCOVA and the correlations (**Table 4**), there were significant effects of Group, $F(2, 79) = 4.86$, $p = 0.01$, partial $\eta^2 = 0.11$, and of Moves, $F(2, 158) = 4.49$, $p = 0.013$, partial $\eta^2 = 0.054$. Percentage solution scores were higher for low move problems ($M = 94.09$; SE = 1.00) than for both the moderate move problems ($M = 70.67$; SE = 2.38) ($p = 0.007$) and the high move problems ($M = 23.37$; SE = 2.84) ($p = 0.012$). The Group × Moves interaction, $F(4, 158) = 1.37$, $p = 0.25$ was not significant. To facilitate comparison with other metrics, group differences were examined at each level of moves. The adjusted means are presented in **Table 5**.

For low move problems, there was a significant effect of group, $F(2, 79) = 7.78$, $p = 0.001$, partial $\eta^2 = 0.165$. Solution accuracy in the unimpaired group and non-frontal stroke did not differ significantly ($p = 0.84$). Solution accuracy in the frontal stroke group was significantly lower than the unimpaired ($p < 0.001$) and non-frontal stroke group ($p = 0.001$). For the moderate moves problems there were significant effects of the covariate, $F(1, 79) = 4.33$, $p = 0.041$, partial $\eta^2 = 0.052$ and of group, $F(2, 79) = 3.90$, $p = 0.024$, partial $\eta^2 = 0.09$. Solution accuracy in the unimpaired group and non-frontal stroke did not differ significantly ($p = 0.76$). Solution accuracy in the frontal stroke group was significantly lower than the unimpaired ($p = 0.009$) and non-frontal stroke group ($p = 0.016$). For the high moves problems there was no significant effect of group, $F(2, 79) = 2.31$, $p = 0.11$, partial $\eta^2 = 0.055$.

A similar approach was used to examine goal ambiguity. There were two problems with low goal ambiguity, nine with moderate goal ambiguity, and four with high goal ambiguity. There were significant effects of Group, $F(2, 79) = 5.40$, $p = 0.006$, partial $\eta^2 = 0.12$, and Goal Ambiguity, $F(2, 158) = 4.32$, $p = 0.015$, partial $\eta^2 = 0.052$. Percentage solution scores were significantly higher for low goal ambiguity ($M = 89.28$; $SE = 2.05$) than high ambiguity ($M = 53.82$; $SE = 2.97$) problems, $F(1, 79) = 6.13$, $p = 0.015$, $\eta^2 = 0.072$, and marginally higher than for problems with the moderate goal ambiguity ($M = 66.01$; SE = 1.44), $F(1, 79) = 3.60$, $p = 0.061$, $\eta^2 = 0.044$. The Group × Goal Ambiguity interaction, $F(4, 158) < 1$, $p = 0.55$, did not approach significance. Group differences for problems with low, moderate, and high goal ambiguity were examined. The adjusted means are presented in **Table 6**.

For problems with low goal ambiguity, there was a significant effect of group, $F(2, 79) = 5.44$, $p = 0.006$, partial $\eta^2 = 0.121$. Solution accuracy in the unimpaired group and non-frontal stroke did not differ significantly ($p = 0.68$). Solution accuracy in the frontal stroke group was significantly lower than the unimpaired ($p = 0.002$) and non-frontal stroke group ($p = 0.005$). For problems with moderate goal ambiguity there was a significant effect of group, $F(2, 79) = 4.92$, $p = 0.01$, partial $\eta^2 = 0.111$. Solution accuracy in the unimpaired group and non-frontal stroke did not differ significantly ($p = 0.56$). Solution accuracy in the frontal stroke group was significantly lower than the unimpaired ($p = 0.003$) and non-frontal stroke group ($p = 0.01$). For problems with high goal ambiguity there was no significant effect of group, $F(2, 79) = 2.40$, $p = 0.097$, partial $\eta^2 = 0.057$.

Search depth was examined in the same way. The five low search depth problems had a depth of zero, the six medium depth problems had depths of 1 or 2, and the four high search depth problems had depths of 3 or 5. There were significant effects of Group, $F(2, 79) = 5.36$, $p = 0.007$, partial $\eta^2 = 0.12$, and Search Depth, $F(2, 158) = 4.55$, $p = 0.012$, partial $\eta^2 = 0.054$. Percentage solution scores were significantly higher for low depth ($M = 92.55$; SE = 1.14) than high depth ($M = 16.33$; SE = 2.38) problems, $F(1, 79) = 8.30$, $p = 0.005$, $\eta^2 = 0.095$. Solution accuracy for the moderate depth ($M = 75.74$; SE = 2.43) problems did not differ significantly from low ($p = 0.11$) or high depth problems ($p = 0.15$). The Group × Search Depth interaction, F

Table 5 | Solution accuracy for TOL problems with low, moderate, and high moves by group.

Group		Moves		
		Low	Moderate	High
Unimpaired	*M*	97.60	76.95	31.63
	SE	1.36	3.23	3.85
Non-frontal stroke	*M*	97.16	75.41	21.42
	SE	1.55	3.68	4.39
Frontal stroke	*M*	87.52	59.64	17.07
	SE	2.72	5.41	6.45

Table 6 | Solution accuracy for TOL4 problems with low, medium, and high goal ambiguity by group.

Group		Goal ambiguity		
		Low	Medium	High
Unimpaired	*M*	95.78	70.60	62.31
	SE	2.78	1.95	4.02
Non-frontal stroke	*M*	93.99	68.84	54.10
	SE	3.17	2.23	4.58
Frontal stroke	*M*	78.09	58.59	45.05
	SE	4.66	3.27	6.73

Table 7 | Solution accuracy for TOL4 problems with low, moderate, and high search depth by group.

Group		Search depth		
		Low	Medium	High
Unimpaired	*M*	96.23	80.69	25.60
	SE	1.54	3.29	3.22
Non-frontal stroke	*M*	97.63	79.55	13.49
	SE	1.76	3.75	3.67
Frontal stroke	*M*	83.79	66.98	9.90
	SE	2.59	5.50	5.39

Table 8 | Solution accuracy for binary-, ternary-, and quaternary-relational TOL4 problems by group.

		Relational complexity		
		Binary	Ternary	Quaternary
Unimpaired	*M*	98.18	83.55	46.73
	SE	1.16	2.78	3.18
Non-frontal stroke	*M*	99.32	83.90	37.09
	SE	1.32	3.17	3.63
Frontal stroke	*M*	90.75	68.43	30.14
	SE	1.94	4.66	5.32

(4, 158) = 2.27, $p = 0.065$, partial $\eta^2 = 0.054$, approached significance. Group differences were examined at each level of search depth. The adjusted means are shown in **Table 7**.

For low depth problems, there were significant effects of the covariate (MMSE), F (1, 79) = 4.47, $p < 0.038$, $\eta^2 = 0.053$, and Group, F (2, 79) = 10.90, $p < 0.001$, $\eta^2 = 0.216$. Solution accuracy in the unimpaired and non-frontal stroke groups did not differ significantly ($p = 0.56$). Solution accuracy in the frontal stroke group was significantly lower than in the non-frontal stroke group ($p < 0.001$) and the unimpaired group ($p < 0.001$). For the moderate depth problems, solution accuracy in the unimpaired, non-frontal, and frontal stroke groups did not differ significantly, F (2, 79) = 2.37, $p = 0.10$, $\eta^2 = 0.057$. For high depth problems, there was a significant effect of Group, F (2, 79) = 4.19, $p = 0.019$, $\eta^2 = 0.096$. Solution accuracy in the unimpaired group was significantly higher than in the non-frontal stroke group ($p = 0.018$) and significantly higher than in the frontal stroke group ($p = 0.018$) but the two stroke groups did not differ significantly ($p = 0.577$).

Relational complexity was examined in the same way. There were three, six, and six problems, respectively at the binary, ternary, and quaternary-relational levels of complexity. The analysis yielded significant effects of Group, F (2, 79) = 5.65, $p = 0.005$, partial $\eta^2 = 0.125$ and Complexity, F (2, 158) = 5.23, $p = 0.006$, $\eta^2 = 0.062$. Solution accuracy was significantly higher for binary- ($M = 96.08$; SE = 0.86) than ternary-relational problems ($M = 78.63$; SE = 2.05), F (1, 79) = 4.07, $p = 0.047$, $\eta^2 = 0.049$, and for binary- than quaternary-relational problems ($M = 37.99$; SE = 2.35), F (1, 79) = 8.85, $p = 0.004$, $\eta^2 = 0.101$. There was also a significant Group × Complexity interaction, F (4, 158) = 2.43, $p = 0.05$, $\eta^2 = 0.058$. Group differences were examined at each complexity level. The adjusted means are shown in **Table 8**.

For binary-relational problems, there were significant effects of the covariate (MMSE), F (1, 79) = 4.57, $p < 0.036$, $\eta^2 = 0.055$, and Group, F (2, 79) = 7.30, $p = 0.001$, $\eta^2 = 0.156$. Solution accuracy in the unimpaired and the non-frontal stroke groups did not differ significantly ($p = 0.53$). Solution accuracy was significantly lower in the frontal stroke group than the non-frontal stroke group ($p < 0.001$) and the unimpaired group ($p = 0.002$). For the ternary-relational problems, there was a significant effect of Group, F (2, 79) = 4.47, $p = 0.015$, $\eta^2 = 0.102$. Solution accuracy in the unimpaired and the non-frontal stroke groups did not differ significantly ($p = 0.94$). Solution accuracy was significantly lower in the frontal stroke group than the non-frontal stroke group ($p = 0.006$) and the unimpaired group ($p = 0.008$). For quaternary-relational problems, there was a significant effect of Group, F (2, 79) = 3.85, $p = 0.025$, partial $\eta^2 = 0.089$. Solution accuracy was marginally higher in the unimpaired group than the non-frontal stroke group ($p = 0.054$) and significantly higher than in the frontal stroke group ($p = 0.011$). The two stroke groups did not differ significantly ($p = 0.274$).

In summary, the foregoing analyses show that patterns of group differences on problems at low, intermediate, and high difficulty levels differ according to how problem difficulty is measured. On the easiest problems, the frontal stroke group performed more poorly than the unimpaired group irrespective of whether problem difficulty was expressed in terms of moves to solution, goal ambiguity, search depth, or relational complexity. The frontal stroke group also performed more poorly than the non-frontal stroke group on the easiest problems.

On problems with an intermediate level of difficulty, the frontal stroke group performed more poorly than the unimpaired group and the non-frontal stroke group when problem difficulty was expressed in terms of moves to solution, goal ambiguity, and relational complexity, but not when difficulty was expressed in terms of search depth. No significant group differences were observed on moderate depth problems.

On problems at the highest level of difficulty, the frontal stroke group performed more poorly than the unimpaired group when problem difficulty was expressed in terms of search depth and relational complexity. The frontal and non-frontal stroke groups performed poorly on high search depth and quaternary-relational problems and there were no significant differences between these two groups. No significant differences were observed between unimpaired, non-frontal stroke, and frontal stroke groups on high move problems and problems with high goal ambiguity.

Thus the pattern of significance for the group effects shows that TOL4 problems at all three levels of the domain-general relational complexity metric were sensitive to frontal lobe damage whereas TOL4 problems at two levels of the task-specific metrics (moves, goal ambiguity, and search depth) were sensitive to frontal lobe damage. Inspection of the effect sizes reported above indicates a similar pattern in that effect sizes were <0.058 for the moderate search depth, high moves, high goal ambiguity problems for which the group effect was not significant, whereas effects sizes exceeded 0.088 in all other conditions.

PLANNING TIMES FOR TOL PROBLEMS SOLVED ON FIRST ATTEMPT

Participants with no first attempt solutions for any problems at a particular difficulty level were excluded from these analyses. This meant that the overall sample sizes were reduced to 39 ($n = 22$ unimpaired; $n = 12$ non-frontal stroke; $n = 5$ frontal stroke) for the analysis examining moves to solution, to 72 ($n = 35$ unimpaired; $n = 27$ non-frontal stroke; $n = 10$ frontal stroke) for the analysis examining goal ambiguity, to 28 ($n = 22$ unimpaired; $n = 4$ non-frontal stroke; $n = 2$ frontal stroke) for the analysis examining search depth and to 75 ($n = 38$ unimpaired; $n = 27$ non-frontal stroke; $n = 10$ frontal stroke) for the analysis examining relational complexity. The losses were due mainly to the more difficult problems, where participants were more likely to require multiple attempts.

Four separate ANOVAs were conducted with moves, goal ambiguity, search depth, or relational complexity as the within-subject factor. There was a significant effect of Moves, F (2, 76) = 20.72, $p < 0.001$, partial $\eta^2 = 0.353$. Planning times (seconds) were significantly shorter for low move problems ($M = 8.22$; SE = 0.52) than for moderate move problems ($M = 15.27$; SE = 1.56), F (1, 38) = 19.87, $p < 0.001$, partial $\eta^2 = 0.343$, which were significantly shorter than for high move problems ($M = 30.71$; SE = 4.71), F (1, 38) = 17.17, $p < 0.001$, partial $\eta^2 = 0.311$.

There was a significant effect of Goal Ambiguity, F (2, 142) = 21.36, $p < 0.001$, partial $\eta^2 = 0.231$. Planning times (seconds) for problems with low ($M = 10.73$; SE = 1.06) and moderate goal ambiguity ($M = 11.93$; SE = 0.76) did not differ significantly, F (1, 71) = 1.18, $p = 0.282$. Planning tomes for problems with low ambiguity were significantly shorter than for problems with high goal ambiguity ($M = 19.69$; SE = 1.91), F (1, 71) = 28.897, $p < 0.001$, partial $\eta^2 = 0.289$.

There was a significant effect of Search Depth, F (2, 54) = 23.48, $p < 0.001$, partial $\eta^2 = 0.465$. Planning times were significantly shorter for low depth problems ($M = 7.81$; SE = 0.60) than for medium depth problems ($M = 13.90$; SE = 1.30), F (1, 27) = 22.68, $p < 0.001$, partial $\eta^2 = 0.467$, which were significantly shorter than for high depth problems ($M = 34.76$; SE = 5.43), F (1, 27) = 20.87, $p < 0.001$, partial $\eta^2 = 0.436$.

There was a significant effect of Relational Complexity, F (2, 148) = 14.52, $p < 0.001$, partial $\eta^2 = 0.164$. Planning times were significantly shorter for binary-relational problems ($M = 8.59$; SE = 0.52) than for ternary-relational problems ($M = 11.18$; SE = 0.73), F (1, 74) = 13.80, $p < 0.001$, partial $\eta^2 = 0.157$, which were significantly shorter than for quaternary-relational problems ($M = 19.19$; SE = 2.08), F (1, 74) = 11.72, $p < 0.001$, partial $\eta^2 = 0.137$.

These findings are generally consistent with the item-based correlations. However, when Group was included as an independent variable along with MMSE as the covariate, the ANCOVAs yielded no significant effects of Group, MMSE, Moves, Goal Ambiguity, Depth or Relational Complexity, and no significant interactions. These null results likely reflect inclusion of the covariate, the small and unequal sizes of the unimpaired, non-frontal stroke and frontal stroke groups, and high within-group variability in planning times.

TOL RULE BREAKS

Analysis of covariance was applied to the number of rule breaks. The analysis yielded a significant effect of Group, F (2, 79) = 5.03, $p = 0.009$, partial $\eta^2 = 0.113$. Contrast analyses showed that the difference between the unimpaired group ($M = 0.35$; SE = 0.21) and the non-frontal stroke group ($M = 0.38$; SE = 0.23) was not significant ($p = 0.938$). The frontal stroke group ($M = 1.56$; SE = 0.34) committed significantly more rule breaks than the non-frontal stroke group ($p = 0.005$), however it should be noted that the absolute number of rule breaks was quite low ($M = 0.77$; SE = 0.15; $N = 83$).

PREDICTING TOL SOLUTION ACCURACY

Table 9 shows the zero-order correlations among the TOL4, LST scores (max = 12) and TMT-Parts A and B. Stroke status (0 = unimpaired; 1 = stroke) and frontal location (0 = no frontal injury; 1 = frontal injury) were dummy variables that together capture the grouping variable used in the ANCOVAs. The TOL4 measure is the average of the binary-, ternary-, and quaternary-relational percentages scores. The results are very similar when the total score (max 45) is used. The negative correlations occur because TMT-A and TMT-B are measures of response times rather than accuracy.

Table 9 | Zero-order correlations ($N = 83$).

	TOL4 (%)	Stroke status	Frontal	MMSE	TMT-A	TMT-B	LST
Stroke status	−0.31**						
Frontal location	−0.40***	0.43***					
MMSE	0.33**	−0.38***	−0.27**				
TMT-A (times)	−0.51***	0.34**	0.35**	−0.51***			
TMT-B (times)	−0.62***	0.38***	0.44***	−0.55***	0.85***		
Latin square	0.51***	−0.32**	−0.33**	0.29**	−0.42***	−0.50***	
Mean	73.00	0.52	0.17	27.74	43.94	146.01	5.78
SD	13.10	0.50	0.38	2.70	27.92	157.06	3.31

TOL, Tower of London; TMT-A, trail making test part A completion times; TMT-B, trail making test part B completion times; LST, Latin square task.

**$p < 0.01$; ***$p < 0.001$.

Table 10 | Multiple regression analyses predicting TOL4 solution accuracy.

	Predictors	*B*	SE (*B*)	ß	Part	*p*
Step 1	Stroke status	−2.72	3.02	−0.10	−0.09	0.372
	Frontal/non-frontal	−10.33	3.87	−0.30	−0.27	0.009
	MMSE	1.01	0.53	0.21	0.19	0.060
	Multiple $R^2 = 0.218$, F (3, 79) = 7.35, $p < 0.001$					
Step 2	Stroke status	−1.47	2.69	−0.06	−0.05	0.586
	Frontal/non-frontal	−4.90	3.60	−0.14	−0.12	0.177
	MMSE	−0.16	0.53	−0.03	−0.03	0.765
	TMT-A	0.03	0.08	0.05	0.03	0.753
	TMT-B	−0.05	0.02	−0.60	−0.29	0.001
	Multiple $R^2 = 0.405$ F (5, 77) = 10.47, $p < 0.001$					
Step 3	Stroke status	−0.67	2.62	−0.03	−0.02	0.800
	Frontal/non-frontal	−4.04	3.50	−0.12	−0.10	0.252
	MMSE	−0.15	0.51	−0.03	−0.03	0.771
	TMT-A	0.02	0.08	0.05	0.03	0.770
	TMT-B	−0.04	0.02	−0.49	−0.23	0.008
	LST	0.98	0.40	0.25	0.21	0.016
	Multiple $R^2 = 0.449$, F (6, 76) = 10.32, $p < 0.001$					

A multiple regression analysis with TOL4 as the criterion variable was conducted. On step 1, the dummy variables stroke status and frontal-non-frontal were entered, along with MMSE. These variables together accounted for significant variance in TOL4 performance. On step 2, TMT-A and TMT-B accounted for an additional 18.6% variance ($p < 0.001$). On step 3, LST accounted for a further 4.41% variance ($p = 0.016$). The unique contribution of TMT-B was reduced from 8.47% at step 2 to 5.38% at step 3, indicating that TMT-B and LST accounted for shared variance in TOL performance. This analysis is summarized in **Table 10**.

DISCUSSION

Our research examined planning assessed using a 4-disk version of the TOL (Shum et al., 2009) following stroke. The overall solution scores provided evidence of impairment but only in those whose strokes resulted in damage to frontal regions of the brain. The overall solution scores, which collapse over problem difficulty, provided no evidence of planning impairments following stroke affecting non-frontal brain regions. These findings are consistent with previous research using the TOL4 (Shum et al., 2009).

We also investigated the extent to which relational complexity theory (Halford et al., 1998), which has been shown to account for performance in many cognitive domains also applies to planning on the TOL4. According to relational complexity theory, integrating the relations between current states, intermediate states, and goal states is a key process in planning. Three aspects of the findings are consistent with relational complexity theory.

First, the observed difficulty of the TOL4 problems increased with the estimated relational complexity of the problems. This was also the case for other complexity metrics. The item-based correlations demonstrate that moves to solution, search depth, and relational complexity are not independent. In the regression analyses, search depth and moves to solution emerged as predictors of solution accuracy and search depth also predicted planning times on problems correctly solved on the first attempt, but in both cases the majority of the variance was shared. Search depth and moves to solution are intrinsic to the TOL4 task but unlike relational complexity they are not applicable across domains.

Search depth quantifies difficulty up to the first goal move. Köstering et al. (2014) showed that search depth is well suited to TOL3 problems. Our findings show that it also captures the difficulty of TOL4 problems that require up to nine moves to solution. The search depth metric and the relational complexity metric both focus on the relations and interdependencies within a sequence of moves and this might underpin the observed positive correlation.

That the number of moves metric predicted solution accuracy is consistent with many previous findings (e.g., Newman et al., 2003; Kaller et al., 2012). The finding is unsurprising in one sense because problems that require more moves to solution also provide more opportunities for errors. Nevertheless, the fact that number of moves was strongly correlated with search depth and relational complexity, which are less vulnerable to this criticism indicates its usefulness as a difficulty metric. One feature of the moves metric that might contribute to its prediction of performance is its scaling. For problems used in the current study, moves ranged from 2 to 9 with most intermediate values represented. The values of search depth (0, 1, 2, 3, 5), goal ambiguity (low, moderate, and high), and relational complexity (binary-, ternary-, and quaternary-relational) were more limited in range. These scaling differences between the metrics should be considered when interpreting the item-based correlations and regression analyses.

It is also likely that metrics that are specific to a task, as moves to solution and search depth are to TOL, will tend to account for more variance in that task. However, because such metrics cannot be applied to other tasks, they cannot be used to compare difficulty of TOL problems with other tasks. The relational complexity approach does allow this. For example, number of moves on the TOL4 task does not have the same meaning as number of moves (steps to solution) on the LST, whereas the relational complexity values are arguably comparable.

The second finding consistent with relational complexity theory is that as in previous studies (Unterrainer and Owen, 2006; Shum et al., 2009) impaired performance was most evident in people with frontal lobe damage. Relational processing is known to rely on the integrity of the frontal lobes (e.g., Waltz et al., 1999, 2004; Kroger et al., 2002; Crone et al., 2009; Andrews et al., 2013), so this finding is consistent with the view that TOL4 problems involve relational processing.

The frontal stroke group was impaired relative to unimpaired controls on TOL4 problems at all three levels of relational complexity. This was not the case when difficulty was expressed in terms of moves, goal ambiguity, and search depth. TOL4 problems with low and moderate numbers of moves, low and moderate goal ambiguity, and low and high search depth were sensitive to frontal lobe damage. Thus relational complexity was more sensitive to frontal lobe damage than the other metrics were.

Relative to the non-frontal stroke group, the frontal stroke group was impaired on low move and moderate move problems, problems with low and moderate goal ambiguity, and problems with low search depth and binary- and ternary-relational problems. Thus none of the metrics was successful in distinguishing patients with frontal versus non-frontal damage at all three levels of difficulty. The significant group effects that were observed on the most difficult quaternary-relational and high search depth problems reflected differences between unimpaired and stroke groups rather than between non-frontal and frontal stroke groups. That this impairment in the non-frontal group was detected only on a subset of the problems illustrates one benefit of analyzing the cognitive demands involved in planning on the TOL4.

Given the demonstrated limit for young adults (Halford et al., 2005), the poor performance of the two stroke groups on the quaternary-relational problems is not surprising. Recent brain imaging of individuals without brain damage showed that limits in relational processing during a deductive reasoning task were manifested in the brain as complexity-dependent modulations of large-scale networks that involved both frontal and non-frontal (e.g., parietal, occipital) regions (Cocchi et al., 2014). If these regions are damaged in individuals in the non-frontal stroke group, their performance on the quaternary-relational TOL4 problems would be adversely affected relative to the unimpaired group. Four of the six quaternary-relational problems were classified as high search depth, and this overlap would explain the similar pattern observed on the high search depth and quaternary-relational problems.

A third finding is consistent with relational complexity theory. As noted, the relational complexity approach has been applied to tasks in many different content domains and cross-domain correspondences in performance have been demonstrated in children (Andrews and Halford, 2002; Halford et al., 2002a,b; Bunch and Andrews, 2012), and adults (Andrews et al., 2006, 2013), suggesting that relational processing is a domain-general capacity. As predicted, relational processing in the LST accounted for variance in TOL4 performance after controlling for stroke status and location, MMSE and completion times on parts A and B of the TMT. The TOL4 and the LST differ substantially in terms of their stimuli and procedural requirements. Therefore the shared variance is unlikely to reflect common surface features of the tasks. We interpret the variance shared by TOL4 and LST as evidence that a common capacity for complex relational processing underpins both tasks.

Completion times for the TMT also accounted for variance in TOL4, but this was due mainly to part B rather than part A. Whereas TMT-A and TMT-B both require non-executive processes involved in visual scanning and speeded motor responses, TMT-B also requires the executive processes involved in set-shifting, maintaining two response sets in working memory, and inhibitory control (Müller et al., 2014). The unique contribution of TMT-B on step 2 of the regression analysis is consistent with the involvement of executive processes in TOL4.

As well as accounting for independent variance in TOL4, TMT-B, and LST also accounted for shared variance in TOL4. This suggests that all three tasks have some common processes. We argued previously that relational processing underpins both TOL4 and LST. TMT-B can also be construed in this way. It requires integration of two well-known sequences, one numerical and the other alphabetic. Each sequence incorporates a succession relation, in that one element is succeeded by the next element, for example, *succeeded by*(3, 4) or *succeeded by*(D, E). Succession is a binary relation because it cannot be defined on fewer than two entities. TMT-B involves integrating the numerical and alphabetic sequences such that the categories (numbers, letters) alternate, for example, *alternating*(3, D, 4). Alternation is ternary-relational because it cannot be defined on fewer than three entities. Thus we propose that the variance shared by the three tasks reflects ternary-relational processing. Some LST and TOL4 problems require quaternary-relational processing, so the unique contribution of LST might reflect this higher complexity.

The research contributes to our understanding of the processes involved in TOL4. It adds to the studies cited previously, which demonstrate that relational processing underpins performance on a wide range of cognitive tasks. Given the ubiquitous nature of relational processing, and the demonstrated effects of relational complexity on performance, relational complexity theory provides a parsimonious approach to conceptualizing human cognition.

The research also has practical implications. To the extent that planning on tower tasks can be construed as relational processing, interventions designed to improve relational processing through for example, structural alignment training (Son et al., 2011; Hribar et al., 2012), use of relational language (Gentner et al., 2011), and techniques to improve access to relational components (e.g., Andrews et al., 2012) might also have beneficial effects on planning. Thus the findings have the potential to inform cognitive rehabilitation of planning deficits following brain injury due to stroke and other factors. Impairments in planning have adverse implications for independent living (Jefferson et al., 2006). For example, without the ability to plan, a person might have problems in achieving independent activities of daily living or their vocational goals. Thus effective interventions would imply considerable benefits for individuals as well as for society more broadly.

ACKNOWLEDGMENTS

We are grateful to the organizations that assisted with recruitment of the stroke and control groups and to participants themselves. This research was funded by an Australian Research Council Discovery Grant (DPO452547) awarded to Halford, Andrews, Birney, Chappell, and Shum.

REFERENCES

Andrews, G. (2010). Belief-based and analytic processing in transitive inference depends on premise integration difficulty. *Mem. Cognit.* 38, 928–940. doi:10.3758/MC.38.7.928

Andrews, G., Birney, D. P., and Halford, G. S. (2006). Relational processing and working memory capacity in comprehension of relative clause sentences. *Mem. Cognit.* 34, 1325–1340. doi:10.3758/BF03193275

Andrews, G., Bunch, K. M., and Tolliday, E. (2008). "Young children's difficulty on the children's gambling task: complexity or variability of losses?," in *Psychology of Gambling*, ed. M. J. Esposito (New York, NY: Nova Science Publishers, Inc), 111–129.

Andrews, G., and Halford, G. S. (1998). Children's ability to make transitive inferences: the importance of premise integration and structural complexity. *Cogn. Dev.* 13, 479–513. doi:10.1016/S0885-2014(98)90004-1

Andrews, G., and Halford, G. S. (2002). A cognitive complexity metric applied to cognitive development. *Cogn. Psychol.* 45, 153–219. doi:10.1016/S0010-0285(02)00002-6

Andrews, G., and Halford, G. S. (2011). "Recent advances in relational complexity theory & its application to cognitive development," in *Cognitive Development and Working Memory: A dialogue between Neo-Piagetian and Cognitive Approaches*, eds P. Barrouillet and V. Gaillard (Hove: Psychology Press), 47–68.

Andrews, G., Halford, G. S., and Boyce, J. (2012). Conditional discrimination in young children: the roles of associative and relational processing. *J. Exp. Child Psychol.* 112, 84–101. doi:10.1016/j.jecp.2011.12.004

Andrews, G., Halford, G. S., Bunch, K. M., Bowden, D., and Jones, T. (2003). Theory of mind and relational complexity. *Child Dev.* 74, 1476–1499. doi:10.1111/1467-8624.00618

Andrews, G., Halford, G. S., Murphy, K., and Knox, K. (2009). Integration of weight and distance information in young children: the role of relational complexity. *Cogn. Dev.* 24, 49–60. doi:10.1016/j.cogdev.2008.07.005

Andrews, G., Halford, G. S., Shum, D., Maujean, A., Birney, D. P., and Chappell, M. (2013). Relational processing following stroke. *Brain Cogn.* 81, 44–51. doi:10.1016/j.bandc.2012.09.003

Andrews, G., and Maurer, J. (2012). "Does perceived power influence relational processing in the Latin square task?" in *Psychology of Power*, eds Q. G. Fry and C. O'Donnell (New York, NY: Nova Science Publishers, Inc), 1–33.

Andrews, G., and Mihelic, M. (2014). Belief-based and analytic processing in transitive inference: further evidence for the importance of premise integration. *J. Cogn. Psychol. (Hove)* 26, 588–596. doi:10.3758/MC.38.7.928

Andrews, G., and Todd, J. M. (2008). "Two sources of age-related decline in comprehension of complex relative clause sentences," in *New Research on Short Term Memory*, ed. N. B. Johansen (New York, NY: Nova Science Publishers, Inc), 93–123.

Birney, D. P., Bowman, D. B., Beckmann, J., and Seah, Y. (2012). Assessment of processing capacity: Latin-square task performance in a population of managers. *Eur. J. Psychol. Assess.* 28, 216–226. doi:10.1027/1015-5759/a000146

Birney, D. P., and Halford, G. S. (2002). Cognitive complexity of suppositional reasoning: an application of the relational complexity metric to the knight-knave task. *Think. Reason.* 8, 109–134. doi:10.1080/13546780143000161

Birney, D. P., Halford, G. S., and Andrews, G. (2006). Measuring the influence of complexity on relational reasoning: the development of the Latin square task. *Educ. Psychol. Meas.* 66, 146–171. doi:10.1177/0013164405278570

Bunch, K. M., and Andrews, G. (2012). Development of relational processing in hot and cool tasks. *Dev. Neuropsychol.* 37, 134–152. doi:10.1080/87565641.2011.632457

Bunch, K. M., Andrews, G., and Halford, G. S. (2007). Complexity Effects on the children's gambling task. *Cogn. Dev.* 22, 376–383. doi:10.1016/j.cogdev.2007.01.004

Cocchi, L., Halford, G. S., Zalesky, A., Harding, I. H., Ramm, B. J., Cutmore, T. H., et al. (2014). Complexity in relational processing predicts changes in functional brain network dynamics. *Cereb. Cortex* 24, 2283–2296. doi:10.1093/cercor/bht075

Cockburn, J. (1995). Performance on the Tower of London test after severe head injury. *J. Int. Neuropsychol. Soc.* 1, 537–544. doi:10.1017/S1355617700000667

Crone, E. A., Wendelken, C., van Leijenhorst, L., Honomichl, R. D., Christoff, K., and Bunge, S. A. (2009). Neurocognitive development of relational reasoning. *Dev. Sci.* 12, 55–66. doi:10.1111/j.1467-7687.2008.00743.x

Dehaene, S., and Changeux, J.-P. (1997). A hierarchical neuronal network for planning behavior. *Proc. Natl. Acad. Sci. U.S.A.* 94, 13293–13298. doi:10.1073/pnas.94.24.13293

English, L. D., and Halford, G. S. (1995). *Mathematics Education: Models and Processes*. Hillsdale, NJ: Erlbaum.

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 12, 189–198. doi:10.1016/0022-3956(75)90026-6

Gentner, D., Anggoro, F. K., and Klibanoff, R. S. (2011). Structure mapping and relational language support children's learning of relational categories. *Child Dev.* 82, 1173–1188. doi:10.1111/j.1467-8624.2011.01599.x

Halford, G. S. (1984). Can young children integrate premises in transitivity and serial order tasks? *Cogn. Psychol.* 16, 65–93. doi:10.1016/0010-0285(84)90004-5

Halford, G. S., and Andrews, G. (2014). Three-year-olds' theories of mind are symbolic but of low complexity. *Front. Psychol.* 5:682. doi:10.3389/fpsyg.2014.00682

Halford, G. S., Andrews, G., Dalton, C., Boag, C., and Zielinski, T. (2002a). Young children's performance on the balance scale: the influence of relational complexity. *J. Exp. Child Psychol.* 81, 417–445. doi:10.1006/jecp.2002.2665

Halford, G. S., Andrews, G., and Jensen, I. (2002b). Integration of category induction and hierarchical classification: one paradigm at two levels of complexity. *J. Cogn. Dev.* 3, 143–177. doi:10.1207/S15327647JCD0302_2

Halford, G. S., Baker, R., McCredden, J. E., and Bain, J. D. (2005). How many variables can humans process? *Psychol. Sci.* 16, 70–76. doi:10.1111/j.0956-7976.2005.00782.x

Halford, G. S., Bunch, K. M., and McCredden, J. E. (2007a). Problem decomposability as a factor in complexity of the dimensional change card sort task. *Cogn. Dev.* 22, 384–391. doi:10.1016/j.cogdev.2006.12.001

Halford, G. S., Cowan, N., and Andrews, G. (2007b). Separating cognitive capacity from knowledge: a new hypothesis. *Trends Cogn. Sci.* 11, 237–242. doi:10.1016/j.tics.2007.04.001

Halford, G. S., and Leitch, E. (1989). "Processing load constraints: a structure-mapping approach," in *Psychological Development: Perspectives Across the Life-Span*, eds M. A. Luszcz and T. Nettelbeck (Amsterdam: Elsevier), 151–159.

Halford, G. S., Wilson, W. H., and Phillips, S. (1998). Processing capacity defined by relational complexity: implications for comparative, developmental, and cognitive psychology. *Behav. Brain Sci.* 21, 803–831. doi:10.1017/S0140525X98001769

Halford, G. S., Wilson, W. H., and Phillips, S. (2010). Relational knowledge: the foundation of higher cognition. *Trends Cogn. Sci.* 14, 497–505. doi:10.1016/j.tics.2010.08.005

Hribar, A., Haun, D. B. M., and Call, J. (2012). Children's reasoning about spatial relational similarity: the effect of alignment and relational complexity. *J. Exp. Child Psychol.* 111, 490–500. doi:10.1016/j.jecp.2011.11.004

Jefferson, A. L., Paul, R. H., Ozonoff, A., and Cohen, R. A. (2006). Evaluating elements of executive functioning as predictors of instrumental activities of daily living (AIDLS). *Arch. Clin. Neuropsychol.* 21, 311–320. doi:10.1016/j.acn.2006.03.007

Kaller, C. P., Rahm, B., Köstering, L., and Unterrainer, J. M. (2011). Reviewing the impact of problem structure on planning: a software tool for analyzing tower tasks. *Behav. Brain Res.* 216, 1–8. doi:10.1016/j.bbr.2010.07.029

Kaller, C. P., Unterrainer, J. M., and Stahl, C. (2012). Assessing planning ability with the Tower of London task: psychometric properties of a structurally balanced problem set. *Psychol. Assess.* 24, 46–53. doi:10.1037/a0025174

Knox, K., Andrews, G., and Hood, M. H. (2010). "Relational processing in children's arithmetic word problem solving," in *ASCS09: Proceedings of the 9th Conference of the Australasian Society for Cognitive Science*. Sydney, Australia.

Köstering, L., Stahl, C., Leonhart, R., Weiller, C., and Kaller, C. P. (2014). Development of planning abilities in normal aging: differential effects of specific cognitive demands. *Dev. Psychol.* 50, 293–303. doi:10.1037/a0032467

Kroger, J. K., Sabb, F. W., Fales, C. L., Bookheimer, S. Y., Cohen, M. S., and Holyoak, K. J. (2002). Recruitment of anterior dorsolateral prefrontal cortex in human reasoning: a parametric study of relational complexity. *Cereb. Cortex* 12, 477–485. doi:10.1093/cercor/12.5.477

Müller, L. D., Guhn, A., Zeller, J. B. M., Biehl, S. C., Dresler, T., Hahn, T., et al. (2014). Neural correlates of a standardized version of the trail making test in young and elderly adults: a functional near-infrared spectroscopy study. *Neuropsychologia* 56, 271–279. doi:10.1016/j.neuropsychologia.2014.01.019

Newman, S. D., Carpenter, P. A., Varma, S., and Just, M. A. (2003). Frontal and parietal participation in problem solving in the Tower of London: fMRI and computational modeling of planning and high-level perception. *Neuropsychologia* 41, 1668–1682. doi:10.1016/S0028-3932(03)00091-5

Perret, P., Bailleux, C., and Dauvier, B. (2011). The influence of relational complexity and strategy selection on children's reasoning in the Latin square task. *Cogn. Dev.* 26, 127–141. doi:10.1016/j.cogdev.2010.12.003

Rasmussen, I. A., Antonsen, I. K., Berntsen, E. M., Xu, J., Lagopoulous, J., and Haberg, A. K. (2006). Brain activation measured using functional magnetic resonance imaging during the Tower of London task. *Acta Neurpsychiatr.* 18, 216–225. doi:10.1111/j.1601-5215.2006.00145.x

Reitan, R. M., and Wolfson, D. (1995). Category test and trail making test as measures of frontal-lobe functions. *Clin. Neuropsychol.* 9, 50–56. doi:10.1080/13854049508402057

Shallice, T. (1982). Specific impairments of planning. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 298, 199–209. doi:10.1098/rstb.1982.0082

Shum, D., Gill, H., Banks, M., Maujean, A., Griffin, J., and Ward, H. (2009). Planning ability following moderate to severe traumatic brain injury: performance on a 4-disk version of the Tower of London. *Brain Impair.* 10, 320–324. doi:10.1375/brim.10.3.320

Shum, D., Short, L., Tunstall, J., O'Gorman, J. G., Wallace, G., Shephard, K., et al. (2000). Performance of children with traumatic brain injury on a 4-disk version of the tower of London and the Porteus Maze. *Brain Cogn.* 44, 59–62.

Son, J. L., Smith, L. B., and Goldstone, R. L. (2011). Connecting instances to promote children's relational reasoning. *J. Exp. Child Psychol.* 108, 260–277. doi:10.1016/j.jecp.2010.08.011

Tunstall, J. (1999). *Improving the Utility of the Tower of London: A Neuropsychological Test of Planning.* Unpublished M. Phil. thesis, Griffith University, Brisbane, QLD.

Unterrainer, J. M., and Owen, A. M. (2006). Planning and problem solving: from neuropsychology to functional neuroimaging. *J. Physiol. Paris* 99, 308–317. doi:10.1016/j.jphysparis.2006.03.014

Viskontas, I. V., Holyoak, K. J., and Knowlton, B. J. (2005). Relational integration in older adults. *Think. Reason.* 11, 390–410. doi:10.1080/13546780542000014

Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Back-Madruga, C., and McPherson, S. (2004). Relational integration and executive function in Alzheimer's disease. *Neuropsychology* 18, 296–305. doi:10.1037/0894-4105.18.2.296

Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., de Menezes Santos, M., et al. (1999). A system for relational reasoning in human prefrontal cortex. *Psychol. Sci.* 10, 119–125. doi:10.1111/1467-9280.00118

Ward, G., and Allport, A. (1997). Planning and problem solving using the five-disk Tower of London task. *Q. J. Exp. Psychol.* 50A, 49–78. doi:10.1080/027249897392224

Zielinski, T. A., Goodwin, G. P., and Halford, G. S. (2010). Complexity of categorical syllogisms: an integration of two metrics. *Eur. J. Cogn. Psychol.* 22, 391–421. doi:10.1080/09541440902830509

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

Received: 16 September 2014; accepted: 08 December 2014; published online: 23 December 2014.

Citation: Andrews G, Halford GS, Chappell M, Maujean A and Shum DHK (2014) Planning following stroke: a relational complexity approach using the Tower of London. Front. Hum. Neurosci. ***8****:1032. doi: 10.3389/fnhum.2014.01032*

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

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 09 February 2015
doi: 10.3389/fnhum.2015.00055

Hemispheric differences in relational reasoning: novel insights based on an old technique

Michael S. Vendetti [1]*† ***, Elizabeth L. Johnson*** [1,2]† ***, Connor J. Lemos*** [2] ***and Silvia A. Bunge*** [1,2]

[1] *Helen Wills Neuroscience Institute, University of California at Berkeley, Berkeley, CA, USA*
[2] *Department of Psychology, University of California at Berkeley, Berkeley, CA, USA*

Edited by:
Jérôme Prado, Centre National de la Recherche Scientifique, France

Reviewed by:
Glenda Andrews, Griffith University, Australia
James Kenneth Kroger, New Mexico State University, USA

****Correspondence:***
Michael S. Vendetti, Helen Wills Neuroscience Institute, University of California at Berkeley, 134 Barker Hall, Berkeley, CA 94720, USA
e-mail: m.vendetti@berkeley.edu

† Shared first authorship.

Relational reasoning, or the ability to integrate multiple mental relations to arrive at a logical conclusion, is a critical component of higher cognition. A bilateral brain network involving lateral prefrontal and parietal cortices has been consistently implicated in relational reasoning. Some data suggest a preferential role for the left hemisphere in this form of reasoning, whereas others suggest that the two hemispheres make important contributions. To test for a hemispheric asymmetry in relational reasoning, we made use of an old technique known as visual half-field stimulus presentation to manipulate whether stimuli were presented briefly to one hemisphere or the other. Across two experiments, 54 neurologically healthy young adults performed a visuospatial transitive inference task. Pairs of colored shapes were presented rapidly in either the left or right visual hemifield as participants maintained central fixation, thereby isolating initial encoding to the contralateral hemisphere. We observed a left-hemisphere advantage for encoding a series of ordered visuospatial relations, but both hemispheres contributed equally to task performance when the relations were presented out of order. To our knowledge, this is the first study to reveal hemispheric differences in relational encoding in the intact brain. We discuss these findings in the context of a rich literature on hemispheric asymmetries in cognition.

Keywords: reasoning, hemispheric specialization, deductive, transitive inference

INTRODUCTION

Relational reasoning is a cognitive process that requires the joint consideration of relations in order to generate an inference to support a conclusion. Although there is a wide range of theoretical models for relational reasoning (for review, see Goodwin and Johnson-Laird, 2005; Knowlton et al., 2012), all of these models present relational reasoning as a unitary system. However, work from neuropsychological and neuroimaging literatures indicates that some cognitive functions may be supported by multiple, redundant systems in the brain (Roser and Gazzaniga, 2004; Marinsek et al., 2014). Here, we sought to test whether one hemisphere displays an advantage over the other during relational encoding, or whether this function can be carried out equally well by each hemisphere.

Hints of a possible left-hemisphere advantage in relational reasoning have emerged over the course of a number of neuroimaging experiments (e.g., Goel and Dolan, 2004; Green et al., 2006; Bunge et al., 2009; Wendelken et al., 2011). Importantly, similar patterns have been observed for tasks involving either verbal or non-linguistic/pictorial stimuli, suggesting that the observed differences are not entirely stimulus-driven and do not completely overlap with regions supporting language (Monti and Osherson, 2012). However, the conclusions we can draw from these fMRI studies about lateralization of function are limited in several ways. Namely, brain imaging provides correlational rather than causal evidence, and results depend on the specific contrasts used as well as the choice of statistical threshold. All of these factors can mask whether both hemispheres are indicated as being involved in a particular task, and thus, any conclusions about localization should converge with experimental findings using multiple approaches.

The neuropsychological literature also hints at possible hemispheric differences in contributions to reasoning. Much of the early work investigating differential hemispheric contributions to cognitive function came from work on split-brain patients (e.g., Sperry et al., 1969). These studies indicated an improved ability for hypothesis testing during problem solving in the left relative to the right hemisphere (LeDoux et al., 1977) and has led to the idea of the left hemisphere being an “interpreter” of events – i.e., the hemisphere with a major role of integrating newly acquired perceived information with previously constructed theories (Gazzaniga, 2000; Marinsek et al., 2014).

Following the seminal work of Gazzaniga et al. (1962) indicating how cognitive function differed in the two hemispheres following sectioning of the commissures, hemispheric asymmetries in cognition have alternately been characterized as a dichotomy between local and global (van Kleeck, 1989), categorical and coordinate (Kosslyn, 1987; van der Ham et al., 2014), or serial and parallel (e.g., Cohen, 1973) processes (for review, see Bradshaw and Nettleton, 1981). In the present study, we did not set out to evaluate these competing accounts of hemispheric specialization; rather, we sought to characterize the contribution of each hemisphere to performance of a relational reasoning task adapted from one used in a prior fMRI study from our group (Wendelken and Bunge, 2010).

There is not a consistent pattern relating relational reasoning ability to damage in a particular hemisphere. Neuropsychological work on relational reasoning has demonstrated the necessity of prefrontal and posterior parietal regions during transitive inference (Waltz et al., 1999; Krawczyk et al., 2008; Waechter et al., 2012), analogical reasoning (Morrison et al., 2004; Krawczyk et al., 2008), and matrix reasoning (Baldo et al., 2010; Woolgar et al., 2010). Additionally, studies employing voxel-based lesion symptom mapping to investigate relationships between patterns of brain damage and resulting cognitive deficits in fluid intelligence (Barbey et al., 2014) have suggested that damage to the right hemisphere plays a more critical role. However, Baldo et al. (2010) demonstrated that patients who have incurred strokes in the left hemisphere have been shown to also have significant deficits in a visuospatial relational reasoning task; therefore, more research is needed to provide a better understanding of each hemisphere's role in relational reasoning.

We designed the current study to test the role of each hemisphere in relational encoding through the use of a visual half-field stimulus presentation procedure. This paradigm was originally developed for use in split-brain patients, who have either minimal or no connection between the two hemispheres (e.g., Gazzaniga et al., 1962). Here, our participants were healthy adults whose hemispheres are presumed to interact closely in the coordination of task performance (Weissman and Banich, 2000). Nevertheless, we sought to test for differences in response times and/or accuracy when relational information is *initially encoded* by the left or the right hemisphere. This visual half-field stimulus presentation procedure allowed us to test whether left and right hemispheres differentially support relational encoding.

In the present study, we used a transitive inference task adapted from an fMRI task that we have used previously (Wendelken and Bunge, 2010). When reasoning using transitive inference, the logical conclusion is deduced through transferring relational inferences among terms expressed in the premises (e.g., if A > B and B > C, then A must be greater than C). On this task, shown in **Figure 1**, participants view a new set of relations on every trial and are expected to integrate them in working memory. There has been a rich literature on this form of reasoning (e.g., Halford, 1984; Cohen et al., 1997; Andrews and Halford, 1998; Greene et al., 2001). Importantly, this form of relational reasoning bears only a passing resemblance to transitive inference paradigms that involve learning paired associations over many trials (e.g., Acuna et al., 2002; Zalesak and Heckers, 2009; Koscik and Tranel, 2012; for discussion, see Wendelken and Bunge, 2010). The major difference between our transitive inference paradigm and those based on learning paired associations is that our task does not rely on remembering associations to be transferred; instead, participants must infer the spatial relationship based on the relations from the most recent trial only. Having to perform this inference anew each trial reduces any tendency to assume an object-order relationship when attempting to solve the task.

Inspired by neuropsychological research demonstrating that prefrontal patients have difficulty with transitive inference when the relations are presented out of order (e.g., "Sam is taller than Roy," "James is taller than Sam"; Waltz et al., 1999;

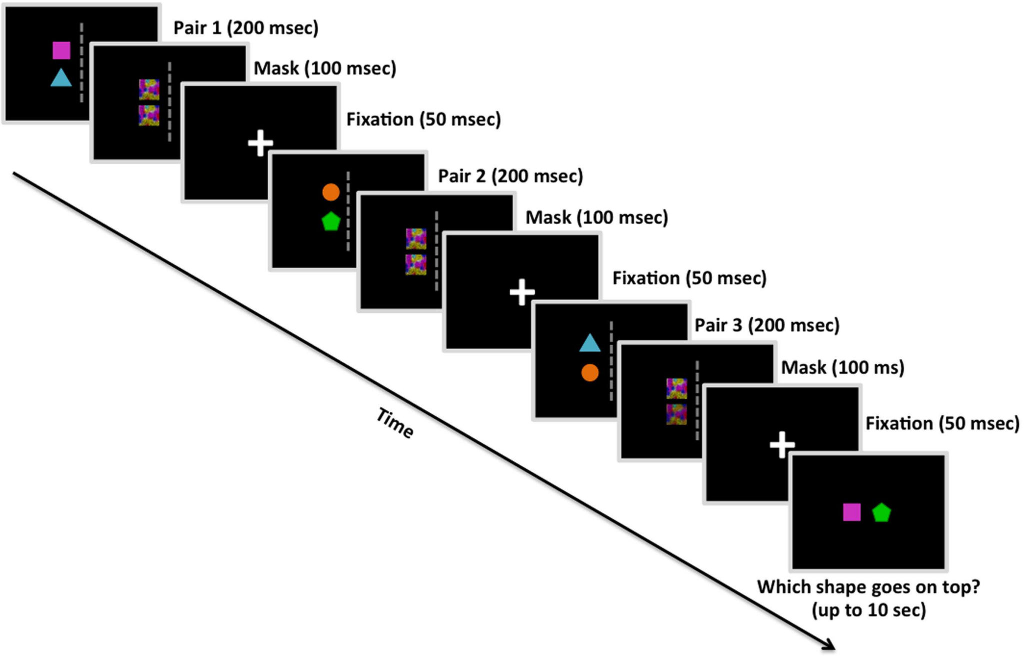


FIGURE 1 | Example trial from Study 2 (including the visual mask). Participants were shown three pairs of colored shapes. Following each pair, participants were shown a visual mask overlaying the previous shapes, and then a fixation cross. After the third pair was presented in a given trial, participants had up to 10 s to decide the correct linear order of two shapes based on the spatial relationships observed among the pairs. This is an example of a reordered trial, in which participants would presumably have to manipulate their memory of the pairs in order to deduce that the square goes on top of the pentagon. Study 1 was similar in design except for the absence of the visual mask presentations.

Krawczyk et al., 2008), we manipulated the sequence of presentation of the three relations. On half of the trials, the relations were *ordered* (A > B; B > C; C > D), and on the other half, they were *reordered* (A > B; C > D; B > C *or* C > D, A > B, B > C). We hypothesized that manipulating encoding in this manner would have an influence on the downstream integration process, and sought to test for hemispheric differences in performance on trials whose relations could be integrated readily (*ordered* trials) and those that could not (*reordered* trials).

MATERIALS AND METHODS

PARTICIPANTS

Experiment 1: Twenty-three healthy adults (14 female, aged 18–34 years; $\bar{X} \pm$ SD age, 22 ±3.08 years). *Experiment 2:* Thirty-one healthy adults (24 female, aged 18–25 years; $\bar{X} \pm$ SD age, 20 ±1.80 years). All participants attended the University of California, Berkeley, and participated in either Experiment 1 or 2 for partial fulfillment of a course requirement. All participants had normal or corrected-to-normal vision, were right-handed, and were fluent in English. Participants had no reported history of neurological or psychiatric disorders. All participants gave their informed consent to participate in the study, which was approved by the Committee for Protection of Human Subjects at the University of California, Berkeley.

DESIGN

We ran two studies with a similar design except for the addition of brief visual masks immediately following presentation of each object pair (100 ms) and an additional 48 trials, both of which were implemented in Experiment 2. We chose to insert the visual masks in Experiment 2 to reduce any after-image perceptual influences on decision making, in effect making the participant's deduction solely based on information stored and manipulated in working memory (Kim and Blake, 2005). The task designs were identical with the exception of these additions in Experiment 2; therefore, all of the information below applied to both studies unless explicitly stated. The stimulus set consisted of four colored shapes: blue triangle, orange circle, green pentagon, and pink square. On each trial, three sets of relations – pairs of shapes arranged vertically, with one colored shape positioned directly above another colored shape – were presented in sequence (**Figure 1**). One-third of the transitive inference trials involved *ordered* problems, in which the source relations were presented in order (e.g., A > B, B > C, C > D; A – D?); the other two-thirds involved *reordered* problems, in which the middle relation was presented last (e.g., A > B, C > D, B > C; A – D? *or* C > D, A > B, B > C; A – D?). Placing the middle relation last instead of the final relation of the sequence assured that participants could not rely on simple memory for the most recent pair when making their decision.

Prior to the onset of each trial, white arrows appeared coming from the four corners of the screen for 400 ms in order to direct eye gaze to the center of the screen. Trials began with a white central fixation cross displayed on screen for 50 ms. Each pair of shapes was presented in the left or right visual hemifield for 200 ms, followed by a visual mask for 100 ms (Experiment 2 only) and a central fixation inter-stimulus interval (ISI) for 50 ms, and then a different pair of shapes in either the same or opposite visual hemifield for 200 ms. After being shown three pairs individually, participants were asked to deduce the correct linear order of two items (e.g., square and pentagon) based on the spatial relations presented in the sequence of object pairs (e.g., square above triangle, triangle above circle, and circle above pentagon). Participants had ≤10 s to make their decision regarding the correct linear order of two colored shapes (i.e., which of the two objects would be on top following the spatial relations represented in the trial).

PROCEDURE

Participants placed their heads in a chinrest affixed at arm's length from the screen, and were instructed to maintain their gaze on a central fixation cross. Vertical pairs of shapes were displayed between 4° and 6° of visual angle from central fixation (Buschman et al., 2011).

In Experiment 1, the task included 96 trials total: 24 in which all three shape pairs were presented to the left hemisphere (LLL), 24 in which they were presented to the right hemisphere (RRR), 24 in which they were presented to alternating hemispheres (12 LRL and 12 RLR trials), and 24 in which they were presented to opposite hemispheres but did not alternate (12 LRR and 12 RLL trials). The LRL, RLR, LRR, and RLL trials were inserted so that participants could not reliably predict where the second and third pairs would be presented. Experiment 2 included an additional 48 trials, but the balance of trial types was consistent with Experiment 1. Trials were evenly counterbalanced by hemispheric presentation and ordering condition, and the trial order was fully randomized.

The final prompt displayed two shapes next to each other and participants were instructed to indicate via key press which shape should "go on top" based on the information in the three pairs of relations. The "z" key corresponded to the shape on the left and the "?/" key to the shape on the right; participants were instructed to keep their left hand on the "z" key and right hand on the "?/" key throughout the trials. In half the trials, the correct answer appeared on the left and half on the right. Participants were given a short break at the mid-point of the task. Experiment 2 contained a third block of trials, so participants were given a second break.

RESULTS

FULLY LATERALIZED TRIALS

We first investigated whether the small differences in task design between Experiments 1 and 2 would lead to any reliable differences in the results. A three-way mixed effects analysis of variance (ANOVA) with experiment number as the between-subjects variable, and hemispheric presentation (LH versus RH) and ordering condition (ordered versus reordered) as within-subjects variables indicated neither a main effect of experiment nor any interaction with other factors, *F's* < 1, *p's* > 0.54. Thus, all subsequent reported effects were generated from models collapsing across studies[1]. We analyzed accuracy and response time data in separate two-way repeated measures ANOVAs, with hemispheric

[1] Including gender as a factor in the full model, we found that the males in this study were more accurate than the females. Given the large gender imbalance in our relatively small sample, this result should not be over-interpreted. Notably, both males and females exhibited higher accuracy when the relations were presented to the left hemisphere than to the right hemisphere.

presentation and ordering condition as within-subjects factors. In this first section, we discuss only those trials that were solely presented to the left or right hemisphere. Behavioral results are presented in **Figure 2**.

The ANOVA revealed a significant main effect of hemisphere on accuracy, $F(1, 53) = 27.15$, $MSE = 0.012$, $p < 0.01$, $\eta^2_{\text{partial}} = 0.34$, such that participants performed better when relational information in the reasoning problem was initially encoded by the left hemisphere ($\bar{X} = 0.76$, $SD = 0.17$) as compared to the right hemisphere ($\bar{X} = 0.68$, $SD = 0.16$). A significant interaction between hemispheric presentation and ordering condition was also observed, $F(1, 53) = 8.2$, $MSE = 0.013$, $p < 0.01$, $\eta^2_{\text{partial}} = 0.13$. *Post hoc t*-tests using Bonferroni correction showed that participants were significantly more accurate when ordered pairs were presented to the left hemisphere ($\bar{X} = 0.79$, $SD = 0.19$) as compared to the right hemisphere ($\bar{X} = 0.66$, $SD = 0.16$), $t(53) = 6.02, p < 0.001, \eta^2_{\text{partial}} = 0.41$. By contrast, no significant differences were found in accuracy between the left hemisphere ($\bar{X} = 0.74, SD = 0.17$) and right hemisphere ($\bar{X} = 0.70, SD = 0.19$) on reordered trials, $t(53) = 1.51$, $p > 0.13$, $\eta^2_{\text{partial}} = 0.04$. We could also describe this interaction by looking at differences between trial types within each hemisphere. Although neither of these comparisons passed Bonferroni correction, in the left hemisphere, performance on ordered trials was better than on reordered trials, whereas the opposite was true in the right hemisphere. These results suggest that, although performance was best when stimuli were presented in order to the left hemisphere, both hemispheres performed similarly when relations were not presented in an order that is conducive to integration before solving the transitive inference problem.

When including response times from correctly performed trials as the dependent variable, the ANOVA produced a marginally significant effect of hemispheric presentation, such that participants were faster to produce the correct decision on trials that were presented to the left hemisphere ($\bar{X} = 1218.41$, $SD = 433.56$) as compared to the right hemisphere ($\bar{X} = 1273.10$, $SD = 448.26$), $F(1, 53) = 3.93$, $MSE = 41115.21$, $p = 0.053$, $\eta^2_{\text{partial}} = 0.07$. No other effects in relation to response time were found to be statistically significant, F's $< 1.26, p$'s > 0.26. These results suggest that the left-hemisphere boost in performance was not due to a speed-accuracy tradeoff; rather, when object pairs were presented to the left hemisphere, participants tended to respond faster than they would have if information had been presented to the right hemisphere.

ALL TRIALS

In this section, we describe analyses investigating performance across both fully lateralized and mixed hemisphere trials (**Figure 3**). We ran 4 × 2 repeated measures ANOVAs with number of times in the left hemisphere (0, 1, 2, 3) and order (*ordered* versus *reordered*) as within-subject factors, predicting accuracy and response time scores in separate models.

No significant effects were found for response times, F's < 1.8, p's > 0.18. In terms of accuracy, we found a significant main effect of number of times in the left hemisphere, $F(3,159) = 8.79$, $MSE = 0.013$, $p < 0.001$, $\eta^2_{\text{partial}} = 0.14$, such that greater accuracy was observed the more often premises were presented in the left hemisphere. We also observed a trend for the effect of order, such that accuracy on ordered trials ($\bar{X} = 0.74$, $SD = 0.16$) was marginally higher than on reordered trials ($\bar{X} = 0.72$, $SD = 0.15$), $F(1,53) = 3.45$, $MSE = 0.017$, $p < 0.07$, $\eta^2_{\text{partial}} = 0.06$. We observed a significant interaction between number of times in the left hemisphere by order, $F(3, 159) = 5.55$, $MSE = 0.013$, $p < 0.001$, $\eta^2_{\text{partial}} = 0.1$. We found that for ordered trials there was a significant monotonic increase in accuracy as premises were presented to the left hemisphere, $F(1, 53) = 38.11$, $MSE = 0.011$,

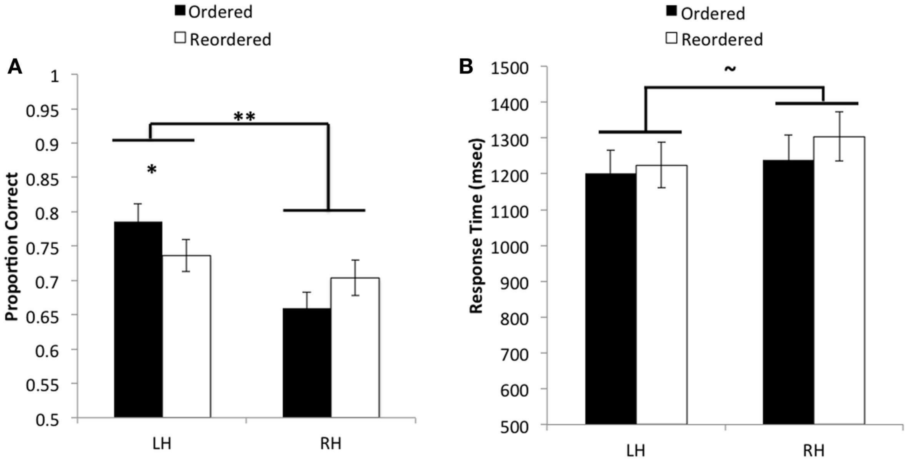


FIGURE 2 | (A) Average proportion correct as a function of hemisphere and ordering condition. A significant interaction was found such that when pairs of objects were presented in order, performance was significantly better when information was initially presented to the left versus the right hemisphere. However, no reliable difference was observed between hemispheres when pairs needed to be reordered in memory. Additionally, an overall main effect was found indicating that accuracy improved when pairs were initially encoded by the left hemisphere as opposed to the right hemisphere. **(B)** Average response time in milliseconds as a function of hemisphere and ordering condition, for correct trials. No reliable differences were observed for response time. $^{**}p < 0.01$.

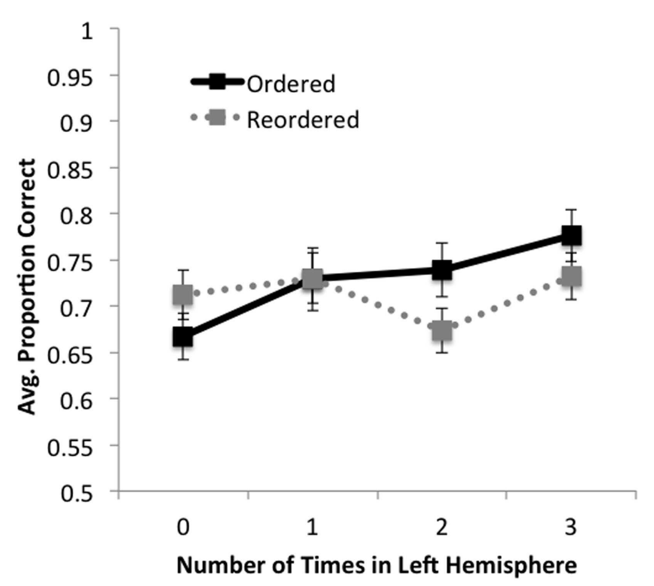


FIGURE 3 | Accuracy as a function of ordering condition and number of times premise was presented in the left hemisphere (0, 1, 2, 3). For ordered trials, accuracy increased monotonically with the number of times a premise was presented in the left hemisphere. For reordered trials, a simple pattern was not observed; rather, accuracy decreased when premises were presented in the left hemisphere two times (i.e., on LRL and RLL trials) relative to one or three times. No effects were observed for response times.

$p < 0.001, \eta^2_{\text{partial}} = 0.42$. For reordered trials, no such linear trend was observed, $F(1,53) < 1, p > 0.5$. These results suggest that when information is already ordered, increases in accuracy can be significantly predicted by how many times the premises are presented in the left hemisphere, and support our finding that participants performed better when ordered trials were presented only to the left hemisphere than to the right.

FOLLOW-UP ANALYSES

In testing for hemispheric differences in performance on this transitive inference task, we sought to ensure that participants were performing this task in the manner expected. When three relations are presented in order, it is possible to produce the correct response even without integrating multiple relations (Bryant and Trabasso, 1971). In our design, this simpler, non-integrative strategy could be undertaken by paying attention only to the top item in the first premise rather than encoding all premises and integrating the relations between them. If participants were to take this strategy, they would be expected to achieve roughly 100% accuracy on ordered trials, but only around 50% accuracy on reordered trials (because the first item of the first premise only appeared in the final prompt on two-thirds of the trials). Six out of 54 participants exhibited a pattern consistent with the use of this strategy. The findings reported here hold even when excluding these six participants.

DISCUSSION

Inspired by findings from the neuroimaging and neuropsychological literatures, we tested whether healthy young adults' performance on a reasoning task would differ on whether the stimuli were presented to the left or right hemisphere. By designing a transitive inference task with visual half-field stimulus presentation, we were able to show differences in reasoning performance as a function of the hemisphere that initially encoded the sets of visuospatial relations. Given that the two hemispheres communicate freely in the intact brain, we had expected only modest differences in response times for left- versus right-hemifield stimulus presentation. As such, we were surprised by the magnitude of the behavioral difference elicited by visual half-field presentation in this study, with an average difference in accuracy of 11% between left-lateralized and right-lateralized ordered trials. Although claims of inter-hemispheric differences in cognition have been made for many years (Gazzaniga et al., 1962; Cohen, 1973), our study is the first to demonstrate hemispheric differences in relational encoding in neurologically intact participants.

Although task performance (i.e., accuracy) improved overall when participants encoded the visuospatial relations in the left hemisphere, this effect was driven by performance on the ordered trials. That is, we observed a left-hemisphere advantage when the relations were ordered linearly and, therefore, could be integrated directly, but not when it was necessary to rearrange the relations before integrating them. For right-hemisphere trials, participants did not show the predicted pattern of worse performance for reordered versus ordered trials. This pattern was unexpected, and warrants further investigation. Surprisingly, given that reordered trials are hypothesized to require additional processing relative to ordered trials (Waltz et al., 1999; Krawczyk et al., 2008), left-hemisphere encoding of *reordered* relations was superior even to right-hemisphere encoding of *ordered* relations. These results suggest that the left hemisphere excels at relational encoding.

The present results fit well with neuroimaging studies that have pointed toward a left-hemisphere specialization in relational reasoning (Wendelken et al., 2008; Bunge et al., 2009; Green et al., 2010). In light of these findings, it is interesting to consider a recent resting-state functional connectivity study showing that the left-hemisphere interacts more exclusively with itself, whereas the right hemisphere demonstrates connectivity patterns associated with both hemispheres (Gotts et al., 2013). This result suggests that the left hemisphere may operate independently, whereas the right hemisphere functions, at least partly, with assistance from the left hemisphere. Given these findings, we would predict a left-hemisphere advantage if relational encoding hinges more on intra-hemispheric interactions, and indeed this prediction was supported by our analysis including the mixed trials.

A LEFT-HEMISPHERE ADVANTAGE FOR RELATIONAL ENCODING

The behavioral improvement observed in our study does not indicate that the right hemisphere cannot encode relational information, but rather suggests that relational encoding may be processed more effectively in the left hemisphere. Although the stimuli were visuospatial in nature, they nonetheless were easily identifiable verbally (e.g., circle, square, pentagon). Given how quickly premises were presented, it does not seem feasible that very many participants would have had enough time to verbally label objects while they solved the task; however, we cannot conclusively rule out this possibility. The present study establishes a paradigm that could be used for further examination of the necessity of verbal labeling for relational reasoning.

Numerous dichotomies have been used to explain hemispheric asymmetries in cognitive functioning (Bradshaw and Nettleton, 1981), and so we do not claim that the left-hemisphere advantage observed in our study is unique to relational encoding, *per se.* Beyond the verbal/non-verbal distinction (Gazzaniga et al., 1962), other theories have focused on local versus global (van Kleeck, 1989), serial versus parallel (Cohen, 1973), holistic versus analytic (Nebes, 1978; Cooper and Wojan, 2000), categorical versus coordinate (Kosslyn, 1987), or syntactical versus intuitive/"gist" (Bogen, 1975; Phelps and Gazzaniga, 1992) processing, to name a few. Such dichotomies are useful in that they demonstrate how a higher level cognitive task such as reasoning might be represented as a combination of lower order cognitive processes. Our transitive inference task could be construed as being syntactical, serial, and analytic, and previous work focusing on these distinctions has consistently demonstrated a left-hemispheric specialization (for review, see Bradshaw and Nettleton, 1981). Additionally, encoding spatial relations in the premises categorically (e.g., identifying the square as above the triangle) would also fit with previous work demonstrating a left-hemispheric advantage for categorical encoding of spatial relations (Kosslyn, 1987; van der Ham et al., 2012).

CONCLUSION AND FUTURE DIRECTIONS

Our results shed light on cognitive theories of relational reasoning, as they provide evidence for differential processing of relations by the two hemispheres. Specifically, we found that participants performed better on our transitive inference task when the premises were presented to the left hemisphere. This effect was driven by an interaction such that there was a greater difference in performance when the premises were ordered than when participants presumably had to reorder the premises before making their conclusion. Theories describing a unitary mechanism of relational reasoning (e.g., Hummel and Holyoak, 2003; Goodwin and Johnson-Laird, 2005) may need to incorporate multiple components in order to fully represent interhemispheric differences used during relational reasoning.

The present results are consistent with theoretical predictions concerning hemispheric specialization of cognitive functions. Specifically, participants are expected to perform better when information is presented to the left hemisphere for tasks that could be solved using a stepwise and analytical strategy. Our findings extend previous work given that our transitive inference task not only exemplifies these types of strategies but also relies on the comparison of relational information between premises in order to arrive at a solution.

These behavioral results warrant further investigation with neuroscientific techniques. First, functional imaging techniques could be used to measure the dynamic interplay between hemispheres during performance of this lateralized transitive inference task. Second, transcranial direct current stimulation could be used to increase or reduce cortical excitability within a hemisphere and test whether relational reasoning performance in each hemisphere changes as a function of cortical excitability (Nitsche and Paulus, 2001; Ardolino et al., 2005). Finally, patients with unilateral brain injuries could be tested on this lateralized task to assess whether relational encoding is primarily a left-hemisphere function, or whether the right hemisphere could specialize in this function after left-hemisphere damage. Thus, reapplying this well-established stimulus presentation procedure in these multiple contexts will help us to better understand the underlying mechanisms required for processing relational information during reasoning.

ACKNOWLEDGMENTS

This work was made possible by a James S. McDonnell Foundation Scholar Award to Silvia A. Bunge. We thank Farida Valji and Kiana Modavi for their assistance with data collection.

REFERENCES

Acuna, B. D., Eliassen, J. C., Donoghue, J. P., and Sanes, J. N. (2002). Frontal and parietal lobe activation during transitive inference in humans. *Cereb. Cortex* 12, 1312–1321. doi:10.1093/cercor/12.12.1312

Andrews, G., and Halford, G. S. (1998). Children's ability to make transitive inferences: the importance of premise integration and structural complexity. *Cogn. Dev.* 13, 479–513. doi:10.1016/S0885-2014(98)90004-1

Ardolino, G., Bossi, B., Barbieri, S., and Priori, A. (2005). Non-synaptic mechanisms underlie the after-effects of cathodal transcutaneous direct current stimulation of the human brain. *J. Physiol.* 568, 653–663. doi:10.1113/jphysiol.2005.088310

Baldo, J. V., Bunge, S. A., Wilson, S. M., and Dronkers, N. F. (2010). Is relational reasoning dependent on language? A voxel-based lesion symptom mapping study. *Brain Lang.* 113, 59–64. doi:10.1016/j.bandl.2010.01.004

Barbey, A. K., Colom, R., Paul, E. J., and Grafman, J. (2014). Architecture of fluid intelligence and working memory revealed by lesion mapping. *Brain Struct. Funct.* 219, 485–494. doi:10.1007/s00429-013-0512-z

Bogen, J. E. (1975). *Educational Aspects of Hemispheric Specialization.* UCLA Educator.

Bradshaw, J. L., and Nettleton, N. C. (1981). The nature of hemispheric specialization in man. *Behav. Brain Sci.* 4, 51–91. doi:10.1017/S0140525X00007548

Bryant, P. E., and Trabasso, T. (1971). Transitive inferences and memory in young children. *Nature* 232, 456–458. doi:10.1038/232456a0

Bunge, S. A., Helskog, E. H., and Wendelken, C. (2009). Left, but not right, rostrolateral prefrontal cortex meets a stringent test of the relational integration hypothesis. *Neuroimage* 46, 338–342. doi:10.1016/j.neuroimage.2009.01.064

Buschman, T. J., Siegel, M., Roy, J. E., and Miller, E. K. (2011). Neural substrates of cognitive capacity limitations. *Proc. Natl. Acad. Sci. U. S. A.* 108, 11252–11255. doi:10.1073/pnas.1104666108

Cohen, G. (1973). Hemispheric differences in serial versus parallel processing. *J. Exp. Psychol.* 97, 349–356. doi:10.1037/h0034099

Cohen, N. J., Poldrack, R. A., and Eichenbaum, H. (1997). Memory for items and memory for relations in the procedural/declarative memory framework. *Memory* 5, 131–178. doi:10.1080/741941149

Cooper, E. E., and Wojan, T. J. (2000). Differences in the coding of spatial relations in face identification and basic-level object recognition. *J. Exp. Psychol. Learn. Mem. Cogn.* 26, 470–488. doi:10.1037/0278-7393.26.2.470

Gazzaniga, M., Bogen, J., and Sperry, R. (1962). Some functional effects of sectioning the cerebral commissures in man. *Proc. Natl. Acad. Sci. U. S. A.* 48, 1765–1769. doi:10.1073/pnas.48.10.1765

Gazzaniga, M. S. (2000). Cerebral specialization and interhemispheric communication: does the corpus callosum enable the human condition? *Brain* 123, 1293–1326. doi:10.1093/brain/123.7.1293

Goel, V., and Dolan, R. J. (2004). Differential involvement of left prefrontal cortex in inductive and deductive reasoning. *Cognition* 93, B109–B121. doi:10.1016/j.cognition.2004.03.001

Goodwin, G. P., and Johnson-Laird, P. N. (2005). Reasoning about the relations between relations. *Q. J. Exp. Psychol.* 59, 1047–1069. doi:10.1080/02724980543000169

Gotts, S. J., Jo, H. J., Wallace, G. L., Saad, Z. S., Cox, R. W., and Martin, A. (2013). Two distinct forms of functional lateralization in the human brain. *Proc. Natl. Acad. Sci. U. S. A.* 110, E3435–E3444. doi:10.1073/pnas.1302581110

Green, A. E., Fugelsang, J. A., Kraemer, D. J. M., Shamosh, N. A., and Dunbar, K. N. (2006). Frontopolar cortex mediates abstract integration in analogy. *Brain Res.* 1096, 125–137. doi:10.1016/j.brainres.2006.04.024

Green, A. E., Kraemer, D. J. M., Fugelsang, J. A., Gray, J. R., and Dunbar, K. N. (2010). Connecting long distance: semantic distance in analogical reasoning modulates frontopolar cortex activity. *Cereb. Cortex* 20, 70–76. doi:10.1093/cecor/bhp081

Greene, A. J., Spellman, B. A., Dusek, J. A., and Eichenbaum, H. B. (2001). Relational learning with and without awareness: transitive inference using nonverbal stimuli in humans. *Mem. Cognit.* 29, 893–902. doi:10.3758/BF03196418

Halford, G. S. (1984). Can young children integrate premises in transitivity and serial order tasks? *Cogn. Psychol.* 16, 65–93. doi:10.1016/0010-0285(84)90004-5

Hummel, J. E., and Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychol. Rev.* 110, 220–264. doi:10.1037/0033-295X.110.2.220

Kim, C. Y., and Blake, R. (2005). Psychophysical magic: rendering the visible 'invisible'. *Trends Cogn. Sci.* 9, 381–388. doi:10.1016/j.tics.2005.06.012

Knowlton, B. J., Morrison, R. G., Hummel, J. E., and Holyoak, K. J. (2012). A neurocomputational system for relational reasoning. *Trends Cogn. Sci.* 16, 373–381. doi:10.1016/j.tics.2012.06.002

Koscik, T. R., and Tranel, D. (2012). The human ventromedial prefrontal cortex is critical for transitive inference. *J. Cogn. Neurosci.* 24, 1191–1204. doi:10.1162/jocn_a_00203

Kosslyn, S. M. (1987). Seeing and imagining in the cerebral hemispheres: a computational approach. *Psychol. Rev.* 94, 148–175. doi:10.1037/0033-295X.94.2.148

Krawczyk, D. C., Morrison, R. G., Viskontas, I., Holyoak, K. J., Chow, T. W., Mendez, M. F., et al. (2008). Distraction during relational reasoning: the role of prefrontal cortex in interference control. *Neuropsychologia* 46, 2020–2032. doi:10.1016/j.neuropsychologia.2008.02.001

LeDoux, J. E., Risse, G. L., Springer, S. P., Wilson, D. H., and Gazzaniga, M. S. (1977). Cognition and commissurotomy. *Brain* 100, 87–104. doi:10.1093/brain/100.1.87

Marinsek, N., Turner, B. O., Gazzaniga, M., and Miller, M. B. (2014). Divergent hemispheric reasoning strategies: reducing uncertainty versus resolving inconsistency. *Front. Hum. Neurosci.* 8:839. doi:10.3389/fnhum.2014.00839

Monti, M. M., and Osherson, D. N. (2012). Logic, language and the brain. *Brain Res.* 1428, 33–42. doi:10.1016/j.brainres.2011.05.061

Morrison, R. G., Krawczyk, D. C., Holyoak, K. J., Hummel, J. E., Chow, T. W., Miller, B. L., et al. (2004). A neurocomputational model of analogical reasoning and its breakdown in frontotemporal lobar degeneration. *J. Cogn. Neurosci.* 16, 260–271. doi:10.1162/089892904322984553

Nebes, R. D. (1978). "Direct examination of cognitive function in the right and left hemispheres," in *Asymmetrical Function of the Brain*, ed. M. Kinsbourne (Cambridge: University Press), 99–140.

Nitsche, M. A., and Paulus, W. (2001). Sustained excitability elevations induced by transcranial DC motor cortex stimulation in humans. *Neurology* 57, 1899–1901. doi:10.1212/WNL.57.10.1899

Phelps, E. A., and Gazzaniga, M. S. (1992). Hemispheric differences in mnemonic processing: the effects of left hemisphere interpretation. *Neuropsychologia* 30, 293–297. doi:10.1016/0028-3932(92)90006-8

Roser, M., and Gazzaniga, M. S. (2004). Automatic brains–interpretive minds. *Curr. Dir. Psychol. Sci.* 13, 56–59. doi:10.1111/j.0963-7214.2004.00274.x

Sperry, R. W., Gazzaniga, M. S., and Bogen, J. E. (1969). "Interhemispheric relationships: the neocortical commissures; syndromes of hemisphere disconnection," in *Handbook of Clinical Neurology*, eds P. J. Vinken and G. W. Bruvn (Amsterdam: North-Holland Publishing Company), 273–290.

van der Ham, I. J. M., Postma, A., and Laeng, B. (2014). Lateralized perception: the role of attention in spatial relation processing. *Neurosci. Biobehav. Rev.* 45, 142–148. doi:10.1016/j.neubiorev.2014.05.006

van der Ham, I. J. M., van Wezel, R. J. A., Oleksiak, A., van Zandvoort, M. J. E., Frijns, C. J. M., Kappelle, L. J., et al. (2012). The effect of stimulus features on working memory of categorical and coordinate spatial relations in patients with unilateral brain damage. *Cortex* 38, 737–745. doi:10.1016/j.cortex.2011.03.002

van Kleeck, M. H. (1989). Hemispheric differences in global versus local processing of hierarchical visual stimuli by normal subjects: new data and a meta-analysis of previous studies. *Neuropsychologia* 27, 1165–1178. doi:10.1016/0028-3932(89)90099-7

Waechter, R. L., Goel, V., Raymont, V., Kruger, F., and Grafman, J. (2012). Transitive inference reasoning is impaired by focal lesions in parietal cortex rather than rostrolateral prefrontal cortex. *Neuropsychologia* 51, 464–471. doi:10.1016/j.neuropsychologia.2012.11.026

Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., de Meneze Santos, M., et al. (1999). A system for relational reasoning in human prefrontal cortex. *Psychol. Sci.* 10, 119–125. doi:10.1111/1467-9280.00118

Weissman, D. H., and Banich, M. T. (2000). The cerebral hemispheres cooperate to perform complex but not simple tasks. *Neuropsychology* 14, 41–59. doi:10.1037/0894-4105.14.1.41

Wendelken, C., and Bunge, S. A. (2010). Transitive inference: distinct contributions of rostrolateral prefrontal cortex an the hippocampus. *J. Cogn. Neurosci.* 22, 837–847. doi:10.1162/jocn.2009.21226

Wendelken, C., Chung, D., and Bunge, S. A. (2011). Rostrolateral prefrontal cortex: domain-general or domain-sensitive? *Hum. Brain Mapp.* 33, 1952–1963. doi:10.1002/hbm.21336

Wendelken, C., Nakhabenko, D., Donohue, S. E., Carter, C. S., and Bunge, S. A. (2008). "Brain is to thought as stomach is to ??": investigating the role of rostrolateral prefrontal cortex in relational reasoning. *J. Cogn. Neurosci.* 20, 682–693. doi:10.1162/jocn.2008.20055

Woolgar, A., Parr, A., Cusack, R., Thompson, R., Nimmo-Smith, I., Torralva, T., et al. (2010). Fluid intelligence loss linked to restricted regions of damage within frontal and parietal cortex. *Proc. Natl. Acad. Sci. U. S. A.* 107, 14899–14902. doi:10.1073/pnas.1007928107

Zalesak, M., and Heckers, S. (2009). The role of the hippocampus in transitive inference. *Psychiatry Res.* 30, 24–30. doi:10.1016/j.pscychresns.2008.09.008

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: 20 January 2015; published online: 09 February 2015.

Citation: Vendetti MS, Johnson EL, Lemos CJ and Bunge SA (2015) Hemispheric differences in relational reasoning: novel insights based on an old technique. Front. Hum. Neurosci. ***9****:55. doi: 10.3389/fnhum.2015.00055*

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

frontiers
in Human Neuroscience

ORIGINAL RESEARCH
published: 14 December 2015
doi: 10.3389/fnhum.2015.00659

Modulation of Neural Activity in the Temporoparietal Junction with Transcranial Direct Current Stimulation Changes the Role of Beliefs in Moral Judgment

Hang Ye[1], *Shu Chen*[1], *Daqiang Huang*[1], *Haoli Zheng*[1], *Yongmin Jia*[1] *and Jun Luo*[2,3]*

[1] *College of Economics, Interdisciplinary Center for Social Sciences at Zhejiang University, Hangzhou, China,* [2] *School of Economics and International Trade, Zhejiang University of Finance and Economics, Hangzhou, China,* [3] *Neuro and Behavior EconLab, Zhejiang University of Finance and Economics, Hangzhou, China*

Judgments about whether an action is morally right or wrong typically depend on our capacity to infer the actor's beliefs and the outcomes of the action. Prior neuroimaging studies have found that mental state (e.g., beliefs, intentions) attribution for moral judgment involves a complex neural network that includes the temporoparietal junction (TPJ). However, neuroimaging studies cannot demonstrate a direct causal relationship between the activity of this brain region and mental state attribution for moral judgment. In the current study, we used transcranial direct current stimulation (tDCS) to transiently alter neural activity in the TPJ. The participants were randomly assigned to one of three stimulation treatments (right anodal/left cathodal tDCS, left anodal/right cathodal tDCS, or sham stimulation). Each participant was required to complete two similar tasks of moral judgment before receiving tDCS and after receiving tDCS. We studied whether tDCS to the TPJ altered mental state attribution for moral judgment. The results indicated that restraining the activity of the right temporoparietal junction (RTPJ) or the left the temporoparietal junction (LTPJ) decreased the role of beliefs in moral judgments and led to an increase in the dependance of the participants' moral judgments on the action's consequences. We also found that the participants exhibited reduced reaction times both in the cases of intentional harms and attempted harms after receiving right cathodal/left anodal tDCS to the TPJ. These findings inform and extend the current neural models of moral judgment and moral development in typically developing people and in individuals with neurodevelopmental disorders such as autism.

Keywords: theory of mind, moral judgment, beliefs and outcomes, temporoparietal junction, transcranial direct current stimulation

OPEN ACCESS

Edited by:
Vinod Goel,
York University, Canada

Reviewed by:
Indrajeet Patil,
Internazionale Superiore di Studi Avanzati, Neuroscience, Italy
Marine Buon,
University College London, UK

Correspondence:
Jun Luo
luojun_zju@hotmail.com

Received: *11 May 2015*
Accepted: *19 November 2015*
Published: *14 December 2015*

Citation:
Ye H, Chen S, Huang D, Zheng H, Jia Y and Luo J (2015) Modulation of Neural Activity in the Temporoparietal Junction with Transcranial Direct Current Stimulation Changes the Role of Beliefs in Moral Judgment. Front. Hum. Neurosci. 9:659. doi: 10.3389/fnhum.2015.00659

INTRODUCTION

In everyday life, a harm caused by an action is morally worse than an equivalent harm caused by omission, and a harm intended as the means to a goal is morally worse than an equivalent harm foreseen as the side effect of a goal (Cushman et al., 2006; Young and Koenigs, 2007). Moral judgment entails judging others' actions on the dimension of right and wrong, but this requires not only the outcomes of these actions but also the cognitive ability to think about another person's beliefs and intentions, which is known as "theory of mind" (Young and Saxe, 2008).

A number of recent studies indeed demonstrate that mental state information (e.g., desire, belief, intention) is one of the crucial inputs into moral decision-making (for a review, see Young and Tsoi, 2013). Evidence from developmental psychology also shows that children (even preverbal infants) start condemning negative intent that does not result in negative outcome (see Baird and Astington, 2004; Killen et al., 2011). But when beliefs and outcomes are incongruent with each other, there are different ways that this incongruence can behaviorally present itself relying on the valence of the conflicting belief and outcome (Patil and Silani, 2014).

Cushman (2008) found that judgments of punishment depended jointly on mental states and the causal relationship of an agent to a harmful consequence. An account of these phenomena has been proposed that distinguished two processes of moral judgment (Young et al., 2007; Cushman et al., 2013; Cushman, 2013): one which begins with harmful outcome and attributes condemnation to the causally responsible agent, and the other which begins with an action and analyses the mental states responsible for that action.

Neuroimaging studies have investigated the selectivity and domain specificity of these brain regions for thinking about another person's thoughts. These regions, which comprise the "theory of mind network, " include the medial prefrontal cortex (MPFC), precuneus (PC), right superior temporal sulcus (RSTS), and bilateral temporal-parietal junction (TPJ; Gallagher et al., 2000; Vogeley et al., 2001; Ruby and Decety, 2003; Saxe and Kanwisher, 2003; Aichhorn et al., 2009).

The precise role of these brain regions in theory of mind for moral judgment has been the topic of recent researches (Young et al., 2007; Young and Saxe, 2008). Specifically, the TPJ exhibits increased activity whenever participants read about a person's beliefs in nonmoral (Saxe and Kanwisher, 2003; Saxe and Powell, 2006) or moral contexts (Young et al., 2007, 2010b). However, fMRI cannot demonstrate direct causal relationships between the activities in these brain regions and mental state attribution for moral judgment.

Noninvasive brain stimulation techniques, such as rTMS, allow for the study of the decision consequences of externally restrained brain activity in healthy participants and thus the establishment of causal connections between the brain and decisions without many of the confounds inherent to natural lesion studies (Rafal, 2001; Robertson et al., 2003). Young et al. (2010a) and Jeurissen et al. (2014) used rTMS to transiently suppress activity in the right temporoparietal junction (RTPJ) and provided evidence for the causal role of this structure in mental state attribution for moral judgment.

Transcranial direct current stimulation (tDCS) has some advantages relative to rTMS because it induces a stronger modulatory effect on brain activity (Nitsche and Paulus, 2000; Romero et al., 2002), allowing for reliable sham stimulation (Gandiga et al., 2006). Importantly, anodal tDCS increases excitability in targeted brain regions, which can transiently enhance decisions and judgment in healthy humans (Fregni et al., 2005; Wassermann and Grafman, 2005).

The goal of the present study was to alter moral judgments by modulating the cortical excitability over the TPJ in healthy adults. To measure the participants' capacities to infer the actor's mental state attributions in moral judgment, we presented the participants with moral scenarios in which (i) the protagonist acts on either a negative belief (e.g., that he or she will cause harm to another person) or on a neutral belief and (ii) the protagonist either causes a negative outcome (e.g., harm to another person) or a neutral outcome (Young et al., 2007; Young and Saxe, 2008). Participants made judgments on a scale of 1 (permissible) to 10 (forbidden), which were regarded as their condemnation ratings towards the behaviors described.

Previous findings have provided direct evidence supporting the critical role of the RTPJ in mediating belief attribution for moral judgment, For example Young et al. (2010a) revealed that the disruption of the RTPJ with TMS led participants to rely their judgments less on the actor's mental states, and Sellaro et al. (2015) found that participants who received anodal tDCS over the RTPJ assigned less blame to accidental harms compared to participants who received sham stimulation. However, a direct causal relationship between left temporoparietal junction (LTPJ) and mental state attribution for moral judgments has not been studied. In the present study, we sought to firstly test whether modulating the activity of the LTPJ activity with tDCS would also influence the role of beliefs on moral judgments. Therefore, we performed an experiment to investigate whether bilateral stimulation of the TPJ (anodal stimulation of the right and cathodal stimulation of the left TPJ or vice versa) would alter mental state attribution for moral judgments. Our findings suggested that restraining the RTPJ or LTPJ with tDCS decreased the role of beliefs in moral judgment. Combining our findings with those of previous work, we infer that the RTPJ and LTPJ commonly represent the ability to use mental states in moral judgment and that both are responsible for the role of belief in moral judgment.

Besides the difference in stimulation electrode positions from previous evidence, the present study has novel assignment for moral judgment task and classification for story context. The previous experiments demonstrated the role of the RTPJ on belief attribution by comparing participants' moral judgments following TMS to the RTPJ and TMS to a control brain region (Young et al., 2010a), or investigating participants' performance on the moral judgment task before and after having received anodal, cathodal, or sham tDCS over the RTPJ (Sellaro et al., 2015). These studies selected and randomly distributed moral stories among different treatments (including active stimulations and sham stimulation) and different tasks (pre-tDCS and post-tDCS task) to test their hypotheses. However, they haven't made sure the balance and similarity of moral stories across the treatments and tasks. In this study, each participant was required to complete a similar (and we demonstrated the similarity) moral judgment task before and after receiving tDCS. Therefore, we combined within-subject and between-subject design in this experiment to test the causal role of the bilateral TPJ regarding mental states in moral judgment.

In addition, how one should act toward another depends on whether the target is a friend, a stranger, a subordinate, or an authority (Dungan and Young, 2012). Therefore, we have assigned two different types of story context that involved economic interests and relationships with friends in moral judgment task to explore the role of TPJ on the actors' mental state attributions for moral judgment across different contexts. Analyses indicated that in conditions of neutral belief, the condemnation ratings of contexts involving economic interests were lower than those of contexts involving relationship with friends. Moreover, in conditions of negative belief with contexts involving economic interests, the condemnation ratings were lower after receiving right anodal/left cathodal tDCS. These findings indicate that the restraining effect of tDCS on the LTPJ in the role of beliefs in moral judgment depends on moral context.

MATERIALS AND METHODS

Subjects

We recruited 54 healthy college students (32 females; mean age 22.11 years, ranging from 19–30 years) to participate in our experiment. All participants were right-handed and naïve to tDCS and moral judgment tasks, and they had no history of psychiatric illness or neurological disorders. The participants were randomly assigned to receive right anodal/left cathodal tDCS over TPJ (*n* = 18, 11 females), left anodal/right cathodal tDCS over TPJ (*n* = 18, 11 females) or sham stimulation over TPJ (*n* = 18, 10 females). Each participant received 50 RMB yuan (approximately 7.995 US dollars) for their participation. Participants gave written informed consent before entering the study, which was approved by the Zhejiang University ethics committee. No participants reported any adverse side effects about pain on the scalp or headaches after the experiment.

Transcranial Direct Current Stimulation (tDCS)

tDCS was induced by two saline-soaked surface sponge electrodes (35 cm^2). Direct current was constant and delivered by a battery-driven stimulator (Multichannel noninvasive wireless tDCS neurostimulator, Starlab, Barcelona, Spain), which was controlled through a Bluetooth signal. It was adjusted to induce cortical excitability of the target area without causing any physiological damage to the participants. Various orientations of the current had various effects on the cortical excitability. Generally speaking, anodal stimulation enhances cortical excitability, whereas cathodal stimulation inhibits it (Nitsche and Paulus, 2000).

TPJ was localized with location CP5 (left) and CP6 (right) on an EEG cap laid out according to the International 10–20 System (**Figure 1A**). Participants were randomly assigned to one of the three single-blinded stimulation treatments. For right anodal/left cathodal stimulation, the anodal electrode was placed over the CP6 according to the international EEG 10–20 system, while the cathodal electrode was placed over the CP5. For left anodal/right cathodal stimulation the placement was reversed. The anodal electrode was placed over CP5 and the cathodal electrode was placed over CP6 (**Figures 1B,C**). Therefore, the target electrode (either the anode or the cathode) was centered over CP6/CP5; the return electrode was placed over CP5/CP6. The reason we chose a bifrontal electrode montage was to provide stimulation able to enhance the activity of one side of the TPJ while simultaneously diminish the other side. For sham stimulation, the procedures were totally the same but the current lasted only for the first 30 s. The participants may have felt the initial itching, but actually there was no current for the rest of the stimulation. This method of sham stimulation has been shown to be reliable (Gandiga et al., 2006). The current had an intensity of 2 mA with 15 s of ramp up and down, the safety and efficiency of which was shown in previous studies.

After the participant finished the first moral judgment task (the computer program for these tasks was written in visual C#) which was similar to Young's design (Young et al., 2010a), the laboratory assistant put a tDCS device on his/her head for

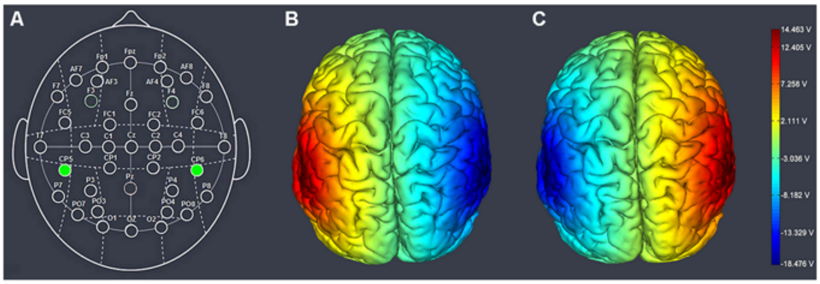


FIGURE 1 | Electrode placements. (A) Schematic of the electrode positions based on the EEG 10–20 system. **(B)** Left anodal/right cathodal stimulation over the temporoparietal junction (TPJ) of the human brain. **(C)** Right anodal/left cathodal stimulation over the TPJ of the human brain. The axis represents the range of input voltage from −18.476 to 14.463 V.

stimulation and removed him/her from the computer screen. After 15 min of stimulation, the participant was then asked to complete the latter moral judgment task with the stimulation being delivered for another 5 min (**Figure 2**).

Task and Procedure

The experiment included two moral judgment tasks. Each participant was required to complete a moral judgment task before receiving tDCS and to complete another moral judgment task after receiving tDCS. To eliminate the sequence effect of the two tasks, we randomly assigned half of the participants (Part I) to complete moral judgment task A (including story S_1 and S_2) before receiving tDCS and to complete moral judgment task B (including story S_1^* and S_2^*) after receiving tDCS; the remaining participants (Part II) completed task B before receiving tDCS and completed task A after receiving tDCS (**Figure 3**). Each story was based on a type of context that involved economic interests (S_1 and S_1^*) or relationships with friends (S_2 and S_2^*).

There were four conditions in each story that included belief (negative vs. neutral) and outcome (negative vs. neutral) factors to yield a 2 × 2 design. Specifically, they were intentional harm (negative belief and negative outcome), accidental harm (neutral belief and negative outcome), attempted harm (negative belief and neutral outcome) and nonharm (neutral belief and neutral outcome). Stories were presented in cumulative segments, each presented for 8 s, describing in a fixed order: (i) background; (ii) foreshadow; (iii) belief; and (iv) action. The background was identical across conditions. Stories were then removed from the screen and replaced with a question about the moral permissibility of the action. Participants made judgments on a scale of 1 (permissible) to 10 (forbidden) using a computer keyboard, which were regarded as their condemnation ratings towards the behaviors described. The time limit for responding was 6 s. The reaction times were recorded and all of the participants had made judgments within the time limit.

The participants were required to read and make judgments about two moral stories with four conditions respectively before receiving tDCS. After completing this moral judgment task, they had a break and received tDCS for 15 min. Subsequently, they were required to read and make judgments about another two stories with four conditions respectively while receiving stimulation for another 5 min. The latter moral judgment task was similar to the first moral judgment task to avoid learning effects in the within-subject design experiment. Both tasks included two stories (S_1 and S_1^*; S_2 and S_2^*) with four conditions (**Figure 4**). The same participant saw all four variations of the same story in both sessions, eight stories pre-stimulation and eight-stories post-stimulation, for a total of 16 stories. On average each story consisted of about 91 words, and the number of words was matched across conditions and tasks. When the subjects completed the two moral judgment tasks, they were asked to complete a questionnaire before finally receiving their payment.

Data Analysis

We first tested the similarity of tasks A and B using repeated measures analyses of variance (ANOVA). Giving the two tasks were equivalent in terms of condemnation ratings and reaction times before receiving tDCS, it ensured us to compare the performance of the participants before and after receiving tDCS. Then we used repeated measures ANOVA to test if the stimulation had changed the participants' moral judgment in different conditions, including condemnation ratings and reaction times. As we distinguished between the contexts that involved economic interests and relationships with friends, all these tests were applied firstly without consideration of the difference between the two contexts (the pooled sample) and then treating context as a within-subjects factor (sample with context). The statistical analyses were performed using SPSS statistical software (SPSS Inc., Chicago, IL, USA).

RESULTS

The Pooled Sample

The mean condemnation ratings and standard deviation information of different conditions and different stimulation types are shown in **Figure 5** and **Table 1**. We first tested

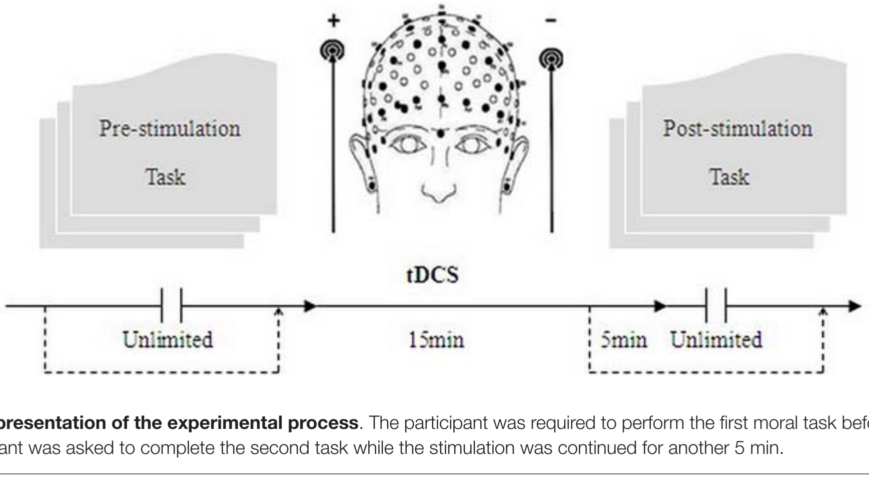


FIGURE 2 | Schematic representation of the experimental process. The participant was required to perform the first moral task before stimulation. After 15 min of stimulation, each participant was asked to complete the second task while the stimulation was continued for another 5 min.

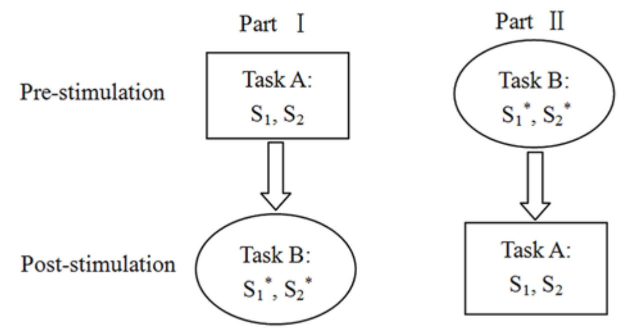


FIGURE 3 | Experimental design. The half of participants (Part I) were required to complete the moral judgment task A (S_1, S_2) before stimulation and complete the moral judgment task B (S_1^*, S_2^*) after stimulation, while the rest of participants (Part II) were required to complete task B (S_1^*, S_2^*) before stimulation and complete task A (S_1, S_2) after stimulation.

whether task A was different from task B before receiving tDCS using repeated measures ANOVA with Belief (neutral vs. negative) and Outcome (neutral vs. negative) as within-subjects factors and Task (A vs. B) as a between-subjects factor. There was significant effect of task neither in condemnation ratings [$F_{(1,106)} = 0.007$, $P = 0.931$] nor in reaction times [$F_{(1,106)} = 0.752$, $P = 0.388$], which made it reasonable to regard the two tasks as equivalent and compare the performance of the participants before and after receiving the stimulations. Meanwhile, we found significant effect of Belief [$F_{(1,106)} = 671.932$, $P < 0.001$], Outcome [$F_{(1,106)} = 419.632$, $P < 0.001$] and a significant interaction of Belief and Outcome [$F_{(1,106)} = 109.063$, $P < 0.001$] in condemnation ratings.

Since there was no significant difference between condemnation ratings and reaction times for the two moral judgment tasks, the difference before and after the stimulations could be attributed to the effect of tDCS. We ran a repeated measures ANOVA with Belief (neutral vs. negative), Outcome (neutral vs. negative) and Time (before vs. after tDCS) as within-subjects factors and stimulation type (right anodal/left cathodal, left anodal/right cathodal or sham) as a between-subjects factor. Significant effects of Belief [$F_{(1,105)} = 845.032$, $P < 0.001$] and Outcome [$F_{(1,105)} = 586.439$, $P < 0.001$] were observed, which meant that the participants' condemnation ratings of moral judgment in conditions of negative belief (mean = 8.671) were higher than that of neutral belief (mean = 4.354). Similarly, conditions of negative outcome (mean = 8.192) were more condemned than conditions of neutral outcome (mean = 4.833). Moreover, the interaction of Belief and Outcome also had a significant effect [$F_{(1,105)} = 4.454$, $P = 0.014$]. *Post hoc* analysis using bonferroni corrections indicated that conditions of intentional harm (mean = 8.755) and attempted harm (mean = 8.588) were less permissible than both conditions of accidental harm (mean = 4.398) and nonharm (mean = 4.310). We also found significant effect of stimulation type [$F_{(2,105)} = 5.289$, $P = 0.006$].

Importantly, we found a slightly significant three-way interaction involving Outcome, Time and stimulation type [$F_{(2,105)} = 3.185$, $P = 0.045$]. Analysis showed that in conditions of negative outcome, participants rated higher in condemnation after receiving right anodal/left cathodal tDCS [before: mean = 7.833; after: mean = 8.292; $P = 0.005$], especially towards intentional harm [$P = 0.001$]. On the other hand, in conditions of neutral outcome, participants rated lower in condemnation after receiving left anodal/right cathodal tDCS [before: mean = 5.764; after: mean = 5.000; $P < 0.001$], both towards attempted harm [$P = 0.001$] and nonharm [$P = 0.015$]. These findings might indicate that restraining the activity of the RTPJ/LTPJ decreased the role of beliefs in moral judgments and led to the participants' moral judgments being more dependent on the actions' consequences.

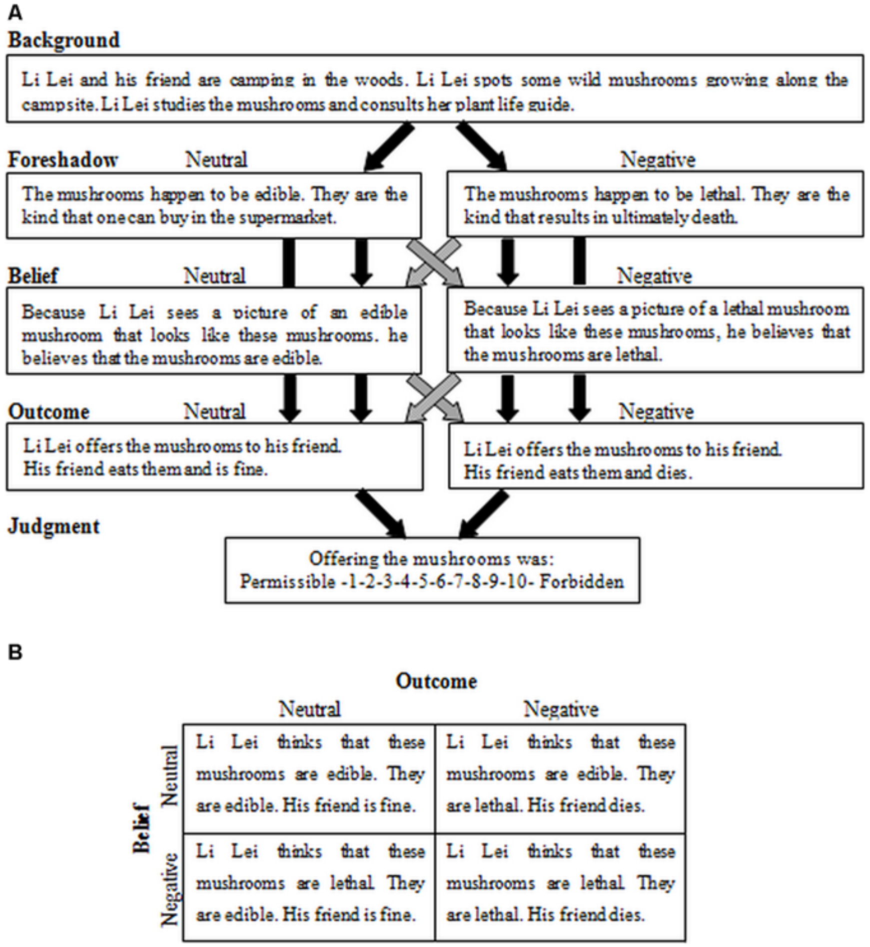
We paid attention to reaction time as well. Applying the above repeated measures ANOVA, we found a significant effect of Time [$F_{(1,105)} = 7.571$, $P = 0.007$]. It is easy to understand that the reaction times after stimulation were shorter than before because that the participants were more familiar with the task. Moreover, the three-way interaction of Belief, Time and stimulation type was trending towards significant [$F_{(2,105)} = 2.749$, $P = 0.069$]. *Post hoc* analysis indicated that the reaction times in conditions of negative belief were significantly shorter after left anodal/right cathodal tDCS [$P = 0.004$], while in conditions of neutral belief the reaction times were significantly shorter after sham stimulation [$P = 0.023$]. The mean reaction time and standard deviation information are displayed in supplementary materials.

Lastly, we checked whether the sequence of the two tasks would influence the participants' moral judgment. Repeated measures ANOVAs showed no significant effect of sequence in condemnation ratings [$F_{(1,102)} = 0.154$, $P = 0.695$] or in reaction times [$F_{(1,102)} = 1.633$, $P = 0.204$].

Sample with Context

To test the effect of context, we added Context (economic interests vs. relationships with friends) as a within-subjects factor into the repeated measures ANOVAs in section "The Pooled Sample". We first tested the similarity of tasks A and B. No significant effect of Task in condemnation ratings [$F_{(1,52)} = 0.004$, $P = 0.947$] or in reaction times [$F_{(1,52)} = 0.407$, $P = 0.526$] was observed. Apart from the significant effects of Belief [$F_{(1,52)} = 427.022$, $P < 0.001$], Outcome [$F_{(1,52)} = 254.778$, $P < 0.001$] and a significant interaction of Belief and Outcome [$F_{(1,52)} = 65.701$, $P < 0.001$] in condemnation ratings as in sections "The Pooled Sample", there was also a significant interaction of Context and Belief [$F_{(1,52)} = 7.379$, $P = 0.009$]. Analysis indicated that in conditions of neutral belief, the condemnation ratings of contexts involving economic interests were lower than those of contexts involving relationships with friends [$P = 0.010$].

We then performed repeated measures ANOVA with Context, Belief, Outcome and Time as within-subjects factors and stimulation type as a between-subjects factor. Again we found significant effects of Belief [$F_{(1,51)} = 473.717$, $P < 0.001$] and Outcome [$F_{(1,51)} = 321.762$, $P < 0.001$], which meant that the participants' ratings of moral judgment in conditions of negative belief were higher than that of neutral belief, as well as conditions



Li Lei and his friend are camping in the woods. Li Lei spots some wild mushrooms growing along the campsite. Li Lei studies the mushrooms and consults her plant life guide.

The mushrooms happen to be edible. They are the kind that one can buy in the supermarket.

The mushrooms happen to be lethal. They are the kind that results in ultimately death.

Because Li Lei sees a picture of an edible mushroom that looks like these mushrooms. he believes that the mushrooms are edible.

Because Li Lei sees a picture of a lethal mushroom that looks like these mushrooms, he believes that the mushrooms are lethal.

Li Lei offers the mushrooms to his friend. His friend eats them and is fine.

Li Lei offers the mushrooms to his friend. His friend eats them and dies.

Offering the mushrooms was:
Permissible -1-2-3-4-5-6-7-8-9-10- Forbidden

Belief	Outcome: Neutral	Outcome: Negative
Neutral	Li Lei thinks that these mushrooms are edible. They are edible. His friend is fine.	Li Lei thinks that these mushrooms are edible. They are lethal. His friend dies.
Negative	Li Lei thinks that these mushrooms are lethal. They are edible. His friend is fine.	Li Lei thinks that these mushrooms are lethal. They are lethal. His friend dies.

FIGURE 4 | Task design and experimental stimuli. (A) Schematic representation of sample scenario. Light-colored arrows mark the combinations of "Foreshadow" and "Belief" for which the belief is false. "Foreshadow" information foreshadows whether the action will result in a neutral or negative outcome. "Belief" information states whether the protagonist holds a belief that he is in a neutral situation and that action will result in a neutral outcome (neutral belief) or a belief that he is a negative situation and that action (or inaction) will result in a negative outcome (negative belief). **(B)** Combination of belief (neutral vs. negative) and outcome (neutral vs. negative) factors yielded a 2 × 2 design with four conditions.

of negative outcome were more condemned than conditions of neutral outcome. The interaction of Belief and Outcome also had a significant effect [$F_{(1,51)} = 65.255$, $P < 0.001$]. In addition, we found significant effects of Context [$F_{(1,51)} = 5.391$, $P = 0.024$], which meant that contexts involving economic interests [mean = 6.419] was less condemned than those of contexts involving relationships with friends [mean = 6.606]. Besides, there was a significant four-way interaction involving Context, Belief, Time and stimulation type [$F_{(2,51)} = 3.871$, $P = 0.027$]. Analysis indicated that in conditions of negative belief with contexts involving economic interests, the condemnation ratings were lower after receiving right anodal/left cathodal tDCS [$p = 0.014$]. There was also a similar but slightly less significant effect in conditions of negative belief with contexts involving economic interests [$P = 0.069$].

As for the reaction time, we found a significant effect of Time [$F_{(1,51)} = 4.517$, $P = 0.038$] similar to section "The Pooled Sample". A significant four-way interaction of Context, Belief, Outcome and stimulation type was also observed [$F_{(2,51)} = 3.908$, $P = 0.026$], indicating that in conditions of accidental harm, the reaction times of contexts involving economic interests were longer than those of contexts involving relationships with friends in sham stimulation [$P = 0.013$]. The mean reaction time and standard deviation information are displayed in supplementary materials. At last, no significant effect of sequence was observed in condemnation ratings [$F_{(1,48)} = 0.083$, $P = 0.774$] or in reaction times [$F_{(1,48)} = 0.855$, $P = 0.360$].

DISCUSSION

Human moral judgment often represents a response that depends on various factors and features that include not only the agent's beliefs but also the agent's desires (Cushman, 2008), their consequences (Greene et al., 2001), the agent's prior record (Kliemann et al., 2008), the cause that leads to harm (Cushman et al., 2008), whether the action was coerced by

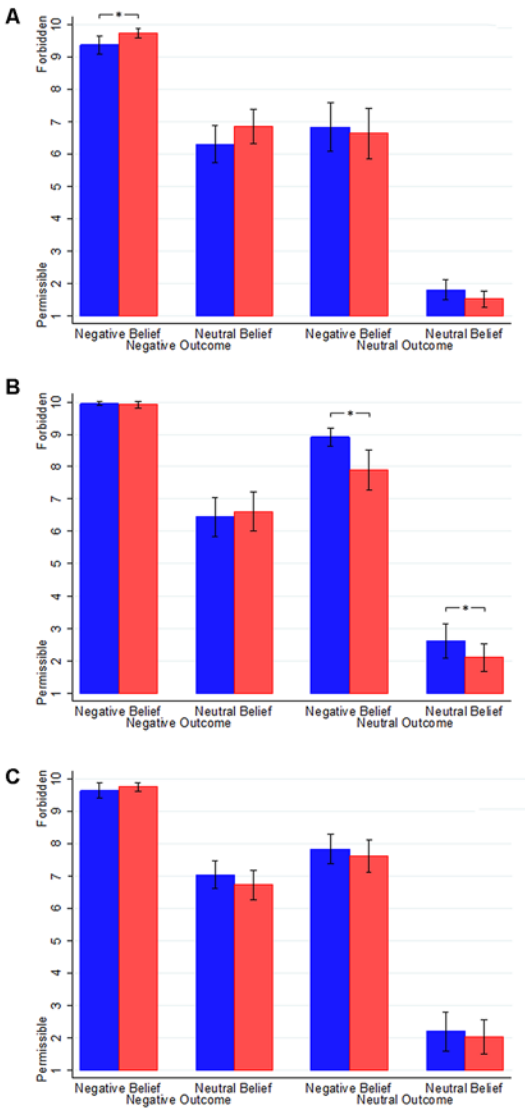


FIGURE 5 | Data of condemnation ratings. (A) The condemnation ratings of participants with four conditions in the right anodal/left cathodal treatment before and after stimulation. **(B)** The condemnation ratings of participants with four conditions in the left anodal/right cathodal treatment before and after stimulation. **(C)** The condemnation ratings of participants with four conditions in the sham treatment before and after stimulation. Blue columns, pre-transcranial direct current stimulation (tDCS); red columns, post-tDCS. Error bars indicate 95% confidence intervals. Asterisks indicate statistical significance of difference within-subject.

external circumstances (Woolfolk et al., 2006; Krebs et al., 2014), (etc., Valdesolo and DeSteno, 2006; Young et al., 2010a). In the present study, we manipulated two of these factors, the agent's belief and the outcomes of the action, and tested whether the effect of modulating activity in the TPJ with tDCS was specific to the agent's mental state attribution for moral judgment.

This study corroborated and complemented the previous finding by Young et al. (2010a), which postulated that disrupting RTPJ function reduces the influence of beliefs on moral judgment. We found that restraining the RTPJ via tDCS caused the participants to judge attempted harms and nonharm as less morally forbidden and more morally permissible, while restraining the LTPJ via tDCS caused the participants to judge accidental harms and intentional harms as more morally forbidden and less morally permissible. Thus, suppressing the activity in the RTPJ or LTPJ disrupted the capacity to use mental states in moral judgment.

To verify the robustness of our results, we modified a related experimental design based on that of Young et al. (2010a). Previous neurostimulation experiments of human decision-making have primarily utilized between-subject design (Knoch et al., 2006; Fecteau et al., 2007a,b; Boggio et al., 2010; Young et al., 2010a). However, the corresponding results lack statistical power due to the heterogeneity of the participants, especially when the samples are small. Our experiment adopted a within-subject design to avoid this interference from the heterogeneity of the participants. Provided that the multiple exposures are independent, this design makes it possible for causal estimates to be obtained by examining how individual decisions change after receiving stimulation.

Furthermore, the previous studies haven't made sure the balance of moral stories across the treatments. In this study, each participant was required to complete similar moral judgment task before and after receiving tDCS (active stimulations and sham stimulation). We also demonstrated that task A was equivalent to task B before receiving tDCS either in terms of condemnation ratings or reaction times, which made it reasonable to compare the performance of the participants before and after receiving the stimulations. Since there was no significant difference between the two moral judgment tasks, the difference before and after the stimulations could be attributed to the effect of tDCS.

Generally, moral judgments are robust to different demographic factors such as gender, age, ethnicity, and religion, but many complexities in moral judgment are still left unresolved. No comprehensive model or taxonomy of moral judgment thus far has accounted for its full diversity. Some models call for a division of the moral space based on the content, and there is work going one on about the role of intentions as a function of the moral content (Shweder et al., 1997; Rozin et al., 1999; Dungan and Young, 2012). This content-based approach also proves fruitful in explaining different emotional responses to different kinds of moral violations. Specifically, there is evidence that individuals have made difference for moral judgment between stranger and friend (Ma, 1989; Smetana et al., 2006; Kurzban et al., 2012).

To consider the context effect on both participants' condemnation ratings and the effects of tDCS for TPJ, we have assigned two different types of moral context that involved friend relationships (harm to her/his friend) and economic interests (harm to her/his customer)—as food-safety problems in China have contributed to a rapid decline of social trust (Yan, 2012)—as stories of moral judgment and separately tested whether the modulation of activity in the TPJ with tDCS changed the agents' mental state attributions for moral judgment in both the friend

TABLE 1 | The mean condemnation ratings and SD across conditions and stimulation types.

Condition	R Anodal/L Cathodal		L Anodal/R Cathodal		Sham	
	Before	After	Before	After	Before	After
Intentional harm	9.36 (0.99)	9.72 (0.51)	9.94 (0.23)	9.92 (0.37)	9.64 (0.83)	9.75 (0.5)
Accidental harm	6.31 (2.01)	6.86 (1.90)	6.44 (2.12)	6.11 (2.16)	7.03 (1.52)	6.72 (1.61)
Attempted harm	6.83 (2.70)	6.64 (2.75)	8.92 (0.97)	7.89 (2.23)	7.83 (1.61)	7.61 (1.82)
Nonharm	1.81 (1.06)	1.53 (0.88)	2.61 (1.89)	2.11 (1.53)	2.19 (2.15)	2.03 (1.9)

relationship and economic interest contexts. Analyses indicated that in conditions of neutral belief, the condemnation ratings of contexts involving economic interests were lower than those of contexts involving relationship with friends. Moreover, in conditions of negative belief with contexts involving economic interests, the condemnation ratings were lower after receiving right anodal/left cathodal tDCS. These findings indicate that the restraining effect of tDCS on the LTPJ in the role of beliefs in moral judgment depends on moral context.

The present study also investigated the participants' reaction times for moral judgments and found that the participants who received restraint of the RTPJ exhibited reduced reaction times in both the cases of intentional harms and attempted harms when the story involved economic interests. Because restraining the RTPJ significantly decreased the capacity to infer the actor's intentions in moral judgment, the participants could easily make judgments that primarily considered the attribution of action's consequence when the role of belief in moral judgment was reduced.

Many studies have shown that both the RTPJ and the LTPJ play essential roles in the theory of mind and that the activities of these two brain regions are associated with the understanding of social intentions (Ciaramidaro et al., 2007; Sommer et al., 2007; Aichhorn et al., 2009; Centelles et al., 2011). Recent fMRI studies have also suggested that the bilateral TPJ are recruited for the encoding and integrating process of beliefs (Young and Saxe, 2008). Specifically, Young et al. (2010a) used TMS to the RTPJ to disrupt the capacity to integrate belief information. Samson et al. (2004) reported evidence from brain-damaged patients that indicated that the patients with lesions in the LPTJ region exhibit impairment in false belief tasks.

In the present study, we also found that restraining the RTPJ or LTPJ via tDCS decreased the role of beliefs in moral judgment. Combining our findings with those of previous work, we infer that the RTPJ and LTPJ commonly represent the capacity to use mental states in moral judgment and that both are responsible for the role of belief in moral judgment. After receiving tDCS to restrain the activities of the RTPJ or LTPJ, the role of beliefs in moral judgment is reduced. In the four conditions of moral stories, the participants placed more weight on the attribution of the action's consequences but not on intentions in moral judgment. Specifically, after restraining the activity of the TPJ, participants judged intentional harms and accidental harms as more morally forbidden and less morally permissible, and the participants judged attempted harms and nonharm as less morally forbidden and more morally permissible. These effects might also depend on stories' context of moral judgment.

In conclusion, our findings provide important information about the effects of tDCS on mental states in moral judgment. These findings might be helpful for the study and treatment of neurodevelopmental disorders, such as autism spectrum disorders (ASDs). Children with ASDs are unable to impute beliefs to others (Baron-Cohen et al., 1985). Even high functioning adults with ASDs have a persistent impairment in spontaneous mentalizing (i.e., the automatic ability to attribute mental states to the self and others; Senju et al., 2009). Furthermore, the impairment in the processing of the mental states of others in autism is associated with reduced RTPJ activity (Kana et al., 2009). Therefore, we believe that this study might inform neural models of moral judgment and moral development in typically developing people and in individuals with neurodevelopmental disorders such as autism (Koster-Hale et al., 2013).

Additionally, both folk moral judgments and legal decisions depend on agent's ability to make judgment for the consequences of an individual's actions to the beliefs and intentions of actions. Our experiments revealed that the mental state attribution of moral judgment, especially in cases involving attempted harm and accidental harm, depends critically on neural activity in the TPJ. Future studies should explore the relevance of these findings for the real-life judgments made by judges and juries who routinely make very detailed distinctions based on mental state information.

Since the same participant saw all four variations of the same story during the experiment, we acknowledged this design may increase demand characteristics for the task as participants could figure out the differences of four conditions. However, we aimed to study whether tDCS to the TPJ (active stimulation treatments) altered mental state attribution for moral judgment. Therefore, the possibility of those demand characteristics which were perceived by the participants would not lead to biased experimental results. In addition, it was noted that the robustness of the current findings across diverse moral contexts remained to be determined because of the limited number of stimuli used in the experiment.

Another limitation of the present study is that we were unable to determine whether the effect on mental state attribution of moral judgment was solely attributable to the modulation of the activity in the RTPJ or whether the changes in moral judgment resulted from altering the balance of activity across the bilateral TPJ. With regard to the tDCS polarity effects, Jacobson et al. (2012) conducted a meta-analytical review aimed to investigate the homogeneity/heterogeneity of the effect sizes of the anodal-excitation and cathodal-inhibition effects dichotomy in both motor and cognitive functions. They found that the anode electrode is applied over a cognitive area, in most cases, it will cause an excitation as measured by a relevant cognitive task. However, the cathodal-inhibition effects seems to be robust only in the motor and sensory cortex but there is wide variation for cognitive studies. Therefore, our finding that the influence of modulating activity in the bilateral TPJ with tDCS on the role of beliefs in moral judgment, to a large extent, may resulted from anodal-excitation effects, rather than cathodal-inhibition effects. Future experiments may include neuroimaging measures to explore the neural changes associated with the neuromodulation that lead to decision-making effects and also to explore other paradigms of stimulation, such as unilateral stimulation.

AUTHOR CONTRIBUTIONS

HY, SC, DH, HZ, YJ and JL designed experiment; DH, HZ, JL performed experiment; SC analyzed data; HY drew figures; SC, DH, HZ and JL wrote the manuscript.

ACKNOWLEDGMENT

This work was supported by the National Social Science Fund, China (Grant number: 13AZD061, 15ZDB134).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fnhum.2015.00659/abstract

REFERENCES

Aichhorn, M., Perner, J., Weiss, B., Kronbichler, M., Staffen, W., and Ladurner, G. (2009). Temporo-parietal junction activity in theory-of-mind tasks: falseness, beliefs, or attention. *J. Cogn. Neurosci.* 21, 1179–1192. doi: 10.1162/jocn.2009.21082

Baird, J. A., and Astington, J. W. (2004). The role of mental state understanding in the development of moral cognition and moral action. *New Dir. Child Adolesc. Dev.* 2004, 37–49. doi: 10.1002/cd.96

Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition* 21, 37–46. doi: 10.1016/0010-0277(85)90022-8

Boggio, P. S., Campanhã, C., Valasek, C. A., Fecteau, S., Pascual-Leone, A., and Fregni, F. (2010). Modulation of decision-making in a gambling task in older adults with transcranial direct current stimulation. *Eur. J. Neurosci.* 31, 593–597. doi: 10.1111/j.1460-9568.2010.07080.x

Centelles, L., Assaiante, C., Nazarian, B., Anton, J. L., and Schmitz, C. (2011). Recruitment of both the mirror and the mentalizing networks when observing social interactions depicted by point-lights: a neuroimaging study. *PLoS One* 6:e15749. doi: 10.1371/journal.pone.0015749

Ciaramidaro, A., Adenzato, M., Enrici, I., Erk, S., Pia, L., Bara, B. G., et al. (2007). The intentional network: how the brain reads varieties of intentions. *Neuropsychologia* 45, 3105–3113. doi: 10.1016/j.neuropsychologia.2007.05.011

Cushman, F. (2008). Crime and punishment: distinguishing the roles of causal and intentional analysis in moral judgment. *Cognition* 108, 353–380. doi: 10.1016/j.cognition.2008.03.006

Cushman, F. (2013). Action, outcome and value a dual-system framework for morality. *Pers. Soc. Psychol. Rev.* 17, 273–292. doi: 10.1177/1088868313495594

Cushman, F., Knobe, J., and Sinnott-Armstrong, W. (2008). Moral appraisals affect doing/allowing judgments. *Cognition* 108, 281–289. doi: 10.1016/j.cognition.2008.02.005

Cushman, F., Sheketoff, R., Wharton, S., and Carey, S. (2013). The development of intent-based moral judgment. *Cognition* 127, 6–21. doi: 10.1016/j.cognition.2012.11.008

Cushman, F., Young, L., and Hauser, M. D. (2006). The role of conscious reasoning and intuitions in moral judgment: testing three principles of harm. *Psychol. Sci.* 17, 1082–1089. doi: 10.1111/j.1467-9280.2006.01834.x

Dungan, J., and Young, L. (2012). "Moral psychology," in *A Companion to Moral Anthropology*, Vol. 32, ed. D. Fassin (Chichester: John Wiley & Sons, Ltd.), 578–594.

Fecteau, S., Knoch, D., Fregni, F., Sultani, N., Boggio, P. S., and Pascual-Leone, A. (2007a). Diminishing risk-taking behavior by modulating activity in the prefrontal cortex: a direct current stimulation study. *J. Neurosci.* 27, 12500–12505. doi: 10.1523/jneurosci.3283-07.2007

Fecteau, S., Pascual-Leone, A., Zald, D. H., Liguori, P., Théoret, H., Boggio, P. S., et al. (2007b). Activation of prefrontal cortex by transcranial direct current stimulation reduces appetite for risk during ambiguous decision making. *J. Neurosci.* 27, 6212–6218. doi: 10.1523/jneurosci.0314-07.2007

Fregni, F., Boggio, P. S., Nitsche, M., Bermpohl, F., Antal, A., Feredoes, E., et al. (2005). Anodal transcranial direct current stimulation of prefrontal cortex enhances working memory. *Exp. Brain Res.* 166, 23–30. doi: 10.1007/s00221-005-2334-6

Gallagher, H. L., Happé, F., Brunswick, N., Fletcher, P. C., Frith, U., and Frith, C. D. (2000). Reading the mind in cartoons and stories: an fMRI study of 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia* 38, 11–21. doi: 10.1016/s0028-3932(99)00053-6

Gandiga, P. C., Hummel, F. C., and Cohen, L. G. (2006). Transcranial DC stimulation(tDCS) a tool for double-blind sham-controlled clinical studies in brain stimulation. *Clin. Neurophysiol.* 117, 845–850. doi: 10.1016/j.clinph.2005.12.003

Greene, D., Sommerville, B., Nystrom, E., Darley, M., and Cohen, D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science* 293, 2105–2108. doi: 10.1126/science.1062872

Jacobson, L., Koslowsky, M., and Lavidor, M. (2012). tDCS polarity effects in motor and cognitive domains: a meta-analytical review. *Exp. Brain Res.* 216, 1–10. doi: 10.1007/s00221-011-2891-9

Jeurissen, D., Sack, A. T., Roebroeck, A., Russ, B. E., and Pascual-Leone, A. (2014). TMS affects moral judgment, showing the role of DLPFC and TPJ in cognitive and emotional processing. *Front. Neurosci.* 8:18. doi: 10.3389/fnins.2014.00018

Kana, R. K., Keller, T. A., Cherkassky, V. L., Minshew, N. J., and Just, M. A. (2009). Atypical frontal posterior synchronization of theory of mind regions in autism during mental state attribution. *Soc. Neurosci.* 4, 135–152. doi: 10.1080/17470910802198510

Killen, M., Lynn Mulvey, K., Richardson, C., Jampol, N., and Woodward, A. (2011). The accidental transgressor: morally-relevant theory of mind. *Cognition* 119, 197–215. doi: 10.1016/j.cognition.2011.01.006

Kliemann, D., Young, L., Scholz, J., and Saxe, R. (2008). The influence of prior record on moral judgment. *Neuropsychologia* 46, 2949–2957. doi: 10.1016/j.neuropsychologia.2008.06.010

Knoch, D., Gianotti, L. R., Pascual-Leone, A., Treyer, V., Regard, M., Hohmann, M., et al. (2006). Disruption of right prefrontal cortex by low-frequency repetitive transcranial magnetic stimulation induces risk-taking behavior. *J. Neurosci.* 26, 6469–6472. doi: 10.1523/jneurosci.0804-06.2006

Koster-Hale, J., Saxe, R., Dungan, J., and Young, L. (2013). Decoding moral judgments from neural representations of intentions. *Proc. Natl. Acad. Sci. U S A* 110, 5648–5653. doi: 10.1073/pnas.1207992110

Krebs, D. L., Vermeulen, S., Carpendale, J., and Denton, K. (2014). "Structural and situational influences on moral judgment: the interaction between stage and dilemma," in *The Handbook of Moral Behavior and Development*, eds W. Kurtines and J. Gewirtz (New York: Psychology Press), 139–169.

Kurzban, R., DeScioli, P., and Fein, D. (2012). Hamilton vs. Kant: pitting adaptations for altruism against adaptations for moral judgment. *Evol. Hum. Behav.* 33, 323–333. doi: 10.1016/j.evolhumbehav.2011.11.002

Ma, H. K. (1989). Moral orientation and moral judgment in adolescents in Hong Kong, Mainland China and England. *J. Cross Cult. Psychol.* 20, 152–177. doi: 10.1177/0022022189202003

Nitsche, A., and Paulus, W. (2000). Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. *J. Physiol.* 527, 633–639. doi: 10.1111/j.1469-7793.2000.t01-1-00633.x

Patil, I., and Silani, G. (2014). Alexithymia increases moral acceptability of accidental harms. *J. Cogn. Psychol.* 26, 597–614. doi: 10.1080/20445911.2014.929137

Rafal, R. (2001). "Bálint's syndrome," in *The Handbook of Neuropsychology*, Vol. 4, ed. M. Behrmann (Amsterdam: Elsevier Science), 121–141.

Robertson, E. M., Théoret, H., and Pascual-Leone, A. (2003). Studies in cognition: the problems solved and created by transcranial magnetic stimulation. *J. Cogn. Neurosci.* 15, 948–960. doi: 10.1162/089892903770007344

Romero, J. R., Anschel, D., Sparing, R., Gangitano, M., and Pascual-Leone, A. (2002). Subthreshold low frequency repetitive transcranial magnetic stimulation selectively decreases facilitation in the motor cortex. *Clin. Neurophysiol.* 113, 101–107. doi: 10.1016/s1388-2457(01)00693-9

Rozin, P., Lowery, L., Imada, S., and Haidt, J. (1999). The CAD triad hypothesis: a mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *J. Pers. Soc. Psychol.* 76, 574–586. doi: 10.1037/0022-3514.76.4.574

Ruby, P., and Decety, J. (2003). What you believe versus what you think they believe: a neuroimaging study of conceptual perspective-taking. *Eur. J. Neurosci.* 17, 2475–2480. doi: 10.1046/j.1460-9568.2003.02673.x

Samson, D., Apperly, I. A., Chiavarino, C., and Humphreys G. W. (2004). Left temporoparietal junction is necessary for representing someone else's belief. *Nat. Neurosci.* 7, 499–500. doi: 10.1038/nn1223

Saxe, R., and Kanwisher, N. (2003). People thinking about thinking people: the role of the temporo-parietal junction in "theory of mind". *Neuroimage* 19, 1835–1842. doi: 10.1016/s1053-8119(03)00230-1

Saxe, R., and Powell, L. (2006). It's the thought that counts: specific brain regions for one component of theory of mind. *Psychol. Sci.* 17, 692–699. doi: 10.1111/j.1467-9280.2006.01768.x

Sellaro, R., Güroğlu, B., Nitsche, M. A., van den Wildenberg, W. P., Massaro, V., Durieux, J., et al. (2015). Increasing the role of belief information in moral judgments by stimulating the right temporoparietal junction. *Neuropsychologia* 77, 400–408. doi: 10.1016/j.neuropsychologia.2015.09.016

Senju, A., Southgate, V., White, S., and Frith, U. (2009). Mind blind eyes: an absence of spontaneous theory of mind in asperger syndrome. *Science* 325, 883–885. doi: 10.1126/science.1176170

Shweder, R. A., Much, N. C., Mahapatra, M., and Park, L. (1997). "The "big three" of morality (autonomy, community and divinity) and the "big three" explanations of suffering," in *Morality and Health*, eds A. Brandt and P. Rozin (New York: Routledge), 119–169.

Smetana, J. G., Campione-Barr, N., and Metzger, A. (2006). Adolescent development in interpersonal and societal contexts. *Annu. Rev. Psychol.* 57, 255–284. doi: 10.1146/annurev.psych.57.102904.190124

Sommer, M., Döhnel, K., Sodian, B., Meinhardt, J., Thoermer, C., and Hajak, G. (2007). Neural correlates of true and false belief reasoning. *Neuroimage* 35, 1378–1384. doi: 10.1016/j.neuroimage.2007.01.042

Valdesolo, P., and DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychol. Sci.* 17, 476–477. doi: 10.1111/j.1467-9280.2006.01731.x

Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happé, F., Falkai, P., et al. (2001). Mind reading: neural mechanisms of theory of mind and self-perspective. *Neuroimage* 14, 170–181. doi: 10.1006/nimg.2001.0789

Wassermann, E. M., and Grafman, J. (2005). Recharging cognition with DC brain polarization. *Trends Cogn. Sci.* 9, 503–505. doi: 10.1016/j.tics.2005.09.001

Woolfolk, R. L., Doris, J. M., and Darley, J. M. (2006). Identification, situational constraint and social cognition: studies in the attribution of moral responsibility. *Cognition* 100, 283–301. doi: 10.1016/j.cognition.2005.05.002

Yan, Y. (2012). Food safety and social risk in contemporary China. *J. Asian Stud.* 71, 705–729. doi: 10.1017/s0021911812000678

Young, L., Camprodon, J. A., Hauser, M., Pascual-Leone, A., and Saxe, R. (2010a). Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proc. Natl. Acad. Sci. U S A* 107, 6753–6758. doi: 10.1073/pnas.0914826107

Young, L., Dodell-Feder, D., and Saxe, R. (2010b). What gets the attention of the temporo-parietal junction? An fMRI investigation of attention and theory of mind. *Neuropsychologia* 48, 2658–2664. doi: 10.1016/j.neuropsychologia.2010.05.012

Young, L., Cushman, F., Hauser, M., and Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proc. Natl. Acad. Sci. U S A* 104, 8235–8240. doi: 10.1073/pnas.0701408104

Young, L., and Koenigs, M. (2007). Investigating emotion in moral cognition: a review of evidence from functional neuroimaging and neuropsychology. *Br. Med. Bull.* 84, 69–79. doi: 10.1093/bmb/ldm031

Young, L., and Saxe, R. (2008). The neural basis of belief encoding and integration in moral judgment. *Neuroimage* 40, 1912–1920. doi: 10.1016/j.neuroimage.2008.01.057

Young, L., and Tsoi, L. (2013). When mental states matter, when they don't and what that means for morality. *Soc. Pers. Psychol. Compass* 7, 585–604. doi: 10.1111/spc3.12044

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.

ORIGINAL RESEARCH
published: 30 June 2015
doi: 10.3389/fnhum.2015.00360

Left inferior-parietal lobe activity in perspective tasks: identity statements

Aditi Arora[1, 2]*, *Benjamin Weiss*[1, 2], *Matthias Schurz*[1, 2], *Markus Aichhorn*[1, 2], *Rebecca C. Wieshofer*[1, 2] and *Josef Perner*[1, 2]

[1] *Department of Psychology, University of Salzburg, Salzburg, Austria,* [2] *Center for Neurocognitive Research, University of Salzburg, Salzburg, Austria*

We investigate the theory that the left inferior parietal lobe (IPL) is closely associated with tracking potential differences of perspective. Developmental studies find that perspective tasks are mastered at around 4 years of age. Our first study, meta-analyses of brain imaging studies shows that perspective tasks specifically activate a region in the left IPL and precuneus. These tasks include processing of false belief, visual perspective, and episodic memory. We test the location specificity theory in our second study with an unusual and novel kind of perspective task: identity statements. According to Frege's classical logical analysis, identity statements require appreciation of modes of presentation (perspectives). We show that identity statements, e.g., "the tour guide is also the driver" activate the left IPL in contrast to a control statements, "the tour guide has an apprentice." This activation overlaps with the activations found in the meta-analysis. This finding is confirmed in a third study with different types of statements and different comparisons. All studies support the theory that the left IPL has as one of its overarching functions the tracking of perspective differences. We discuss how this function relates to the bottom-up attention function proposed for the bilateral IPL.

Keywords: identity, false belief, episodic memory, visual perspective taking, fMRI, IPL, overarching function

OPEN ACCESS

Edited by:
Ira Andrew Noveck,
Centre Nationale de la Recherche
Scientifique, France

Reviewed by:
Matteo Feurra,
National Research University, Russia
Maria Spychalska,
Ruhr University Bochum, Germany
Erica Cosentino,
Ruhr University Bochum,
Germany/University of Messina, Italy

Correspondence:
Aditi Arora,
Center for Neurocognitive Research
and Department of Psychology,
University of Salzburg, Hellbrunnerstr.
34, 5020 Salzburg, Austria
aditi.arora@sbg.ac.at

Received: *11 August 2014*
Accepted: *03 June 2015*
Published: *30 June 2015*

Citation:
Arora A, Weiss B, Schurz M, Aichhorn M, Wieshofer RC and Perner J (2015) Left inferior-parietal lobe activity in perspective tasks: identity statements. Front. Hum. Neurosci. 9:360. doi: 10.3389/fnhum.2015.00360

Introduction

There is growing evidence that the dorsal part of the left temporo-parietal junction (TPJ), which overlaps with the left inferior parietal lobe (IPL), is reliably activated by perspective tasks (Goel et al., 1995; Ruby and Decety, 2003). Perspective tasks are tasks that require tracking of (potential or actual) perspective differences[1]. Findings from cognitive development indicate that these tasks share a common cognitive basis. They are mastered around the age of 4 years. Brain imaging

[1] With "tracking perspective differences" or, for short, "perspective tracking" we want to merely grasp the existence of this concept required for registering an actual or potential conflict between perspectives. The more common term "perspective taking" suggests the ability to put oneself into another perspective than the perspective one currently has. This would require the tracking of a particular perspective not just the tracking of a potential perspective difference. For, one can be aware of perspectives being involved without being able to switch between them. One can be aware that another person has or may have a different perspective without actually being able to figure out what that perspective is.

Abbreviations: +IDENT, identity condition; −IDENT, control of identity condition; +REVISION, belief revision condition; −REVISION, control of belief revision condition; IDENTc, identity-with-context; PREDc, predication-with-context; C, context-only; IDENTo, identity only; BL, baseline condition; FB, false belief; vPT, visual perspective taking; EM, episodic memory.

studies of perspective tasks also point to a common neural basis. Existing evidence suggests regional specificity (Kanwisher, 2010) of different kinds of perspective tasks activating the left IPL[2] . Our aim is to test this specificity hypothesis in three steps. In the first step we carry out a meta-analysis of existing data from three different kinds of perspective tasks to test the regional specificity hypothesis. Partial activation overlap of the different kinds of tasks within left IPL counts in favor of the hypothesis. In the second step we test the hypothesis further with the prediction that a novel and unusual perspective task, processing identity statements, should activate within the region identified by the meta-analysis. In a third step we confirm this finding with novel stimulus material. To carry through with this project we need to be more specific about what perspective tasks are and about the criteria that define the region of overlap, for which we adopt the overarching view proposed by Cabeza et al. (2012).

What are Perspective Tasks?

In response to this question we follow the intuition elaborated by Perner et al. (2003), who links the notion of perspective to the notion of representation and modes of presentation. A representation represents something (object, target) as being in a certain way (content). The content provides a perspective of the target. Hence, if two representations represent the same target (e.g., the spatial relation between objects A and B) but differ in their content, i.e., how they represent the target as being ("A is in front of B" vs. "A is behind B") then we face a perspective difference. Similarly, if one person individualizes an entity as a *mouse*, another person the same entity as an *animal*, they differ in how they think of the same target object. Psycholinguists express this point by saying that the choice of label for an object puts a different perspective on that object (see Clark, 1997; Tomasello, 1999). In general, a perspective task can be characterized as a task where one becomes aware of the distinction between the target and content. We now need to show that this can cover the different cases in which all visual perspective tasks are thought to play a role.

Visual Perspective

If two people look at different scenes their visual representations are likely to differ because they see different scenes and not because they have different visual perspectives of the same scene. In contrast, if they stand side by side looking at the same scene they see the same things in the world but their visual representations still differ. Since they are looking at the same scene that difference cannot be attributed to a difference in the scenes they are looking at (the target) but only to how that single scene presents itself differently to them due to their different viewing positions. In the developmental literature children's understanding of perspective in this sense has been captured by the notion of Level 2 perspective taking (Masangkay et al., 1974; Flavell et al., 1981). At around 4 years of age children become able to understand that people who look at the same objects may see them related in different ways due to their different viewing position. The classic example is a simple drawing of a turtle positioned on a table between experimenter and child, who face each other across the table. Children before the age of 4 years understand that the turtle "stands on its feet" when its feet are pointing toward the child, and that it is "lying on its back" when the drawing has been turned by 180○. However, when asked whether the experimenter sees the turtle as standing on its feet or lying on its back they cannot give a correct answer until around 4 years of age. In contrast, much younger children have no problems with Level 1 perspective taking tasks, which test the understanding that people may see different things from different vantage points. For instance, if on a piece of paper, e.g., a car is drawn on one side and a lion on the other side, children correctly point out that the experimenter can see the car when they can see the lion.

Unfortunately, brain imaging studies do not systematically observe this distinction between Levels 1 and 2 tasks. Most of them contrast questions about what another person can see with what the participants themselves can see. Although this often only requires a Level 1 understanding, it is still likely that instruction to pay attention to what others see naturally triggers Level 2 perspective taking processes.

False Belief

The false belief test (Wimmer and Perner, 1983) has become the most popular way of assessing understanding and processing of other people's mental states both in developmental (Wellman et al., 2001) and brain imaging research (Saxe and Kanwisher, 2003). Brain imaging studies present short vignettes in which people develop a false belief (e.g., Aichhorn et al., 2009): "Julia sees the ice cream van go to the lake. She doesn't see that the van turns off to the town hall. Therefore, Julia will look for the ice cream van at the... lake/town hall?" To understand that Julia is mistaken about the location of the ice cream van one has to understand that she represents the van as being at the lake, while we know that it is at the town hall. Both, Julia and we represent the current location of the van (target) but she represents it as being at the lake while we represent it as being at the town hall. This is a difference in content hence a difference in perspective.

In contrast most imaging studies use the so-called "false photo" task[3] (originally designed for children; Zaitchik, 1990), e.g., "Julia takes a picture of the ice cream van in front of the pond. The ice cream van moves to the market place; the picture gets developed. In the picture the ice cream van is by the... pond/market place?" (Aichhorn et al., 2009). Although this task parallels in many ways the belief task—an object changes location and a representation of the object in its original location (photo/belief) persists—there are crucial differences. Unlike the belief the photo is not false and, unlike the belief, one does not have to understand the photo as giving a differing perspective on the object's location from its actual location. One just has to describe where the object is in the photo (notice: one could not ask "In Julia's belief the ice cream van is . . . ?").

[2] With IPL we denote the inferior parietal lobe consisting of the ventral region comprised by BA 40 located in the supramarginal gyrus and BA 39 located in the angular gyrus (Caspers et al., 2006, 2008).

[3] The common name for this task is an unfortunate misnomer because the photo correctly represents the object's earlier location (Perner and Leekam, 2008).

Episodic Memory

Episodic memory is defined in Tulving's tradition by Wheeler et al. (1997) as re-experiences of earlier experiences. Re-experience requires tracking of perspective. When simply experiencing an event one just takes in the event without reflecting on the fact that one has had an experience. In contrast, when re-experiencing a past event one has to understand that the experience one currently has provides but a view (perspective) of an actual past event. Without this awareness one would either mistake the re-experience for an actual experience resulting in severe delusion, or one would mistake it for an experience of an imagined, fictional event. In neither case would it count as remembering the past.

The strictest way to test for episodic memory is the remember-know judgment (Tulving, 1989). When able to retrieve a learned item or able to recognize it, participants are asked to judge whether they really remember the item, i.e., can relive their experience, or whether they just know that the item had been presented. Unlike knowing of an event the critical element of remembering an event is the double awareness of re-experiencing the event and of the fact that the event happened in one's past. In order not to mistake the re-experience as experiencing the same event again (Martin, 2001) one has to understand the ongoing re-experience as providing a perspective on something that has happened in the past.

False Signs

This task has been developed for children (Parkin, 1994) and was adopted for brain imaging by Aichhorn et al. (2009), e.g., The ice cream vendor's sign points to the lake. The ice cream van goes to the town hall without changing the sign. According to the sign post the ice cream van is at the... lake/town hall?" The false sign vignettes share with the false belief vignettes misinformation or misconception about the current state of things. In the belief vignette Julia thinks the van is at the lake, and in the false sign vignette the sign shows that the van is at the lake, when it really is at the town hall. Both vignettes differ from the "false photo" vignettes in this respect. The photo does not show where the ice cream van is, and participants are asked where in Julia's photo the van is. As pointed out earlier, this question is not possible for Julia's belief (Where in Julia's mind is the ice cream van?) and it is not possible for the false sign (Where in the sign is the ice cream van?). The two imaging studies that used false sign vignettes tested whether these vignettes activated the same brain regions as false beliefs in contrast to the "false photo" vignettes.

Commonality of Perspective Tasks

Developmental Synchrony

The four kinds of perspective tasks listed above are those for which we could find brain imaging data. All of them have been used in child appropriate versions in developmental studies. They all tend to be mastered between the age of 3–5 years (e.g., episodic remembering: Perner and Ruffman, 1995; Naito, 2003). Moreover, several studies have used the false belief task together with other perspective tasks and consistently found correlations between these tasks when controlling for differences in age and verbal intelligence (for overview see Perner and Roessler, 2012). In particular, passing the false belief task correlates with passing the level 2 visual perspective task (Hamilton et al., 2009—also in children with autism) and with passing the false sign task (Parkin, 1994; Bowler et al., 2008—also in children with autism; Sabbagh et al., 2006; Leekam et al., 2008; Iao and Leekam, 2014). Another perspective task used with children, which has not been used for brain imaging, is the appearance reality task (Flavell et al., 1983), in which children are explicitly asked what a deceptive object (a piece of sponge that looks like a rock) looks like and what it really is. Children's ability to draw this distinction also correlates with passing the false belief task (Gopnik and Astington, 1988; Taylor and Carlson, 1997; Courtin and Melot, 2005).

Cerebral Overlap: The Overarching View

Many of the developmental perspective tasks have been used in brain imaging experiments on adults. We now look for evidence whether their common development is also reflected in shared brain activity. A strict criterion for sharing brain activation would be activation overlap of all perspective tasks. This may, however, be an overly conservative criterion as Cabeza et al. (2012) argued for a similar case. Instead of looking for complete overlap they proposed the "overarching function" view that allows for subdivisions within a broader brain region. The broad region (in our case, the left IPL) has a global, overarching function (tracking perspective) and its various sub-regions mediate different aspects (false beliefs, visual perspectives, etc.) of the global function. The expected pattern of finding is that each perspective task should activate the broad region and partially overlap with activations by other perspective tasks. To check whether existing data support this view we extended an existing meta-analysis for false belief studies and visual perspective taking by Schurz et al. (2013) by also including episodic memory studies testing for remember-know judgments.

Study 1: Meta-analysis

For false belief studies and visual perspective studies we used the meta-analysis data from the work by Schurz et al. (2013) based on 25 false belief and 14 visual perspective taking (vPT)[4] studies. To this we added a meta-analysis of episodic memory (EM) studies that contrast items judged as "remembered" or "recollected" (the sense of being able to re-experience the learning phase) with items judged as just "known" or "of high confidence familiarity" (the sense of the item being old without a re-experience of learning the item). We found 16 studies that make the relevant contrasts (see details in Table S1 in supplementary material).

[4] In order to find enough studies to allow for a meta-analysis, Schurz et al. (2013) included level 1 as well as level 2 perspective tasks. Although this is conceptually less than optimal, a follow-up review by the authors showed that the main areas for vPT (e.g., the left IPL and precuneus) were equally often reported in Level 1 and in Level 2 tasks (see Table 3 on p.7 in Schurz et al., 2013). Although level 1 tasks are easy for children because they can be solved without understanding different views of the same target (simply by judging whether the object is within or outside the other person's field of vision), the same activations by level 1 and level 2 tasks in the meta-analysis suggest that level 1 tasks trigger level 2 perspective thoughts in adults. The level 1 question of what the other *sees* tends also to activate concerns about *how* the other sees the object, a level 2 concern.

All meta-analytic maps were thresholded at a voxel-wise threshold of $p < 0.005$ uncorrected and a cluster extent threshold at 10 voxels. **Figure 1** shows the activation maps for each meta-analysis. As one can see there is a potential overlap among all three kinds of tasks only on the left lateral hemisphere (2nd and 4th column) in the parietal lobe and medially (3rd column) in the posterior parts around the precuneus. **Figure 2** shows these two areas in detail. Overlap in **Figure 2** was determined by conjunction analysis between maps of significant meta-analytic activation (i.e., conjunction determined areas significantly activated in map1 AND map 2). This was done with the image calculator in SPM8 (www.fil.ion.ucl.ac.uk/spm/).

The observed pattern of overlap among activations from the three meta-analyses conforms to the view by Cabeza et al. (2012) that the IPL and possibly also parts of the anterior (y close to −60) precuneus have the overarching function of tracking perspective: All three kinds of tasks overlap in a central area but also activate individually surrounding areas. We can now use the activations shown in the meta-analyses to check whether other perspective tasks, which were tested only in a few studies, overlap with the meta-analysis. Since activations in individual studies tend to be variable we cannot expect each single study to show overlap with the central area where the three meta-analytic activations overlap. Hence our criterion for supporting evidence is that the activation of perspective tasks from individual studies must overlap with at least one of the activation areas of the meta-analysis.

As a first test case we have two studies that used false sign vignettes (Perner et al., 2006; Aichhorn et al., 2009). They looked at the regions of interest defined by the false belief vs. photo vignettes (Saxe and Kanwisher, 2003). In both studies the false sign vignettes activated the right IPL less than the false belief with no difference to the photo vignettes. In the left IPL the vignettes activated more strongly than the photo with no difference to the false belief. The same held true for the precuneus as expected under the regional specificity hypothesis that perspective tasks like the false sign task should overlap with other perspective tasks in the left IPL and precuneus.

Moreover, the left IPL was also reported in studies using conceptual perspective tasks (Goel et al., 1995; Ruby and Decety, 2003). Goel et al. (1995) asked participants to describe how, e.g., a person like Columbus from the perspective of the 15th century could infer the function of a modern artifact, e.g., hair drier. They reported activation in the left IPL and precuneus. Ruby and Decety (2003) asked medical students to respond to health-related questions either from their own perspective or from the perspective of a "lay person." Third person vs. first person activated the IPL/TPJ on the left and also on the right (to be expected since the third person perspective relied heavily on what the lay person believes about the issues). No precuneus activation was reported. So these studies confirm that the left IPL and (with less certainty) precuneus have the overarching function of tracking perspective.

In the following we test the prediction. We argue that processing identity statements requires the tracking of perspectives and thus should activate these areas in the left IPL and in precuneus whose overarching function is to track perspective.

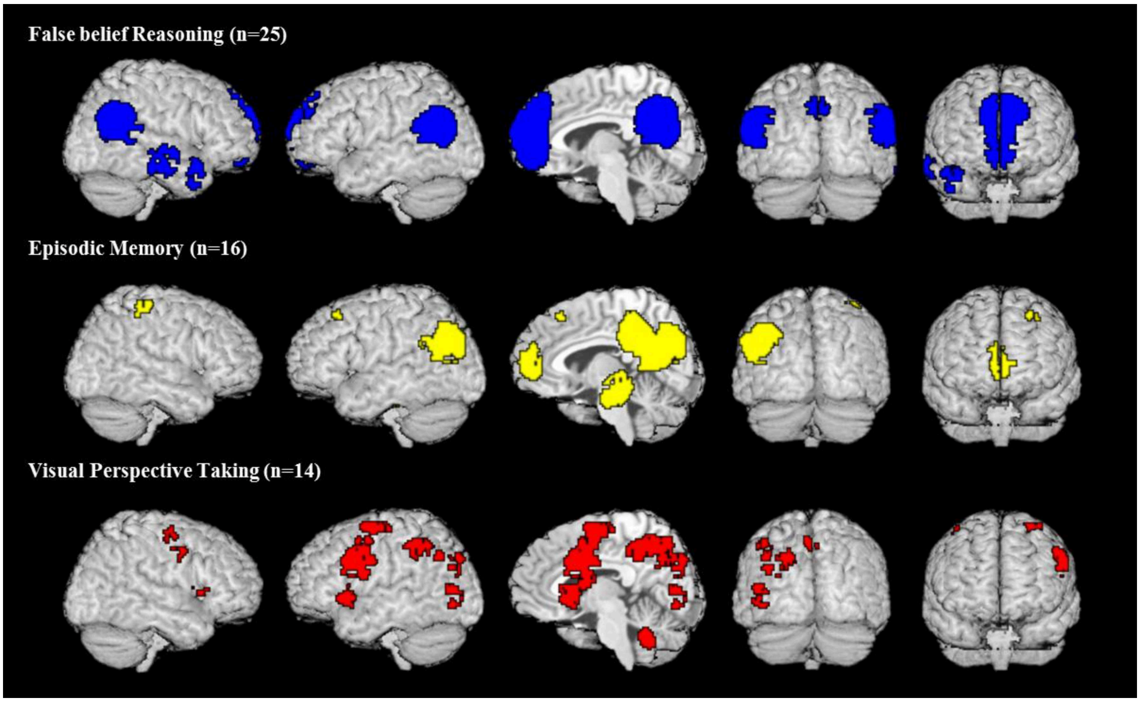

False belief Reasoning (n=25)
Episodic Memory (n=16)
Visual Perspective Taking (n=14)

FIGURE 1 | Activation maps of meta-analyses for three different domains. All maps are thresholded at voxel-wise threshold of $p < 0.005$ uncorrected and a cluster extent threshold of 10 voxels. Activations of all meta-analyses are superimposed on the Talairach template.

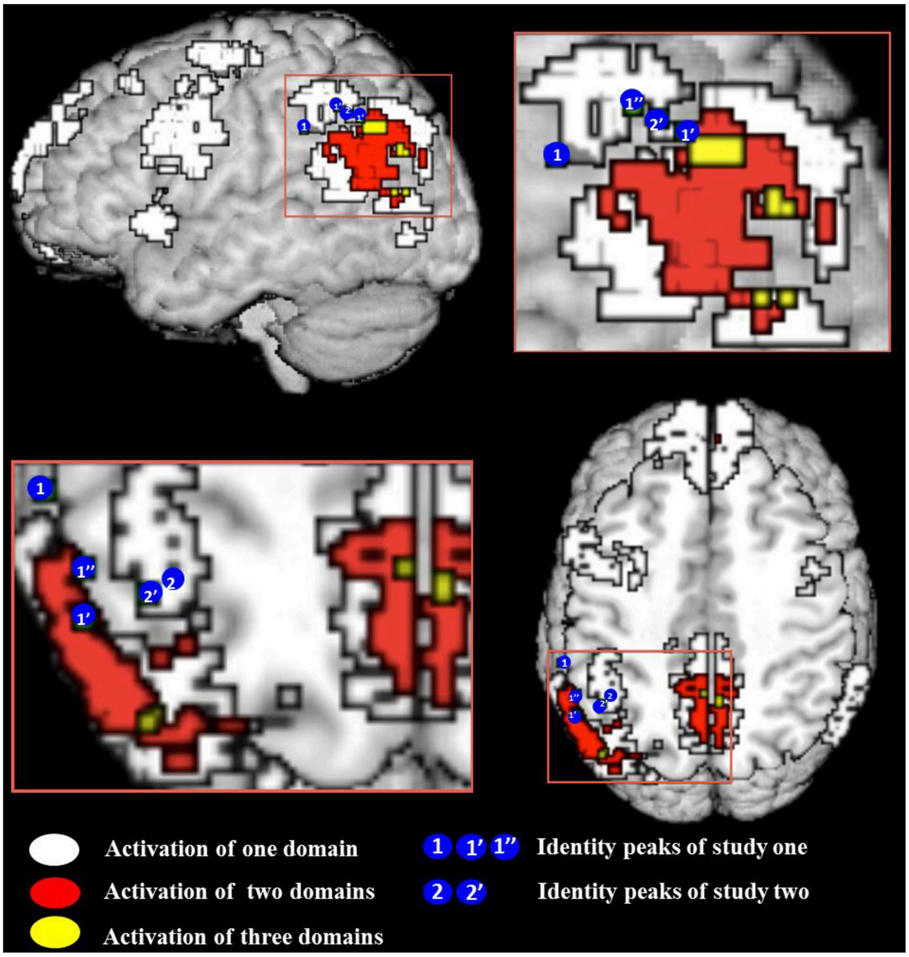

1
1'
1''
2
2'
Activation of one domain
Activation of two domains
Activation of three domains
1 1' 1'' Identity peaks of study one
2 2' Identity peaks of study two

FIGURE 2 | Conjunction map of all meta-analyses false belief (FB), episodic memory (EM), and visual perspective taking (vPT). White indicates the regions activated by one meta-analysis, red and yellow indicate the conjunction of at least two and three meta-analyses. Location of activation peaks for the identity contrast are shown as blue circles with the number of the study—see **Table 4** for peak coordinates and overlap details (areas of the blue circles do not reflect the actual size of the activation). All meta-analytic maps were thresholded at voxel-wise threshold of $p < 0.005$ uncorrected and a cluster extent threshold of 10 voxels. Activations of all meta-analyses are superimposed on the Talairach template.

Study 2: Identity 1

We want to provide a new test of the regional specificity hypothesis that the left IPL and possibly the anterior precuneus have the overarching function of tracking perspectives. For this test we try to identify an unusual candidate for a perspective task and then investigate whether it, too, activates the predicted areas. For our test we focus on identity statements, which on first blush seem to have little affinity to perspective. However, identity statements, e.g., "the driver is the tour guide" involve different labels ("driver," "tour guide") for the same individual. Psycholinguists often say that identifying an object under different labels puts a different perspective on that object (see Clark, 1997; Tomasello, 1999). Frege's (1967) and May's (2001) famous analysis of identity statements brings out the importance of perspective in the form of modes of presentation. In the identity statement "the driver is the tour guide" the expressions "the driver" and "the tour guide" refer to the same individual (person X). If the meaning of these expressions were understood only in terms of their referent (person X) then the identity statement would not be informative, for it would reduce to "person X is person X." The statement only makes sense if one is sensitive to the fact that each constituent expression provides a different mode of presentation (sense or perspective) of that particular individual to which they both refer.

Mental files (Perry, 2002; Recanati, 2012) provide a helpful alternative approach for seeing how perspective enters identity statements and why they have an affinity to understanding belief (Perner and Leahy, 2015). Use of the referential expressions "the driver" and "the tour guide" in discourse create two mental files for the same referent. They capture the two ways how one conceives of person X. The files contain the information that one has accumulated for the person under each conception. The identity statement makes clear that these are but different conceptions of a single person. One can then either keep the two files separate but link them (Perry, 2002) or merge them into a single file for person X[5]. Similarly when representing what someone mistakenly thinks, e.g., Julia in the false belief vignettes about the ice cream van, two mental files are created, a regular file registering what one knows about the van, and a vicarious file indexed to Julia. The vicarious file is linked to the regular file (Recanati, 2012) to represent sameness of referent, and on the file one registers what Julia thinks about the van. In other words, the regular file captures how oneself conceives of the van and the vicarious file how Julia conceives of it. Both, understanding identity statements and attributing false beliefs, require linked files for a single referent. This common requirement can explain why understanding identity and belief emerges at the same age (Perner et al., 2011; Perner and Leahy, 2015).

If one wants to assess brain activation due to identity statements, one has to make sure that the stimulus material induces the relevant processing. There is a danger that listeners to a statement like, "the driver is the tour guide," do not—as intended—think of two individuals, the driver and the tour guide, and then understand that there is but a single individual who is the driver and the tour guide. Instead, especially under the repetitive presentation conditions typical for fMRI, participants may gloss the sentence as "the driver is a tour guide," i.e., they only ever think of one individual as driver and then encode that he works as a tour guide. This would ruin our identity condition.

Therefore, we took care that participants naturally thought of two different individuals before they were given the critical identity information, e.g.:

> S1: "On this bus trip the tour guide talks to the passengers as much as the driver."

The listener now thinks of two people, the tour guide and the driver. Then the identity statement is given:

> S2: (+IDENT): "The tour guide is also the driver."

This informs the listener that there are not two people involved but only one person. This should—according to our Fregean analysis—make the listener aware that "tour guide" and "driver" are just two different perspectives (modes of presentation, conceptions) of that one person. A suitable control statement needs to be syntactically and in other aspects as similar as possible to our critical statement without involving an identity relation, e.g.:

> S2: (−IDENT): "The tour guide has an assistant[6]."

Unfortunately, in addition to the minor linguistic differences, there is another not so negligible difference between these two versions of sentence S2 to contend with. When two different referential expressions like "the tour guide" and "the driver" are used we naturally think (build a mental model) of two distinct people. Although natural, it is strictly speaking a rash interpretation, as the ensuing identity information makes clear. There are not two but only one person talked about. In other words, the listener has to revise her rashly formed belief of two distinct people on this bus trip to believing that there is only one person filling both positions. Quite plausibly the listener will also notice that she has been briefly misled, which amounts to attributing a false belief to herself in the immediate past. So we need to control for this in order to prevent misinterpreting activations due to the listener attributing a false belief to herself as activations caused by identity statements. In order to control for this possibility we introduced two further variations of sentence S2 one involving belief revision without any identity information:

> S2: (+REVISION): "Today, the tour guide talks more than the driver."

This would also lead to revision of the belief created by the first sentence that both people always talk the same amount. In contrast to S2 (+IDENT) it does not involve an identity statement. In order to identify activations due to this belief revision we also used a control that was syntactically similar to S2 (+REVISION) without involving a belief revision. It just adds more information:

> S2: (−REVISION): "The tour guide also earns as much as the driver."

The objective of our study is to see whether the identity contrast (+IDENT > −IDENT contrast) activates identifiable regions of the brain. The most general question (1) is whether there is any such region. More specifically (2) we expect activations in areas relevant for perspective awareness, specifically the network in the left IPL identified in **Figures 1**, **2** by meta-analyses of other perspective tasks.

However, these expectations have to be modulated by results of our belief revision control contrast (+REVISION > −REVISION), which indicates that belief revision leads to self-attribution of a false belief. In this case the identity contrast (+IDENT > −IDENT) can only be interpreted outside these regions unless the (+IDENT > +REVISION) contrast is also significant, i.e., the identity statement activates

[5] Anderson and Hastie (1974) showed in a reaction time experiment that people who have learned seemingly about two people and then learn that they are the same person keep the representation (files) for each person separate at first and later tend to merge them into a single file.

[6] Ideally the control sentence should be improved in two ways. One improvement would be to use the same names as in the identity statements: "The tour guide also has a driver," but that would clash with the first sentence. However, this difference in name is expected to be controlled for by the use of many different sentences using different names for the identity and the control. However, it leaves a systematic difference; the two names mentioned in S1 are both mentioned again in S2 in +IDENT but only one of them in −IDENT. We therefore verified whether repetition of names might activate the left IPL and precuneus in our study. Almor et al. (2007) contrasted a condition where the name introduced in the first sentence was repeated in the second sentence with a condition where a pronoun was used in the second sentence instead. This contrast did not show any activation in the IPL. There was activation in the precuneus, but in quite a different part than the activations in the present study. Another improvement would be to use "is" instead of "has," e.g., "the tour guide is a driver," but this creates the danger that participants might gloss this statement as an identity statement and annihilate any activation difference between identity and control condition.

the region in addition to any false belief attribution caused by belief revision[7].

Method

Participants

Twenty-one university students (6 males, mean age 23.95 years, *SD* = 3.96) participated in this study for course credits and small monetary reimbursement. All participants were native German speakers, had normal or corrected-to-normal vision, and had no history of neurological disorder. A written informed consent was obtained from all the participants before scanning. The ethics committee of the University of Salzburg approved the study.

Stimuli

The stimuli consisted of written German sentences (example sentences translated in English are presented in **Table 1**). During the whole experiment, 18 different scenarios were used to administer the four conditions of interest (+IDENT, −IDENT, +REVISION, –REVISION). For a particular scenario there was a standard first sentence S1. The second sentence (S2) differed for each of the four conditions. This yielded 72 different vignettes. The whole scanning session was split into three runs consisting of six trials of each condition. To avoid sequence effects vignettes derived from the same scenario were never presented near each other. Moreover, participants were instructed that all vignettes could be treated as independent and nothing had to be remembered for longer than one trial. Thirty percent of the vignettes were followed by a control question. Whether the question was about the first or the second sentence, the side of "Yes" and "No" response, and the side of the correct answer-key was randomized. Stimulus presentation, timings and response recording were controlled by Presentation software (Neurobehavioral System, Albany, CA, USA).

Procedure and Design

Participants were asked to read short vignettes. Every trial consisted of at least two sentences. At the beginning only the first sentence S1 (e.g., "On the bus trip the tour guide talks as much as the driver") was presented for 5 s. Then the second sentence S2 (e.g., "The tour guide is the driver") was added and both sentences remained for a further 6 s on the screen. In 70% of the trials of each scanning run the vignette was followed by the word "CONTINUE" (500 ms) to indicate that the trial had finished and the next one was about to start. To ensure the compliance of participants, they had to answer in the remaining trials a simple question within 6 s (e.g., "Thus a driver is on the trip: Yes?/No?) by pressing a key. Between trials a fixation cross was presented with varying duration, ranging from one to 4 s. Correct affirmative and negative answers were balanced within conditions.

The no-question trials lasted for an average of 14 s and question trials for an average of 19.5 s. Before the start of each trial there was an inter-stimulus interval of 1–4 s. The sequence of the trial and the inter-stimulus interval was optimized using Russ Poldrack's script (we optimized a fixed time span for four conditions of interest and one rest condition; http://sourceforge.net/projects/fmri-toolbox/files/optimize_design/1.1/).

fMRI Data Acquisition

Functional and structural imaging was acquired with a Siemens 3 Tesla Tim-Trio Scanner, located at Christian-Doppler-Clinic, Salzburg. Functional images sensitive to the BOLD contrast were obtained with a T2*-weighted gradient echo-planar imaging (EPI) sequence using a 32 channel head coil. Per subject, three sessions, and a total of 239 EPI images including 6 dummy scans at the beginning of the functional images were scanned to allow transient signals to diminish (TR = 2000 ms; TE = 30 ms; matrix size = 96 × 96; voxel size = 2.187 × 2.187 × 3.58 mm^3; slice thickness = 3.0 mm; slice gap 0.6 mm; FOV = 210 mm; flip angle = 70°). Thirty-six axial slices were acquired in descending order parallel to the bicommissural (co-planar with AC–PC) line along the z-axis. In addition to functional scanning, sagittally oriented high-resolution structural scan was acquired (T1-weighted MP-RAGE sequence; TR = 6.73 ms; TE = 3.14 ms; voxel size 0.797 × 0.797 × 1.2 mm^3; slice-thickness = 1.2 mm; matrix 256 × 256; FOV = 204 mm; 170 slices per volume; flip angle = 8°).

fMRI Data Processing

Preprocessing and statistical data analysis was performed by Statistical Parametric Mapping (SPM8, http://www.fil.ion.ucl.ac.uk/spm), implemented in MATLAB 7.3 [R2006b] (Matworks, Sherborn, MA) runtime environment. Images were slice-time and motion corrected by standard SPM8 algorithms. Functional images were registered to the SPM8 EPI template. The structural scan was co-registered onto the mean functional images of each session and segmented. Segmentation parameters were used for normalization of structural and functional images to MNI space (Montreal Neurological Institute, McGill, Montreal, Canada) template. The normalized images were resampled to isotropic 3 × 3 × 3 mm voxels and smoothed with an 8 mm full width at half maximum (FWHM) Gaussian kernel.

The preprocessed data were analyzed using a general linear model (GLM) approach. The functional data were high-pass filtered in order to remove frequencies below 1/128 Hz to reduce low frequency drift. The serial correlation was taken into account using the autocorrelation AR (1) model, as implemented in SPM8. On individual level contrast the four conditions relative to fixation baseline were modeled. The condition sentence (S2) was modeled as an event of interest for all four conditions separately. The context sentence (S1) and the verification questions were modeled as regressors of no interest. Additionally, realignment parameters and session mean were included as covariates. The

[7] As one of our reviewers rightly pointed out the contradiction in the control task (+REVISION) is a direct incompatibility between S1 and S2, while the contradiction in the identity task only occurs due to natural pragmatic assumptions about S1 of there being two separate individuals. On this basis one would expect stronger activations for belief revision in the control than in the identity task. This safeguards against false positives, i.e., that we would not detect the effects of belief revision in the control task when it is present in the identity task.

TABLE 1 | Example sentences of Study 2 (translated from German; see Table S2. in supplementary material for more original examples in German).

Conditions	Context sentence (S1) 5 s	Condition sentence (S2) 6 s	Verification sentence 6 s
Identity (+IDENT)	On this bus trip the tour guide talks to the passengers as much as the driver[a]	The tour guide is also the driver	Thus, a tour guide is on the bus. <yes>
No Identity (−IDENT)		The tour guide has an assistant	Thus, the assistant always comes along. <yes>
Belief Revision (+REVISION)		Today, the tour guide talks more than the driver	Thus, today one of them does more of the talking. <yes>
No Belief Revision (−REVISION)		The tour guide also earns as much as the driver	Thus, both earn different amounts of money. <no>

[a] *The same context sentence was used for all conditions.*

TABLE 2 | Behavioral results of Study 2: mean accuracy in percent hit rate (SD).

	Conditions			
	+IDENT	−IDENT	+REVISION	−REVISION
Hit-Rate (%)*SD*	91.3(14.3)	87.2(8.5)	90.3(12.4)	94.3(9.1)

first level contrast images of each subject were used for the second level (random effects) analysis, that allows for the generalization to the population. The statistical comparisons were inspected at a voxelwise threshold of $p < 0.001$ together with a cluster extent threshold of $p < 0.05$, corrected for family-wise error (FWE).

Results

Behavioral Results

The overall accuracy was around 90% (see **Table 2**), indicating that the participants were attentive and understood the task. We computed a One-Way repeated measure ANOVA using participants' hit-rates. There was no statistically significant difference in accuracy across the four conditions [$F_{(3,\ 60)} = 1.488$, $p = 0.22$, $\eta^2 = 0.069$]. This implies that the difficulty level was similar across all conditions.

We will not report reaction times (RT) for the sake of brevity. This is because RTs were collected on the Yes/No responses to the questions presented within the response window of 6 s, they do not reflect the actual time taken to comprehend the vignettes but rather the time taken to read the question and respond "yes" or "no" to the visual cue.

Neuro-imaging Results

We report all regions for identity and belief revision contrasts at FWE cluster level corrected $p < 0.05$ in **Table 3**.

Of main interest was the identity contrast comparing identity with its control condition (+IDENT > −IDENT). Only one parietal activation in the left inferior parietal lobe (left IPL) with its main peak and one of the sub-peaks in the supramarginal gyrus (SMG) and another sub-peak in angular gyrus (AG) was FWE cluster level corrected significant at $p < 0.05$. Comparison in the opposite direction (−IDENT > +IDENT) did not reveal any significant cluster.

TABLE 3 | Supra-threshold whole brain activation of identity and belief revision in Study 2.

Region	H	*k*	Max *Z*	MNI coordinates		
				x	*y*	*z*
IDENTITY: +IDENT > −IDENT						
Supramarginal Gyrus (PF L)	L	90	4.46	−60	−34	37
Angular Gyrus (PFm L)	*L*	–	*3.84*	*−54*	*−52*	*43*
Supramarginal Gyrus (PF L)	*L*	–	*3.55*	*−54*	*−43*	*46*
BELIEF REVISION: +REVISION > −REVISION						
Angular/Lateral occipital cortex, superior division (Pga L)	L	136	4.53	−51	−64	34
Angular Gyrus (PFm L)	*L*	–	*4.26*	*−42*	*−58*	*31*
Middle Frontal Gyrus (BA6 L)	L	95	4.43	−39	11	52
Middle Frontal Gyrus (BA44 L)	*L*	–	*3.65*	*−39*	*17*	*40*
Middle Frontal Gyrus (BA44 L)	*L*	–	*4.03*	*−42*	*20*	*49*

Significant cluster are reported at p < 0.05 FWE cluster level corrected.
Regions are reported from posterior to anterior. Regions, Anatomical labeling corresponding to the cluster peak and sub-peak (according to Harvard-Oxford cortical and subcortical structural atlases). Regions in brackets, Anatomical labeling corresponding to the cluster peak and to sub-peaks are also reported according to Jülich histological cyto-and myelo-architectonic atlas by Eickhoff et al. (2005, 2006, 2007); H, Hemisphere of peak; k, cluster extent in voxel; Max Z, Maximum Z-value; sub-peaks of the regions with cluster level below p < 0.05 FWE corrected are reported in italics.

The belief revision contrast (+REVISION > −REVISION) activated two clusters FWE corrected at $p < 0.05$; one in the left IPL (angular gyrus) and the other in the left middle frontal gyrus. The inverse contrast (−REVISION > +REVISION) did not show any significant activation. For each relevant contrast, overlap with meta-analytic activations was tested in the following way: Based on the peak-voxel coordinate activated in this contrast, we checked for each meta-analysis map if significant activation was found here. The left angular gyrus cluster peak and sub-peaks of the belief revision contrast were also significantly activated in our false belief and in our episodic memory meta-analysis. This overlap suggests that by becoming aware of having to revise one's belief one attributes a false belief to oneself[8].

[8] This is a novel finding with interesting implications. Attribution of false beliefs to oneself could be a reason why invalid cue trials on the Posner task activate the belief attribution region in the TPJ (Mitchell, 2008).

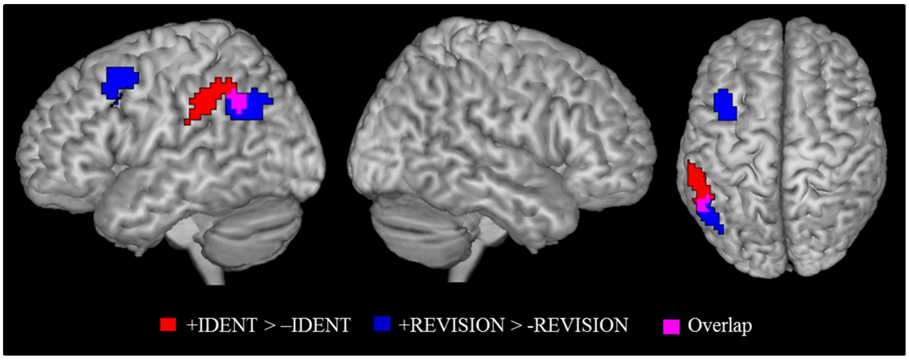


FIGURE 3 | Identity contrast (red); belief revision contrast (blue), and overlap between the two contrasts (magenta). Activation cluster are superimposed on an MNI template. All contrasts were shown at $p < 0.05$ FWE cluster level corrected threshold.

Identity and Belief Revision

As argued earlier in the explanation of our experimental design, we needed to check if brain activity for identity statements can also be found for other statements that cause belief revision. This is necessary in order to not misinterpret activations as caused by identity statements when in fact they may be due to the listener attributing a false belief to herself. **Figure 3** shows the activation patterns for the identity contrast (+IDENT > −IDENT) and for belief revision contrast (+REVISION > −REVISION). Overlap was determined by inclusively masking the belief revision contrast with the identity contrast (at the default threshold of $p < 0.001$). We found overlap in the left angular gyrus (−54, −52, 43) $k = 15$ and in the right lateral occipital cortex (48, −64, 40) $k = 5$. Given this overlap, we cannot rule out that our identity statements were activating these left IPL areas because they caused a belief revision. Therefore, to detect areas activated by the identity contrast independently of belief revision, we removed (exclusively masked) all regions activated by belief revision ($p < 0.001$ uncorrected) from the identity contrast. The identity contrast outside the belief revision mask stayed significant in the left IPL, ($k = 74$) at FWE cluster level corrected $p < 0.05$ with the cluster peak (−60, −34, 37) and two sub-peaks in the left supramarginal gyrus (−54, −43, 46; −54, −49, 43).

This result confirms our expectation based on developmental data that identity statements activate the left left IPL, as the region is sensitive to perspective differences. To answer our more specific question, whether identity statements activate a more specific "perspective region" in the left IPL, we need to define a region of interest. Here we adopt the *overarching view* of Cabeza et al. (2012) that allows for subdivisions within a broad brain region. The broad region (in our case, the left IPL) has a global function (representing perspective differences) and its various sub-regions mediate different aspects (false beliefs, visual perspectives, etc.) of the global function. The expected pattern of finding is that each perspective task should activate the broad region and partially overlap with activations by the other tasks. For this purpose we used the results from our meta-analysis. We checked for each peak voxel if a meta-analysis showed significant activation at the given coordinate. Results of this examination are given in **Table 4**. **Figure 2** show the overlay of identity contrast peaks with the activations shown in the meta-analyses.

TABLE 4 | Overlap (+) of identity activations in Study 2 and 3 with false belief, episodic memory, and visual perspective taking.

	Peak label	Overlap with					
		MNI coordinates			FB	EM	vPT
STUDY 2							
Cluster peak: SMG	1	−60	−34	37	−	−	−
Sub-peak: AG	1′	−54	−52	43	+	−	−
Sub-peak: SMG	1″	−54	−43	46	−	−	−
STUDY 3							
Cluster peak: SMG	2	−39	−46	43	−	−	+
Sub-peak: SMG	2′	−42	−49	46	−	−	+

Peak label. corresponds to the labeling in ***Figure 2****. MNI peak coordinates of Study 2 and Study 3 were converted into Talairach space to have the same stereotactic space as the meta-analysis. FB, False belief reasoning; EM, Episodic memory; vPT, Visual perspective taking.*

We were unable to directly compare results because our imaging studies and the meta-analyses were analyzed in different coordinate systems. All meta-analyses had to be performed in Talairach space, as the default coordinate system of Effect-Size Signed Differential Mapping (ES-SDM) software, version 2.31 for meta-analysis (Radua et al., 2010, 2012); http://www.sdmproject.com), while our data were normalized in MNI space. We thus converted our left SMG cluster peak and sub-peaks into Talairach space (see **Table 4**). We constructed a 3 mm in diameter sphere—which corresponds to the voxel-size of our images – around those peaks using the WFU PickAtlas (http://fmri.wfubmc.edu/software/PickAtlas). One of the sub-peak spheres in the angular gyrus that overlapped with belief revision also overlapped significantly with false belief meta-analysis areas [a coordinate-wise search for foci that were significantly activated in both analyses, performed in MRIcron (http://www.mccauslandcenter.sc.edu/mricro/mricron/)]. This confirms our prediction that the processing of identity statements might have led participants to correct their rashly formed belief about the "tour guide" and the "driver" as being two distinct people to believing that there is only one person filling both positions.

Discussion

Our initially formulated expectations for the identity contrast received a fairly clear answer. (1) We were able to identify at least one region that is significantly (FWE-corrected) activated by the identity contrast. (2) This identity cluster lies in the left IPL as predicted in the hypothesis; tasks that require awareness of perspective will activate this region. (3) Although the main peak and one of the sub-peaks of the identity cluster were in the left supramarginal gyrus another sub-peak was in the angular gyrus that overlapped with false belief activation of the meta-analysis.

This pattern of results fits the overarching view (Cabeza et al., 2012) that the left IPL has the overarching function of registering (actual or potential) perspective differences. Different tasks modulate this function, showing activation in different parts of the IPL but such that they partially overlap, as the meta-analysis of perspective tasks (false belief, visual perspective taking, and episodic memory) show. Our results extend this picture to identity tasks.

Overlap of the identity contrast in our study happened to occur in the meta-analytic areas for false belief activation. One problem of interpretation occurred because our identity task involved belief revision. Belief revision, as we were able to show, also activates in the meta-analytic false belief area, suggesting that belief revision, at least when one is aware of it, amounts to attributing a past false belief to oneself. This raises the possibility that the overlap between the identity contrast and false belief may be due to the belief attribution caused by the belief revision inherent in our identity condition. Therefore, it would be reassuring if overlap with perspective tasks can be found without the involvement of belief revision in identity tasks. This was investigated in the next study.

Study 3: Identity 2

The objective of this experiment is to check whether the central results of Study 2 can be replicated by avoiding the confounding of identity statements with belief revision. The confound resulted from our decision to prevent participants glossing a simple identity statement like, "the mayor is the lawyer," as an attributive statement, "the mayor is a lawyer." While the former mentions two people (the mayor and the lawyer) and then says something about their identity, the latter only mentions one person (the mayor) and then informs about that person's profession. To avoid such a gloss we used a context sentence to establish the mayor and the lawyer as two different individuals in participants' minds. With the identity statement participants then learned that mayor and lawyer are the same person. This led inevitably to a belief revision.

For this current experiment we decided to run the risk of participants glossing some of the identity statements as attributive assertions. If this results in similar activations as in Study 2 (especially of the left IPL) we can conclude that these activations are not due to belief revision. Trying to minimize the risk of an attributive gloss, each statement used a common description (the lawyer) as its first referential term and as the second term a proper name (Mr Müller). Although one can easily gloss "Mr Müller is the lawyer" as "Mr Müller is a lawyer)," it is harder to do so with "The lawyer is Mr Müller[9]."

Method

Participants

Seventeen (5 males; mean age 24.6 years, $SD = 4.9$ years) right-handed university students participated in this study for course credits and small monetary reimbursement. All participants were native German speakers, had normal or corrected-to-normal vision, and had no history of neurological disorder. A written informed consent was obtained from all the participants before scanning. The ethical committee of the University of Salzburg approved the study.

Design and Stimuli

The study had five conditions (see **Table 5**) consisting of written German sentences. Three context conditions were introduced

[9] Although not impossible; one could gloss it as "The lawyer is called Mr Müller." Also, the use of a proper name in the identity condition raises the potential danger that the proper name is responsible for the left IPL and precuneus activation and not the identity statement itself. Fortunately, existing data from clinical and imaging studies speak against this possibility (for review see Semenza, 2011). Processing of proper names compared to common names was linked to activation in bilateral temporal poles, and, somewhat less consistently, to anterior parts of the superior temporal sulcus, ventral mPFC, and the anterior cingulate. In contrast, the left IPL and precuneus were not associated with processing of proper names.

TABLE 5 | Example sentences of Study 3 (translated from German; see Table S3 in supplementary material for more original examples in German).

Conditions	Context sentence (S1) 4.5 s	Condition sentence (S2) 3 s	Comprehension questions? 5.5 s
Identity-with-context (IDENTc)	The doctor saves the lawyer after the accident	The lawyer is Mr. Moser	Who is Mr. Moser?[a]
Predication-with-context (PREDc)		The lawyer is young	Who is young?
Context only (C)		–	Who saved the lawyer?
Identity only (IDENTo)	–	The neurologist is Dr. Phillips	Who is the neurologist?
Baseline (BL)		The chair is old-fashioned	What is old-fashioned?

[a] *The comprehension question in conditions with context sentence varied accordingly (see design and stimuli section of Study 3 for details).*

with a context sentence mentioning two people, e.g., a doctor and a lawyer. In the *identity-with-context* (IDENTc) condition an identity statement followed which expressed that one of these people (lawyer) was identical to, e.g., Mr. Müller. In the *predication-with-context* (PREDc) condition the second sentence predicated some attribute of, e.g., the lawyer. In the *context-only* (C) condition this second sentence was omitted. This condition served as a parameter of no interest for comparing IDENTc with PREDc. Two additional conditions served to replicate a finding of a pilot study using simple identity statements without any background context (*identity only*, IDENTo). The pilot activation was difficult to interpret, as the design didn't have any explicit low-level baseline. We therefore included, a low-level baseline condition (BL) with simple sentences (e.g., the glasses are old-fashioned).

Twenty-seven different sentences were used per condition, resulting in a total of 135 trials in the experiment. All sentences of IDENTc and PREDc conditions were formed by linking a referential noun phrase, e.g., "The lawyer" by the particle "is" with either a proper name to form an identity statement or with an adjective to form predicative sentences. The noun phrases were counterbalanced for the two conditions.

We controlled for sentence length in all conditions. The mean number of letters in the context sentences (S1) varied between conditions from 40.7 (±6.0) in IDENTc to 40.5 (±6.0) in PREDc to 41.2 (±7.0) in C, and the average letter count in the identity sentences (S2) varied from 22.19 (±2.6) in IDENTo to 23.5 (±2.9) in IDENTc. There was no significant difference across conditions for context or for identity sentences (all p's $\geq$ 0.35).

The presentation times for sentences S1 and S2 are shown in **Table 5**. On 30% of trials a comprehension question was asked. In the context conditions this question could be about any of the three names mentioned (for example see **Table 5**: "Who saved the lawyer?" or "Who did the doctor save?" or "Who is Mr. Moser?"). This variation was to ensure that participants had to integrate sentences S1 and S2 in a single model. In the conditions without context the question only varied between the two names that referred to the same individual (e.g., "Who is Mr. Moser?" or "Who is the neurologist?"). The total time provided was 5500 ms: the question was presented for 3000 ms, followed by 1000 ms of black screen, and finally the answer option for 250 ms (e.g., <the lawyer> <the doctor>). Correct and incorrect options to the question were balanced across conditions to avoid confounds of any strategies to answer the questions and habitual finger use. Stimulus presentation, timings and response recording were controlled by the Presentation Software (Neurobehavioral System, Albany, CA, USA).

Functional neuroimaging was divided into three sessions. Each session comprised 45 trials, 9 pre-condition trials and 14 comprehension questions. The order of the presentation of sessions was counterbalanced across participants. A single trial without question lasted for 11 s in the conditions with context, 6.5 s in the identity only and in the baseline condition, and 8 s in the context only trials. Each single session lasted for 10.35 min, and the whole functional scanning of the experiment took 31.07 min.

Procedure

The participants were given a training session before the start of the scanning. They were specifically instructed to read and understand the sentences carefully, and that they would sometimes be asked to answer a question to verify their attention and comprehension of the vignettes. Behavioral responses were collected using an MRI-compatible response box.

fMRI Data Acquisition

Functional and structural imaging was acquired with a Siemens 3 Tesla Tim-Trio Scanner, located at the Christian-Doppler-Clinic, Salzburg. Functional images sensitive to the BOLD contrast were obtained with a T2*-weighted gradient EPI sequence using a 32 channel head coil. Per subject, three sessions, a total of 260 EPI images including 6 dummy scans at the beginning of the functional images were scanned to allow transient signals to diminish ($TR = 2250$ ms; $TE = 30$ ms; matrix size $= 64 \times 64$; voxel size $= 3.0 \times 3.0 \times 3.0\,\text{mm}^3$; slice thickness $= 3.0$ mm; slice gap 0.3 mm; $FOV = 192$ mm; flip angle = 70°). Thirty-six axial slices were acquired in descending order parallel to the bicommissural (co-planar with AC–PC) line along the z-axis. In addition for each subject sagittally oriented high-resolution structural scan was acquired (T1-weighted MP-RAGE sequence; $TR = 2300$ ms; $TE = 2.91$ ms; voxel size $1.0 \times 1.0 \times 1.0\,\text{mm}^3$; slice-thickness $= 1.00$ mm; matrix 256×256; $FOV = 256$ mm; 192 slices per volume; flip angle $= 9°$).

fMRI Data Processing

Preprocessing and statistical data analysis was performed using Statistical Parametric Mapping (SPM8, http://www.fil.ion.ucl.ac.uk/spm), implemented in MATLAB 7.6.0.324 [R2008a] (Matworks, Sherborn, MA) runtime environment. Images were slice-time and motion corrected by standard SPM8 algorithms. Functional images were registered to the SPM8 EPI template. The structural scan was co-registered onto the mean functional images of each session and segmented. The structural and functional images were normalized to MNI (Montreal Neurological Institute, McGill, Montreal, Canada) template. The normalized images were resampled to isotropic $3 \times 3 \times 3$ mm voxels and smoothed with an 8 mm full width at half maximum (FWHM) Gaussian kernel.

The preprocessed data were analyzed using a GLM approach. Per subject, and session, IDENTc, PREDc, IDENTo, and BL condition sentence (S2) was modeled as a separate regressor of interest with the duration of 3 s and convolved with the hemodynamic response function. The S1 of conditions with context (IDENTc and PREDc) and C were modeled with the duration of 4.5 s as a single regressor of no interest. We also modeled the comprehension question with the duration of 5.5 s as a separate regressor of no interest. Additionally, realignment parameters and session means were included in the design matrix as covariate. The low frequency noise was removed by high-pass filter with a cut-off of 128 s, and serial correlation was taken into account using an autocorrelation AR (1) model, as implemented in SPM8. At the individual level of contrasts the four conditions were modeled separately relative to an implicit baseline.

TABLE 6 | Behavioral results of Study 3: mean accuracy in percent hit rate (SD).

	Conditions				
	IDENTc	**PREDc**	**IDENTo**	**BL**	**C**
Hit-Rate (%)SD	98.4(4.6)	97.4(5.9)	97.6(5.5)	100(0)	94.1(5.7)

Data at the second level were subject to a random effects analysis to allow for population inference. We computed paired t-tests between contrasts of interest. Whole brain results are reported at a voxel-wise threshold of $p < 0.001$ together with a FWE cluster level corrected threshold of $p < 0.05$.

Results and Discussion

Behavioral Results

Overall accuracy was very high 97.51% (see **Table 6**), with an overall miss rate of 5.06%. The high accuracy was a good indicator that participants were attentive and understood the task. Given that accuracy was at ceiling in this study, it was unnecessary to carry out statistical tests here.

We do not report reaction time (RT), since the RTs depended on the time spent to answer the comprehension question they do not reliably reflect the actual time taken to comprehend the vignettes.

Neuro-imaging Results

The main contrast of interest is the one between *identity-with-context* and *predication-with-context* (IDENTc > PREDc). The whole brain analysis for this contrast showed two significant FWE-corrected clusters (see **Table 7**). One cluster lies in the precuneus on the left side, the other, in the left supramarginal gyrus as predicted.

The inverse contrast (PREDc > IDENTc) showed activations in quite distant parts of the brain (see **Table 7** and **Figure 4**). Two large FWE corrected clusters were located in the left and right temporal pole area associated with social scripts and social concepts (Zahn et al., 2007; Ross and Olson, 2010) and prevalent in theory of mind studies (Schurz et al., 2014). This is plausibly due to the fact that predicative information about a person (the lawyer is young) stimulates social thoughts more strongly than a statement that this person is identical to someone (Mr. Moser) about whom one has no information.

The identity statement without context compared to the baseline condition (IDENTo > BL) showed significant activation of the left supplementary motor area (SMA), left precentral gyrus, left lateral occipital cortex, bilateral cerebellum, left inferior frontal gyrus (IFG) and right superior parietal lobe activation at FWE cluster level corrected at $p < 0.05$[10].

TABLE 7 | Supra-threshold whole brain activation of identity vs. predication in context conditions of Study 3.

Region	**H**	***k***	**Max *Z***	**MNI coordinates**		
				x	***y***	***z***
IDENTITY: IDENTc > PREDc						
Precuneus Cortex (7M L)	L	88	4.07	−12	−67	28
Lateral Occipital Cortex, superior division (7P L)	*L*	–	*3.61*	*−15*	*−76*	*46*
Precuneus Cortex	*L*	–	*3.58*	*−18*	*−70*	*22*
Supramarginal Gyrus (hIP1)	L	67	4.09	−39	−46	43
Supramarginal Gyrus (hIP1)	*L*	–	*4.06*	*−42*	*−49*	*46*
INVERSE IDENTITY: PREDc > IDENTc						
Temporal Pole (No label)	R	146	4.90	48	8	−26
Superior Temporal Gyrus (No label)	*R*	–	*4.24*	*57*	*−10*	*−8*
Superior Temporal Gyrus (No label)	*R*	–	*4.58*	*54*	*2*	*−17*
Temporal Pole (No label)	L	191	5.69	−51	11	−20
Superior Temporal Gyrus (No label)	*L*	–	*4.12*	*−51*	*−4*	*−17*
Temporal Pole (No label)	*L*	–	*5.19*	*−45*	*14*	*−26*

Significant cluster are reported at p < 0.05 FWE cluster level corrected.
Regions are reported from posterior to anterior. Regions, Anatomical labeling corresponding to the cluster peak and sub-peak (according to Harvard-Oxford cortical and subcortical structural atlases). Regions in brackets, Anatomical labeling corresponding to the cluster peak and sub-peak are also reported according to Jülich histological cyto-and myelo-architectonic atlas by Eickhoff et al. (2005, 2006, 2007); H, Hemisphere of peak; k, cluster extent in voxel; Max Z, Maximum Z-value; sub-peaks of the regions with cluster level below p < 0.05 FWE corrected are reported in italics.

Relation to Study 2 and Meta-analysis

The predicted activation by the identity contrast (IDENTc > PREDc) in Study 3 was in close vicinity to the activation observed in the left IPL for the identity contrast (+IDENT > − IDENT) in Study 2. After masking the belief revision clusters the average Euclidian distance between the sub-peaks of Study 2 (−54, −43, 46, and −54, −49, 43) and the cluster peak and sub-peak (−39, −46, 43, and −42, −49, 46) of Study 3 was 14.16 mm. In order to assess the support for the claim that all perspective tasks activate the overarching region in the left IPL we tested

[10] This finding poses two questions for us. The first one is problematic for our account: Why does this contrast not activate in the left IPL? We can offer the following two post-hoc explanations. Intuitively, in the no-context condition (see **Table 5**) "The neurologist is Dr. Phillips," can plausibly be glossed as "The neurologist is called Dr. Phillips," hence no identity is expressed, and consequently no left IPL activation. This gloss is intuitively less likely when a context is provided: S1 "The doctor saves the lawyer after the accident," followed by S2 "The lawyer is Mr. Moser," is less likely to be glossed in a similar way as indicated by the fact that a glossed version of sentence 2 "The lawyer is called Mr. Moser," would provide an unexpected and less informative content than the un-glossed original version. Our second explanation pertains to the fact that the comprehension questions in the context conditions varied. They could be, e.g., "Who saved the lawyer?," "Who did the doctor save?," or "Who is Mr. Moser?" which can only be answered if sentences S1 and S2 have been integrated within a model. In contrast, sentences S2 without context, e.g., "The neurologist is Dr. Phillip," the questions were always about the person mentioned in S2, "Who is the neurologist?" or "Who is Dr. Phillip?" This question could be answered on the basis of the sentence's surface form without interpreting it within a mental model, i.e., without thinking of different individuals and identity—hence no left IPL activation. The second question raised by this finding is: Why does this contrast activate five areas which are not activated by the sentences in the context conditions? This is an interesting question but not directly problematic for our account. One feature that distinguishes IDENTo > BL from IDENTc > PREDc contrast is that the former contrast is confounded with a contrast of person vs. no person, which could account for at least some of these activations.

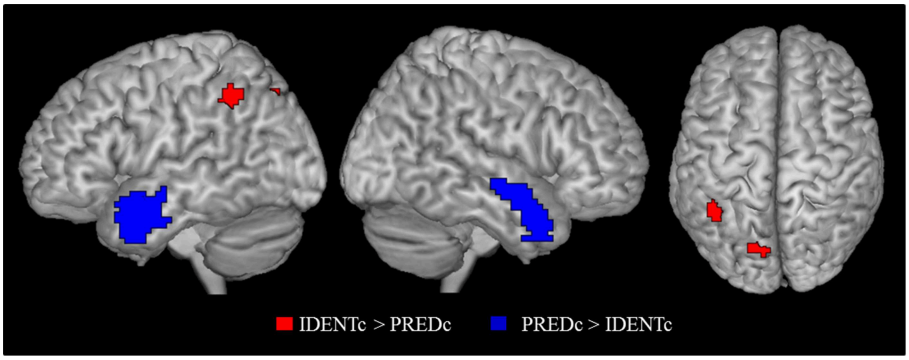


FIGURE 4 | Contrast of identity-with-context > predication-with-context (red) and the inverse contrast (blue). Activation cluster are superimposed on an MNI template. All contrasts were shown at $p < 0.05$ FWE cluster level corrected threshold.

for overlap of the identity contrast (IDENTc > PREDc) with areas shown in the meta-analyses. Using the same method as in Study 2 we converted the left SMG cluster peak and sub-peaks of Study 3 (see **Table 4**, **Figure 2** for overlap details) into Talairach space, and constructed a 3 mm in diameter sphere around it, which overlapped with the regions of the visual perspective taking meta-analysis and bordered on the areas activated by false belief vignettes and episodic memory.

The fact that the activations found in the two identity studies did not directly overlap is mitigated by the strong connectivity between the subareas of the IPL in which the activations occurred. The SMG cluster peak (−60, −34, 37) and the sub-peaks (−54, −52, 43, and −54, −43, 46) of the identity contrast in Study 2 (see **Table 3**) fall into the cytoarchitectonic area of the left PF and PFm region (Caspers et al., 2006, Jülich Histological Atlas). The SMG cluster peak (−39, −46, 43) and sub-peak (−42, −49, 46) of Study 3 (see **Table 7**) are located in the left intraparietal sulcus (subregion: hIP1; Choi et al., 2006; Jülich Histological Atlas). According to Caspers et al. (2011) structural connectivity fingerprints show a strong connection between PF, PFm, and hIP1 region. The strong connectivity among the different areas activated by our studies supports the conclusion that different activation points reflect activity of an overarching functionally related network.

General Discussion

Main Achievements

Our studies produced two main achievements. (a) We were able to establish that the ability to track perspective, which marks an important advance in child development around 4 years of age, manifests itself in a common brain activity. Based on existing data we hypothesized that such commonality might be reflected in mutual activations of a particular brain region. The results show that, indeed, all different kinds of perspective tasks that, to our knowledge, have been used in brain imaging activate the left IPL and precuneus; although the evidence for the latter remains less solid.

(b) Our second achievement was to turn the "overarching view" of a region's broader function, which Cabeza et al. used to summarize existing results, into a predictive instrument. We proceeded in the following way. In a meta-analysis we established that activations of three kinds of perspective tasks show triple overlap in the left IPL and precuneus. This result establishes that the left IPL and precuneus, qualify as areas with the overarching function of tracking perspective. To test the general validity that these regions are responsible for tracking perspective we looked for further perspective tasks. We found several single studies, too few for a meta-analysis. We then needed to check whether the reported activations overlap with the meta-analytic areas, ideally within the area of triple overlap. However, results from single studies do not show the stability of meta-analyses and total overlap with all three tasks from the meta-analysis would be unreasonably conservative. So we settled for the following criterion: The results satisfy the expectations from the overarching view if the activations are found in the target areas (the left IPL and precuneus) and overlap with at least one of the meta-analytic activations in those areas.

With this procedure we were able to show that existing data conform to the hypothesis that the left IPL and precuneus qualify as areas with the overarching function of tracking perspective. We then used the same technique for prediction of identity statements, which qualify as perspective tasks on the basis of a technical account, activate within the overarching regions of the left IPL and precuneus. This prediction was confirmed and with it the hypothesis that these areas track perspective.

The concept of an overarching function helps with the problem of low power of individual studies. For instance, the lack of overlap of activations in our two identity studies can be explained by two factors accounted for in the overarching view. Due to their low power, activations happen to be detected at different points within the overarching region. Another reason for the discrepancy is that the belief revision induced in our first study drew the center of activation more toward the region where false beliefs are processed than in the second study where no belief revision occurred.

Relation to Competing Theories

The main competitor for our claim that the left IPL has the overarching function of tracking perspective is the BUA (bottom-up attention) model for the ventral part of the parietal cortex (VPC = IPL) put forward by Cabeza et al. (2012). As an extension of Corbetta and Shulman (2002) dual attention model BUA sees the VPC (IPL) bilaterally responsible for detecting salient and behaviorally relevant stimuli in the environment, especially when they were previously unattended (exogenous, or stimulus-driven attention). Cabeza et al. (2012) extended this model from attention capture by environmental stimuli to capture by internal (memory-based) information. Three interesting aspects arise about the relationship between BUA and perspective tracking: similarities, reducibility, and differences.

Similarities

Perspective tasks can be seen as a special kind of internal attention capture. In our thinking and conversations we usually stick to a single perspective because mixing different perspectives is a source of confusion[11] . Therefore, (external or internal reasoning) cues that indicate the need for a change in perspective are exogenous stimuli, and should activate the IPL according to BUA. Attention capture by cues for potential perspective differences is, however, special as it does not require reorienting attention to information about a new topic but reorienting to a new way of informing about (view, mode of presentation, perspective of) the same topic. On these grounds we may consider two possible views of how activation of the left IPL by perspective tasks relates to BUA.

Differences

Perspective tracking differs strikingly from BUA in terms of lateralization. Perspective tracking evidently has regional specificity only for the left IPL, while BUA is claimed to operate bilaterally. Cabeza et al. (2012) noticed a prevalence of the left IPL (VPC) activation reports for some tasks in their review and give two possible reasons for it. Left activation reports prevail when predominantly verbal stimulus material is used. However, this explanation does not quite fit the finding that false belief vignettes, which are purely verbal, activate bilaterally (Schurz et al., 2014) while visual perspective tasks, which use a much stronger visual presentation mode, activate exclusively on the left side (Schurz et al., 2013).

Cabeza et al. also suggested that authors often focus on one hemisphere for historical reasons linked to work on patients with lesions, e.g., neglect being observed with right hemisphere parietal lesions. This explanation does not apply to the evidence from perspective tasks we have reviewed, which stems exclusively from fMRI studies without any historical bias. Although few studies test for hemispheric asymmetry the sheer number of studies that report activation only in left and not in the right IPL is remarkable. Of the 14 visual perspective tasks included in the meta-analysis by Schurz et al. (2013) all of them reported activity in left, only Wraga et al. (2010) found bilateral IPL activation. Similarly in our meta-analysis of 16 remember-know studies all of them report left and only Eldridge et al. (2000) reported bilateral IPL activation. This is clear evidence of stronger activation in the left, as only two out of thirty studies (combined vPT + EM) showed bilateral activation and no other study showed activation in the right hemisphere (binomial test $z = -4.56, p < 10^{-6}$).

Moreover, the two false sign studies (Perner et al., 2006; Aichhorn et al., 2009) only showed effects in the left IPL, and our two studies with identity statements also showed significant reliable activation in the left IPL[12]. The noticeable exception to this left asymmetry are false belief vignettes, which activate the TPJ (including the IPL) on the right as much as on the left (Schurz et al., 2014; see our **Figure 1**). One reason for this may be that the false belief task engages theory of mind, which activates areas in temporal lobe immediately adjacent and overlapping with the left and right IPL. In contrast, the other perspective tasks show no activations in adjacent areas, only in rather distant areas. All of them tend to activate the precuneus in an overlapping fashion (see **Figure 2**). Episodic remembering activates bilateral para-hippocampal gyrus areas [e.g., Daselaar et al., 2006; our episodic memory meta-analysis (see **Figure 1**)], whereas visual perspective tasks activate, the precuneus, left IPL, precentral, and middle frontal region.

Reducing Perspective Tracking to BUA

As outlined above perspective tasks can be seen as a special case of exogenous attention capture, because endogenous thinking usually maintains to the same perspective. One obvious exception to this occurs when perspective itself becomes the topic of thinking. For instance, in visual perspective tasks the instructions are to judge how another viewer sees the display. So taking the other person's perspective is endogenous to the set task and should, according to BUA, activate dorsal parts of the parietal cortex and not the IPL. Another problem case for BUA is a fact persistently ignored in the discussion of why theory of mind tasks activate the TPJ (or IPL) as a consequence of attention reorienting in false belief tasks (Decety and Lamm, 2007; Corbetta et al., 2008; Mitchell, 2008; Cabeza et al., 2012). It is never made clear why the act of reorienting plausibly required in the false belief vignettes (shifting attention from where an object actually is to where an agent mistakenly thinks it is) is not also required in the photo control vignettes (shifting from where the object actually is to where it is in a photo), a contrast introduced by Saxe and Kanwisher (2003) and since used in many studies with exceedingly strong meta-analytic effects (Schurz et al., 2014).

These two problem cases for BUA can be explained by perspective tracking. Visual perspective tasks require perspective tracking hence activate the left IPL. False belief tasks do so too and reliably activate the left IPL, while the photo control tasks do not. A photo taken of the ice cream van in an earlier location does not give a different perspective on where the van is now (unlike a false belief or a flipped direction sign which does give a different view of where the van is now). In sum, although perspective tracking shows a close affinity to bottom-up attention processes it is unlikely that the activation in the left IPL perspective tasks can be completely explained by BUA.

[11] A good example are conceptual pacts (Brennan and Clark, 1996) which help ensure that a particular object of conversation is referred to under the same label, since a change of label also entails a change of perspective (Clark, 1997).

[12] However right IPL activation for the identity only condition suggests that the lateralization is one of degree and not one without any involvement of the right hemisphere.

Reducing Left Lateralized BUA to Perspective Tracking

A different view on perspective tracking and BUA is to claim that only perspective tracking is the overarching function of (at least) the left IPL. To defend this view one would need to show that the evidence recited by Cabeza et al. (2012) in favor of BUA can also be used as evidence for perspective tracking, i.e., that all the tasks that activate the left IPL can be argued to be perspective tasks. Up to now we have considered only tasks that had been independently claimed to be perspective tasks in the developmental literature. Hence, whether a task should or should not activate the left IPL was a predictive enterprise from an existing classification. To retrospectively decide whether a task, which activates the IPL, is a perspective task or not is a much more unconstrained enterprise. We will therefore restrain our analysis to some exemplary illustrations taken from the categories discussed by Cabeza et al.

Number Processing

Equations can be viewed as identity statements (numerical facts: 4+5 is identical to 9) or computational procedures (if you have 4 and add 5 you get 9). So retrieval of numerical facts should activate the left IPL since an identity is likely involved which induces perspective tracking. And, indeed, the IPL is being activated (Dehaene et al., 2003). In contrast, calculation of the result should not activate the IPL or, at least, less so. This also turns out to be the case (Grabner et al., 2009). So, some findings in this area clearly relate to perspective tracking.

Episodic Retrieval

In contrast to three contenders discussed by Cabeza et al. BUA can explain a characteristic U-function of recognition certainty. The IPL activation is stronger for items judged “definitely old” or “definitely new” than for uncertain answers (data only for the left IPL; Yonelinas et al., 2005; Daselaar et al., 2006). This activation pattern can also result from perspective tracking. Correct recognition can come about for two reasons at least (Jacoby, 1991). One can make a conscious judgment of whether the presented test item has been on the learning list. In some of these cases one may use an episodic approach (Tulving, 1989) and try to re-experience ones’ earlier experience of having seen this item during learning. Re-experience requires awareness that one’s re-experience of seeing the item is a representation, which gives a perspective, of the past event (Perner et al., 2007). Plausibly if this approach gives a clear answer it will provide high confidence that the item has or has not been experienced. Since awareness of perspective is involved, the confident judgments will activate the left IPL. In other cases no clear judgment may be possible but one can still rely on a feeling of familiarity. Depending on the strength of this feeling one will respond with “old” or “new,” but the subjective confidence will be low. Familiarity judgments do not need awareness of perspective; hence the resultant low confidence answers will not be associated with activations of the left IPL.

Conclusion

Tracking and monitoring perspectives is a skill whose acquisition has important consequences on children’s reasoning and social competence around the age of 4 years. In a meta-analysis of brain imaging in adults we were able to show that this important developmental factor is also reflected in a common cerebral resource: the left IPL and precuneus track perspective. In two empirical studies we were able to extend this finding and confirm that these brain regions are reliably involved in other and novel kinds of perspective tasks, e.g., processing identity statements.

Author Contributions

AA was responsible for Study 3, contributions to the meta-analyses of episodic remembering, and the coordination of all contributions and writing of the manuscript. BW conducted Study 2 in partial fulfillment of his master’s degree at the Department of Psychology, University of Salzburg. RW performed the analysis of episodic memory studies. MS provided the general meta-analytic expertise and MA the technical support for collecting and analysing the fMRI data of both studies. JP provided the theoretical framework.

Acknowledgments

The project received financial support from the Austrian Science Fund (FWF) Project I93-G15: “Metacognition of Perspective Differences” as part of the ESF EUROCORES CNCC (Consciousness in a Natural and Cultural Context) initiative for BW to conduct the fMRI study, and from the Austrian Science Fund funded Doctoral College (DK W 1233-G17) “Imaging the Mind” for AA.

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fnhum.2015.00360

References

Aichhorn, M., Perner, J., Weiss, B., Kronbichler, M., Staffen, W., and Ladurner, G. (2009). Temporo-parietal junction activity in theory-of-mind tasks: falseness, beliefs, or attention. *J. Cogn. Neurosci.* 21, 1179–1192. doi: 10.1162/jocn.2009.21082

Almor, A., Smith, D. V., Bonilha, L., Fridriksson, J., and Rorden, C. (2007). What is in a name? Spatial brain circuit are used to track discourse reference. *Neuroreport* 18, 1215–1219. doi: 10.1097/WNR.0b013e32810f2e11

Anderson, J., and Hastie, R. (1974). Individual and reference in memory: proper names and definite descriptions. *Cogn. Psychol.* 6, 495–514. doi: 10.1016/0010-0285(74)90023-1

Bowler, D. M., Gaigg, S. B., and Gardiner, J. M. (2008). Effects of related and unrelated context on recall and recognition by adults with high-functioning autism spectrum disorder. *Neuropsychologia* 46, 993–999. doi: 10.1016/j.neuropsychologia.2007.12.004

Brennan, S. E., and Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *J. Exp. Psychol. Learn.*

Mem. Cogn. 22, 1482–1493. doi: 10.1037/0278-7393.22.6.1482

Cabeza, R., Ciaramelli, E., and Moscovitch, M. (2012). Cognitive contributions of the ventral parietal cortex: an integrative theoretical account. *Trends. Cogn. Sci.* 16, 338–352. doi: 10.1016/j.tics.2012.04.008

Caspers, S., Eickhoff, S. B., Geyer, S., Scheperjans, F., Mohlberg, H., Zilles, K., et al. (2008). The human inferior parietal lobule in stereotaxic space. *Brain. Struct. Funct.* 212, 481–495. doi: 10.1007/s00429-008-0195-z

Caspers, S., Eickhoff, S. B., Rick, T., von Kapri, A., Kuhlen, T., Huang, R., et al. (2011). Probabilistic fibre tract analysis of cytoarchitectonically defined human inferior parietal lobule areas reveals similarities to macaques. *Neuroimage* 58, 362–380. doi: 10.1016/j.neuroimage.2011.06.027

Caspers, S., Geyer, S., Schleicher, A., Mohlberg, H., Amunts, K., and Zilles, K. (2006). The human inferior parietal cortex: cytoarchitectonic parcellation and interindividual variability. *Neuroimage* 33, 430–448. doi: 10.1016/j.neuroimage.2006.06.054

Choi, H.-J., Zilles, K., Mohlberg, H., Schleicher, A., Fink, G. R., Armstrong, E., et al. (2006). Cytoarchitectonic identification and probabilistic mapping of two distinct areas within the anterior ventral bank of the human intraparietal sulcus. *J. Comp. Neurol.* 495, 53–69. doi: 10.1002/cne.20849

Clark, E. V. (1997). Conceptual perspective and lexical choice in acquisition. *Cognition* 64, 1–37. doi: 10.1016/S0010-0277(97)00010-3

Corbetta, M., Patel, G., and Shulman, G. L. (2008). The reorienting system of the human brain: from environment to theory of mind. *Neuron* 58, 306–324. doi: 10.1016/j.neuron.2008.04.017

Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215. doi: 10.1038/nrn755

Courtin, C., and Melot, A. M. (2005). Metacognitive development of deaf children: lessons from the appearance-reality and false belief tasks. *Dev. Sci.* 8, 16–25. doi: 10.1111/j.1467-7687.2005.00389.x

Daselaar, S. M., Fleck, M. S., Dobbins, I. G., Madden, D. J., and Cabeza, R. (2006). Effects of healthy aging on hippocampal and rhinal memory functions: an event-related fMRI study. *Cereb. Cortex* 16, 1771–1782. doi: 10.1093/cercor/bhj112

Decety, J., and Lamm, C. (2007). The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. *Neuroscientist* 13, 580–593. doi: 10.1177/1073858407304654

Dehaene, S., Piazza, M., Pinel, P., and Cohen, L. (2003). Three parietal circuits for number processing. *Cogn. Neuropsychol.* 20, 487–506. doi: 10.1080/02643290244000239

Eickhoff, S. B., Heim, S., Zilles, K., and Amunts, K. (2006). Testing anatomically specified hypotheses in functional imaging using cytoarchitectonic maps. *Neuroimage* 32, 570–582. doi: 10.1016/j.neuroimage.2006.04.204

Eickhoff, S. B., Paus, T., Caspers, S., Grosbras, M. H., Evans, A. C., Zilles, K., et al. (2007). Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. *Neuroimage* 36, 511–521. doi: 10.1016/j.neuroimage.2007.03.060

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., et al. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 25, 1325–1335. doi: 10.1016/j.neuroimage.2004.12.034

Eldridge, L. L., Knowlton, B. J., Furmanski, C. S., Bookheimer, S. Y., and Engel, S. A. (2000). Remembering episodes: a selective role for the hippocampus during retrieval. *Nat. Neurosci.* 3, 1149–1152. doi: 10.1038/80671

Flavell, J. H., Everett, B. A., Croft, K., and Flavell, E. R. (1981). Young chindren's knowledge about visual perception: further evidence for level 1-level 2 distinction. *Dev. Psychol.* 17, 99–103. doi: 10.1037/0012-1649.17.1.99

Flavell, J. H., Flavell, E. R., and Green, F. (1983). Development of the appearance-reality distinction. *Cogn. Psychol.* 15, 95–120. doi: 10.1016/0010-0285(83)90005-1

Frege, G. (1967). "Begriffsschrift, a formula language, modeled upon that of arithmetic for pure thought (1879)," in *From Frege to Gödel: Mathematical Logic*, ed J. Van Heijenoort (Cambridge: Harvard University Press), 1–82.

Goel, V., Grafman, J., Sadato, N., and Hallett, M. (1995). Modeling other minds. *Neuroreport* 6, 1741–1746. doi: 10.1097/00001756-199509000-00009

Gopnik, A., and Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child. Dev.* 59, 26–37. doi: 10.2307/1130386

Grabner, R. H., Ischebeck, A., Reishofer, G., Koschutnig, K., Delazer, M., Ebner, F., et al. (2009). Fact learning in complex arithmetic and figural-spatial tasks: the role of the angular gyrus and its relation to mathematical competence. *Hum. Brain. Mapp.* 30, 2936–2952. doi: 10.1002/hbm.20720

Hamilton, A. F. D. C., Brindley, R., and Frith, U. (2009). Visual perspective taking impairment in children with autistic spectrum disorder. *Cognition* 113, 37–44. doi: 10.1016/j.cognition.2009.07.007

Iao, L. S., and Leekam, S. R. (2014). Nonspecificity and theory of mind: new evidence from a nonverbal false-sign task and children with autism spectrum disorders. *J. Exp. Child. Psychol.* 122, 1–20. doi: 10.1016/j.jecp.2013.11.017

Jacoby, L. L. (1991). A process dissociation framework: separating automatic from intentional uses of memory. *J. Mem. Lang.* 30, 513–541. doi: 10.1016/0749-596X(91)90025-F

Kanwisher, N. (2010). Functional specificity in the human brain: a window into the functional architecture of the mind. *Proc. Natl. Acad. Sci. U.S.A.* 107, 11163–11170. doi: 10.1073/pnas.1005062107

Leekam, S., Perner, J., Healey, L., and Sewell, C. (2008). False signs and the non-specificity of theory of mind: evidence that preschoolers have general difficulties in understanding representations. *Br. J. Dev. Psychol.* 26, 485–497. doi: 10.1348/026151007X260154

Martin, M. G. F. (2001). "Out of the past: episodic recall as retained acquaintance," in *Time and Memory: Issues in Philosophy and Psychology*, eds C. Hoerl and T. McCormack (Oxford: Clarendon Press), 257–284.

Masangkay, Z. S., McCluskey, K. A., McLntyre, C. W., Sims-Knight, J., Vaughn, B. E., et al. (1974). The early development of inferences about the visual percepts of others. *Child. Dev.* 45, 357–366. doi: 10.2307/1127956

May, R. (2001). "Frege on identity statements," in *Semantic Interfaces: Reference, Anaphora, and Aspect*, eds C. Cecchetto, G. Chierchia, and M. T. Guasti (Stanford, CA: CSLI Publications), 1–62.

Mitchell, J. P. (2008). Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cereb. Cortex* 18, 262–271. doi: 10.1093/cercor/bhm051

Naito, M. (2003). The relationship between theory of mind and episodic memory: evidence for the development of autonoetic consciousness. *J. Exp. Child. Psychol.* 85, 312–336. doi: 10.1016/S0022-0965(03)00075-4

Parkin, L. J. (1994). *Children's Understanding of Misrepresentation*. Unpublished D.Phil thesis, University of Sussex.

Perner, J., Aichhorn, M., Kronbichler, M., Staffen, W., and Ladurner, G. (2006). Thinking of mental and other representations: the roles of left and right temporo-parietal junction. *Soc. Neurosci.* 1, 245–258. doi: 10.1080/17470910600989896

Perner, J., Brandl, J. L., and Garnham, A. (2003). What is a perspective problem? Developmetal issues in belief ascription and dual identity. *Facta. Philos.* 5, 355–378.

Perner, J., Kloo, D., and Stöttinger, E. (2007). Introspection and remembering. *Synthese* 159, 53–270. doi: 10.1007/s11229-007-9207-4

Perner, J., and Leahy, B. (2015). Mental files in development: dual naming, false belief, identity and intensionality. *Rev. Philos. Psychol.* 1–18. doi: 10.1007/s13164-015-0235-6

Perner, J., and Leekam, S. (2008). The curious incident of the photo that was accused of being false: issues of domain specificity in development, autism, and brain imaging. *Q. J. Exp. Psychol.* 61, 76–89. doi: 10.1080/17470210701508756

Perner, J., Mauer, M. C., and Hildenbrand, M. (2011). Identity: key to children's understanding of belief. *Science* 333, 474–477. doi: 10.1126/science.1201216

Perner, J., and Roessler, J. (2012). From infants' to children's appreciation of belief. Trends *Cogn. Sci.* 16, 519–525. doi: 10.1016/j.tics.2012.08.004

Perner, J., and Ruffman, T. (1995). Episodic memory and autonoetic consciousness: developmental evidence and a theory of childhood amnesia. *J. Exp. Child. Psychol.* 59, 516–548. doi: 10.1006/jecp.1995.1024

Perry, J. (2002). *Identity, Personal Identity, and the Self*. Indiana: Hackett Publishing Company, Inc.

Radua, J., MataixCols, D., Phillips, M. L., El-Hage, W., Kronhaus, D. M., Cardoner, N., et al. (2012). A new meta-analytic method for neuroimaging studies that combines reported peak coordinates and statistical parametric maps. *Eur. Psychiatry* 27, 605–611. doi: 10.1016/j.eurpsy.2011.04.001

Radua, J., van den Heuvel, O. A., Surguladze, S., and Mataix-Cols, D. (2010). Meta-analytical comparison of voxel-based morphometry studies in obsessive-compulsive disorder vs. other anxiety disorders. *Arch. Gen. Psychiatry* 67, 701–711. doi: 10.1001/archgenpsychiatry.2010.70

Recanati, F. (2012). *Mental Files*. Oxford, UK: Oxford University Press.

Ross, L. A., and Olson, I. R. (2010). Social cognition and the anterior temporal lobes. *Neuroimage* 49, 3452–3462. doi: 10.1016/j.neuroimage.2009.11.012

Ruby, P., and Decety, J. (2003). What you believe versus what you think they believe: a neuroimaging study of conceptual perspective-taking. *Eur. J. Neurosci.* 17, 2475–2480. doi: 10.1046/j.1460-9568.2003.02673.x

Sabbagh, M. A., Moses, L. J., and Shiverick, S. (2006). Executive functioning and preschoolers' understanding of false beliefs, false photographs, and false signs. *Child. Dev.* 77, 1034–1049. doi: 10.1111/j.1467-8624.2006.00917.x

Saxe, R., and Kanwisher, N. (2003). People thinking about thinking people the role of the temporo-parietal junction in "theory of mind." *Neuroimage* 19, 1835–1842. doi: 10.1016/S1053-8119(03)00230-1

Schurz, M., Aichhorn, M., Martin, A., and Perner, J. (2013). Common brain areas engaged in false belief reasoning and visual perspective taking: a meta-analysis of functional brain imaging studies. *Front. Hum. Neurosci.* 7:712. doi: 10.3389/fnhum.2013.00712

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., and Perner, J. (2014). Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neurosci. Biobehav. Rev.* 42, 9–34. doi: 10.1016/j.neubiorev.2014.01.009

Semenza, C. (2011). Naming with proper names: the left temporal pole theory. *Behav. Neurol.* 24, 277–284. doi: 10.1155/2011/650103

Taylor, M., and Carlson, S. M. (1997). The relation between individual differences in fantasy and theory of mind. *Child. Dev.* 68, 436–455. doi: 10.2307/1131670

Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press, 161–200.

Tulving, E. (1989). Memory: performance, knowledge, and experience. *Eur. J. Cogn. Psychol.* 1, 3–26. doi: 10.1080/09541448908403069

Wellman, H. M., Cross, D., and Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child. Dev.* 72, 655–684. doi: 10.1111/1467-8624.00304

Wheeler, M. A., Stuss, D. T., and Tulving, E. (1997). Toward a theory of episodic memory: the frontal lobes and autonoetic consciousness. *Psychol. Bull.* 121, 331–354. doi: 10.1037/0033-2909.121.3.331

Wimmer, H., and Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13, 103–128. doi: 10.1016/0010-0277(83)90004-5

Wraga, M., Flynn, C. M., Boyle, H. K., and Evans, G. C. (2010). Effects of a body-oriented response measure on the neural substrate of imagined perspective rotations. *J. Cogn. Neurosci.* 22, 1782–1793. doi: 10.1162/jocn.2009.21319

Yonelinas, A. P., Otten, L. J., Shaw, K. N., and Rugg, M. D. (2005). Separating the brain regions involved in recollection and familiarity in recognition memory. *J. Neurosci.* 25, 3002–3008. doi: 10.1523/JNEUROSCI.5295-04.2005

Zahn, R., Moll, J., Krueger, F., Huey, E. D., Garrido, G., and Grafman, J. (2007). Social concepts are represented in the superior anterior temporal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 104, 6430–6435. doi: 10.1073/pnas.0607061104

Zaitchik, D. (1990). When representation conflict with reality: the preschooler's problem with false belief and "false" photograph. *Cognition* 35, 41–68. doi: 10.1016/0010-0277(90)90036-J

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.

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 21 January 2015
doi: 10.3389/fnhum.2014.01042

Meta-analysis: how does posterior parietal cortex contribute to reasoning?

Carter Wendelken *

Helen Wills Neuroscience Institute, University of California, Berkeley, CA, USA

Edited by:
Vinod Goel, York University, Canada

Reviewed by:
Marian Berryhill, University of Nevada, Reno, USA
Fabio Richlan, University of Salzburg, Austria
Barbara J. Knowlton, University of California, Los Angeles, USA

****Correspondence:***
Carter Wendelken, Helen Wills Neuroscience Institute, University of California, Barker Hall 134A, Berkeley, CA 94720, USA
e-mail: cwendelken@berkeley.edu

Reasoning depends on the contribution of posterior parietal cortex (PPC). But PPC is involved in many basic operations—including spatial attention, mathematical cognition, working memory, long-term memory, and language—and the nature of its contribution to reasoning is unclear. Psychological theories of the processes underlying reasoning make divergent claims about the neural systems that are likely to be involved, and better understanding the specific contribution of PPC can help to inform these theories. We set out to address several competing hypotheses, concerning the role of PPC in reasoning: (1) reasoning involves application of formal logic and is dependent on language, with PPC activation for reasoning mainly reflective of linguistic processing; (2) reasoning involves probabilistic computation and is thus dependent on numerical processing mechanisms in PPC; and (3) reasoning is built upon the representation and processing of spatial relations, and PPC activation associated with reasoning reflects spatial processing. We conducted two separate meta-analyses. First, we pooled data from our own studies of reasoning in adults, and examined activation in PPC regions of interest (ROI). Second, we conducted an automated meta-analysis using Neurosynth, in which we examined overlap between activation maps associated with reasoning and maps associated with other key functions of PPC. In both analyses, we observed reasoning-related activation concentrated in the left Inferior Parietal Lobe (IPL). Reasoning maps demonstrated the greatest overlap with mathematical cognition. Maintenance, visuospatial, and phonological processing also demonstrated some overlap with reasoning, but a large portion of the reasoning map did not overlap with the map for any other function. This evidence suggests that the PPC's contribution to reasoning may be most closely related to its role in mathematical cognition, but that a core component of this contribution may be specific to reasoning.

Keywords: deductive reasoning, posterior parietal cortex, IPL, SPL, numerical cognition, spatial cognition, meta-analysis

INTRODUCTION

Reasoning, the capacity to reach novel conclusions on the basis of existing premises, is among the most complex of cognitive operations. It necessarily depends on multiple underlying capacities, but the extent of this reliance on specific mechanisms is a subject of considerable debate. One possibility is that reasoning, generally or in some cases, utilizes syntactic representations of premises and application of formal logical rules (Rips, 1994; Braine and O'Brien, 1998). If this is the case, then the representations afforded by language are likely to be central to reasoning (Kertesz and McCabe, 1975; Carruthers, 2002). Another possibility is that reasoning proceeds via the use of quasi-perceptual mental models, in which case the high-level spatial and perceptual representations upon which the models are built would be critical for reasoning (Johnson-Laird, 1983, 2001). Recent work has emphasized the role of probabilistic mechanisms, in contrast to deterministic logical rule-following, in much of human reasoning (Oaksford and Chater, 2009). To the extent that reasoning proceeds via estimation and probabilistic computation, mechanisms for number processing should be critical. Of course, multiple mechanisms are possible (see e.g., Goel et al., 2000), so these theories are not mutually exclusive.

Reasoning often depends on attention to relational structure, so the mechanisms that support basic relational processing are also likely to be key. Relational representations might depend upon semantic understanding of relational terms, in which case mechanisms of semantic processing can be expected to come into play during reasoning. Alternatively, relational representations may be built upon the representation of space and spatial relationships, in which case the mechanisms of visuospatial processing may be more central to reasoning. In addition, working memory, long-term memory, and attention are all basic cognitive mechanisms that are likely to contribute to reasoning.

Many investigations of reasoning, including our own, have highlighted the role of rostrolateral prefrontal cortex (RLPFC; Christoff et al., 2001; Bunge et al., 2005; Wendelken and Bunge, 2010; Wendelken et al., 2012). In particular, these studies have shown that RLPFC contributes to second-order relational reasoning, which involves the joint consideration or integration of multiple relations and is thought to be a core component of the reasoning capacity (Gentner and Holyoak, 1997; Halford et al., 1998; Penn et al., 2008; Chuderski, 2014). However, posterior parietal cortex (PPC) is also consistently engaged during reasoning tasks (Crone et al., 2009; Eslinger et al., 2009; Watson and Chatterjee, 2012; Wendelken et al., 2012). Like RLPFC, PPC is sensitive to the need to integrate relations, but PPC is also sensitive to the number of relations considered (Crone et al., 2009) and the specificity of those relations (Wendelken and Bunge, 2010). Furthermore, there is mounting evidence from lesion studies pointing toward a critical role for PPC in reasoning. One study of left-hemisphere stroke patients revealed that performance on a matrix reasoning task was affected by damage to the inferior parietal lobe (IPL; Baldo et al., 2010). In another recent investigation, involving patients with damage to RLPFC or parietal cortex, only patients with parietal damage were significantly impaired on a transitive inference task (Waechter et al., 2013).

That PPC makes an important contribution to reasoning is apparent; but PPC is involved in numerous cognitive functions besides reasoning. To understand PPC's contribution to reasoning, it is critical to understand how it relates to the other functions of PPC. We summarize primary functions attributed to PPC briefly here. For more extensive review of parietal function, see Grefkes and Fink (2005), Nickel and Seitz (2005), Seghier (2013), and Humphreys and Lambon Ralph (2014).

A key function of PPC is the implementation of visuospatial attention (Mesulam, 1981; Hopfinger et al., 2001; Wager et al., 2004), and of spatial processing more generally (Marshall and Fink, 2001; Husain and Nachev, 2007; Sack, 2009; Amorapanth et al., 2010). The intraparietal sulcus (IPS), which separates the inferior and superior parietal lobes, has been shown to contribute to the maintenance of spatial location information (Todd and Marois, 2004; Xu and Chun, 2006; Ackerman and Courtney, 2012). IPL, by contrast, has been implicated as a locus of spatial relational processing (Ackerman and Courtney, 2012).

PPC has also been linked to various language processes (Binder et al., 2009; Wu et al., 2012). For example, posterior IPL, angular gyrus, particularly on the left side, has been implicated as a key locus for semantic processing (Binder et al., 2009; Seghier, 2013). Moreover, just as IPS has been implicated as the locus of visuospatial maintenance, more anterior and ventral parts of IPL have been implicated in maintenance of verbal information (Paulesu et al., 1993; Awh et al., 1996; Becker et al., 1999).

In addition to its apparent role in the maintenance of both spatial and verbal information, PPC, and in particular SPL, has also been implicated in manipulation of the contents of working memory (Marshuetz et al., 2000; Wager and Smith, 2003; Wendelken et al., 2008). Moreover, PPC contributes not only to various aspects of working memory, but also to episodic memory (for review, see Berryhill and Olson, 2012). In episodic memory, parietal activation is most commonly associated with the endorsement of stimuli as having been previously encountered (Wagner et al., 2005; Nelson et al., 2013), though associations with memory encoding (e.g., Uncapher and Wagner, 2009) and memory confidence (e.g., Johnson et al., 2013) have also been noted.

Finally, though this list is by no means exhaustive, PPC is a primary contributor to mathematical cognition (Dehaene et al., 2003; Rosenberg-Lee et al., 2011). Some aspects of mathematical cognition may be linked to verbal and spatial representations within PPC (Dehaene et al., 1999). But evidence suggests that a core numerical system, localized to IPS, may be independent of these (Dehaene et al., 2003; Cohen Kadosh et al., 2005; Nieder et al., 2006).

Whether these various functions of parietal cortex on the one hand rely on shared circuitry and similar operations, or on the other hand represent separable circuits and distinct functionality, is a subject of much debate. A number of studies have sought to parcellate PPC into distinct subdivisions with differing functional roles (e.g., Nelson et al., 2010, 2013; Mars et al., 2011), while others have sought to explain apparently diverse functions in terms of a core mechanism (e.g., Bueti and Walsh, 2009; Cabeza et al., 2012).

It is possible that PPC supports reasoning through one dominant mechanism, be it numerical processing, relational representation, language, attention, working memory, or some other function; but it is also possible that different subdivisions of PPC support reasoning in different ways (see e.g., Goel, 2007; Prado et al., 2011). Regardless, understanding the way or ways in which PPC supports reasoning is critical for understanding not only the neural implementation of reasoning, but also for understanding the extent to which reasoning depends on different cognitive mechanisms.

Here, we re-examined previously collected data to better characterize the contribution of PPC to reasoning. We pursued two broad approaches. First, we examined parietal data from our own fMRI studies of deductive reasoning, all of which included a contrast between second-order and first-order relational reasoning conditions, to determine which parietal subdivisions are most selectively engaged by the higher-order reasoning condition. Second, we expanded our investigation to a much broader collection of studies to find characteristic activation patterns across PPC for reasoning as well as for a number of other parietal functions. We compared the spatial overlap of activation patterns associated with reasoning and with other cognitive functions, to determine whether or not parietal engagement during reasoning could be best understood in relation to its involvement in these other functions of parietal cortex.

METHODS

All of our analyses, described below, were focused on activation patterns within PPC. Our specific parietal regions of interest (ROIs) were based on the parietal subdivisions defined in Mars et al. (2011) on the basis of tractography (**Figure 1**). The set of ROIs included, on each side of the brain, five subdivisions of the IPL, arrayed from anterior to posterior, and five subdivisions of the SPL, similarly arrayed from anterior to posterior;

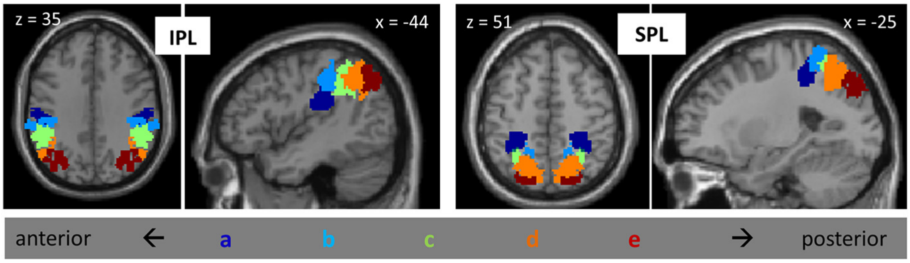


FIGURE 1 | Posterior parietal ROIs from Mars et al., 2011, including five subdivisions of IPL and five subdivisions of SPL, on the left and on the right. Subdivisions are labeled "a" through "e", from anterior to posterior.

thus, there were a total of 20 parietal ROIs. For convenience, we label these regions IPLa—IPLe and SPLa—SPLe, with "a" referring to the most anterior subdivisions and "e" referring to the most posterior subdivisions. IPLa, with a center of gravity at (±49, −25, 30), is located ventral to the other IPL regions, in the parietal opercular region (Caspers et al., 2006). IPLb, with center of gravity at (±53, −32, 44), corresponds to anterior supramarginal gyrus, while IPLc, with a center of gravity at (±50, −44, 43) corresponds to posterior supramarginal gyrus. IPLd, with a center of gravity at (46, −55, 45), is located in the anterior part of the angular gyrus, and IPLe, with a center of gravity at (37, −67, 39), comprises posterior angular gyrus and the most anterior parts of the lateral occipital complex. All of these IPL regions, with the exception of IPLa, are bordered by the IPS. The anterior-most SPL region (SPLa), with a center of gravity at (30, −41, 53), was located on the anterior medial bank of the IPS. SPLb, with center of gravity at (12, −50, 63), was adjacent and medial to SPLa. SPLc, with center of gravity at (28, −55, 55), comprised the middle-to-posterior medial bank of the IPS. SPLd, with center of gravity at (19, −63, 53), was medial and posterior to SPLc. Finally, SPLe, with a center of gravity at (21, −78, 43), included the most posterior part of the medial bank of the IPS.

We first examined data from four different studies of relational reasoning that we have previously conducted in young adults (Bunge et al., 2009; Crone et al., 2009; Wendelken and Bunge, 2010; Wendelken et al., 2012). These deductive reasoning tasks included matrix reasoning (Raven's Progressive Matrices), transitive inference, relational shape matching, and relational picture matching (see **Figure 2**). All tasks included a contrast between second-order and first-order relational reasoning conditions. For matrix reasoning, a second-order problem required consideration of both row and column to determine the correct missing element from a visuospatial array. For transitive inference, a second-order problem required combining multiple premises. The transitive inference task included problems that required consideration of directional (inequality) relations (pictured in **Figure 2**) as well as problems that required only consideration of non-directional (equality) relations. For both relational matching tasks, the second-order condition required participants to determine whether the top pair of stimuli matched along the same dimension as the bottom pair. All three of the above tasks involved visuospatial stimuli. By contrast, the relational picture matching task included evaluation of semantic relationships (pictured) as well as visuospatial relationships. We obtained contrast activation values for each participant, from each of the four studies, for each parietal ROI. We then submitted these contrast values to statistical analysis in SPSS, wherein we conducted an ANOVA that included parietal region, subdivision, and side as within-subjects factors and task/study as a between subjects factor.

For the broader analysis of reasoning-related activation and its relationship with other parietal functions, activation maps were obtained using Neurosynth, which provides automated meta-analyses based on Keywords (Yarkoni et al., 2011). The Neurosynth algorithm extracts clusters associated with specific key words across a large database (thousands) of neuroimaging studies. First, for a given key word (e.g., "reasoning"), it calculates frequency of appearance within an article, and identifies studies for which the key word appears at a high frequency (more than once per thousand words). Second, it automatically extracts activation coordinates from tables reported in these studies. Third, the set of coordinates extracted from studies that have been linked to a key word are submitted to multilevel kernel density analysis (MKDA) to produce activation maps (c.f. Wager et al., 2009). Finally, taking into consideration maps generated for a large number of different key words, machine learning (naïve Bayes classification) is used to estimate the likelihood that activations were associated with specific psychological terms.

In addition to "reasoning", we utilized the following terms associated with functions of PPC: "numerical" and "calculation" for mathematical cognition, "visuospatial" and "attention" for visuospatial processing and attention, and "phonological", "lexical", and "semantic" for language-related processes. We also examined activation maps associated with the terms "maintenance" and "manipulation" (working memory), and "memory encoding" and "memory retrieval" (long-term memory). **Table 1** gives the number of studies included for each term. For each of these terms, we obtained the reverse inference map, which displays regions that are reported more often in studies that load highly on the selected term than in studies that do not load highly on the term. In other words, the reverse inference maps display regions that are diagnostic of the term or feature. In addition, to obtain a broader representation of reasoning-related activation, we also obtained the forward inference map associated with reasoning. The forward inference map includes regions that are consistently activated in studies that load highly on the term.

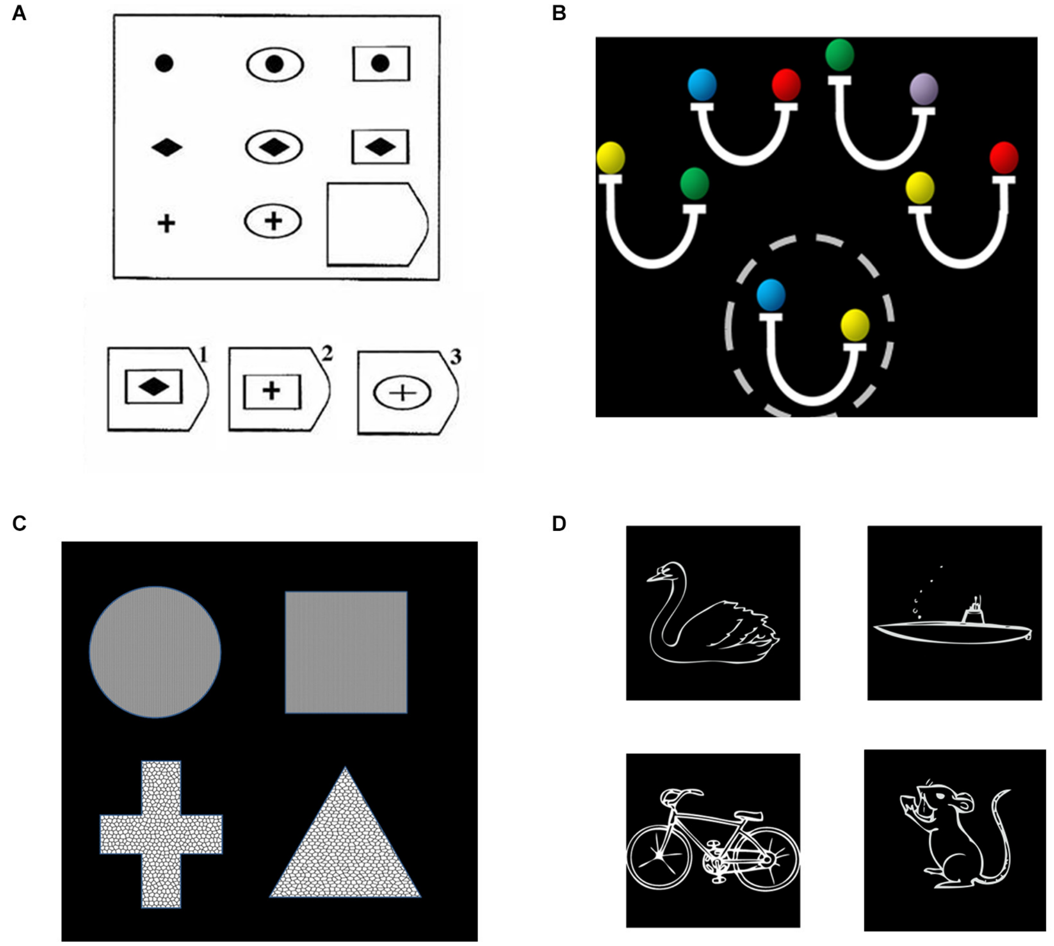


FIGURE 2 | Relational reasoning tasks, including (A) matrix reasoning; (B) transitive inference; (C) relational shape matching; and (D) relational picture matching. (A) For relational matching, the given stimulus depicts a second-order problem, in which one must consider the relationships in both the bottom row and rightmost column to determine that the correct answer is #2. **(B)** For transitive inference, a second-order problem is shown, for which one to evaluate the validity of the probe (circled, "yellow is heavier than blue"), one must combine both the second and fourth premises ("blue is same as red" and "yellow is heavier than red"). **(C)** For the relational matching task, equivalent stimuli were used across conditions. The given stimulus is a texture match, because the top two shapes share the same texture, a shape mismatch, because neither pair share the same shape, and a relational match, because the same dimension of match (texture) is present for both the top and bottom pairs. **(D)** The semantic picture matching task follows the same logic as the relational matching task, but utilized animal vs. vehicle and land vs. water as dimensions of possible match or mismatch. The example depicts a relational match, in that the dimension of match for the top pair (land vs. water) is the same as the dimension of match for the bottom pair.

Thus, the forward inference reasoning map included regions that are typically activated during reasoning tasks, not all of which are particularly diagnostic of reasoning.

Calculations of image characteristics were done using FSL (FMRIB Software Library, Oxford Center for Functional Magnetic Resonance Imaging of the Brain). We first computed, for each term, the extent of activation within each parietal ROI. Next, we computed overlap volume between each reasoning map (forward and reverse inference) and every other feature map (reverse inference only). This was done separately for each parietal ROI. From these initial values, we computed similarity scores relating the reasoning maps to every other feature. Similarity between two maps was defined as the volume of activation in the intersection of the two maps divided by the total volume of activation in the union of the two maps; thus, non-overlapping maps would have a similarity score of 0 and maps that are the same would have a similarity score of 1. We also computed the percentage of the reasoning activation that was accounted for by each feature; this differs from the similarity score in that a large activation cluster that effectively contains the reasoning cluster, but which includes many non-reasoning voxels as well, would have a high percent-of-reasoning score but a lower similarity score.

RESULTS

POSTERIOR PARIETAL ENGAGEMENT DURING RELATIONAL REASONING

First, we sought to characterize patterns of reasoning-related activation across the posterior parietal ROIs. **Figure 2A** shows average

Table 1 | The number of studies included in the Neurosynth meta-analysis, for each term.

Term	Number of studies
Reasoning	124
Visuospatial	184
Attention	1199
Memory retrieval	144
Memory encoding	101
Manipulation	204
Maintenance	224
Numerical	64
Calculation	55
Semantic	701
Phonological	260
Lexical	212

percent signal change in each ROI. Notably, there was engagement across posterior IPL, and to a lesser extent across left posterior SPL. We conducted an ANOVA that included parietal region (IPL or SPL), subdivision (1–5), and hemisphere (left or right) as within-subjects factors, and task (matrix reasoning, transitive inference, shape matching, or picture matching) as a between-subjects factor. First, there was a main effect of hemisphere ($F_{(1,65)} = 10.31$, $p = 0.002$), such that activation on the left was stronger than activation on the right. Second, there was a main effect of subdivision ($F_{(4,260)} = 17.64$, $p < 0.001$). *Post hoc* tests indicated that this was driven by greater activation in the middle and posterior subdivisions (c, d, and e) relative to the anterior subdivisions (a and b; all p's < 0.001). There was no main effect of region ($p > 0.2$). However, there was a significant region × subdivision interaction ($F_{(4,260)} = 3.62$, $p = 0.007$), such that increased activation for IPL vs. SPL was observed in the middle and posterior but not in the anterior subdivisions. There was also an interaction between subdivision and side ($F_{(4,12)} = 5.06$, $p = 0.01$), such that the increased activation within left vs. right PPC was strongest in the posterior subdivisions and was not present in the anterior subdivisions.

Although our purpose here was to determine commonalities across studies, we note that there were differences between these studies in terms of both the parietal subdivisions and hemisphere that were most strongly engaged, as reflected in a subdivision × task interaction ($F_{(12,260)} = 4.62$, $p < 0.001$) as well as a hemisphere × task interaction ($F_{(3,65)} = 7.01, p < 0.001$). Notably, the transitive inference task did not demonstrate the preferential engagement of more posterior subdivisions that was present for the other three tasks. Moreover, while three out of four tasks engaged left PPC more than right PPC, the picture matching task, which included a visuospatial component, engaged right PPC to a greater extent.

POSTERIOR PARIETAL REGIONS ASSOCIATED WITH REASONING AND OTHER TASKS

Next, we turned to the large-scale meta-analysis and examined the extent of reasoning-related activations within each posterior parietal ROI. For the reverse inference reasoning map, which shows voxels that are most selective for reasoning, activations were almost entirely limited to the third and fourth subdivisions of left IPL (IPLc and IPLd: 51% and 42% of total active voxels, respectively;). For the forward inference map, activations were more extensive (**Figure 3B**), with greater volume on the left vs. right (69% left; **Figure 4A**) and greater volume within IPL vs. SPL (77% IPL; **Figure 4B**). Again, active voxels were concentrated in left IPLc and IPLd (26% and 20% of active voxels, respectively), but also spread to IPLe as well as to the more posterior subdivisions of SPL.

No other tested function demonstrated a similar concentration of active voxels within left IPLc. Memory retrieval, like reasoning, had a large share of activated voxels in left IPLd; but unlike for reasoning, memory retrieval activations were more concentrated in left IPLe. **Figure 4** shows relative numbers of voxels for left vs. right PPC and for IPL vs. SPL, for each of the examined features. Language and memory activations, like reasoning, were heavily left-lateralized. In contrast, attention and visuospatial activations, as well as those for manipulation, were heavily right lateralized. Voxels associated with mathematical cognition as well as maintenance were evenly balanced across left and right. Memory retrieval and semantic processing, along with reasoning, demonstrated the strongest preferential engagement of IPL over SPL. In contrast, visuospatial processing and attention, as well as memory encoding, demonstrated notable preferential engagement of SPL.

SIMILARITY OF REASONING TO OTHER FUNCTIONS IN POSTERIOR PARIETAL CORTEX

Our primary Neurosynth-based analysis involved examination of overlap between the activation maps associated with reasoning and those associated with other parietal functions. For each function (i.e., key word), in relation to reasoning, we examined: (1) overlap volume; (2) percentage of the reasoning volume accounted for by the overlap ("percent-of-reasoning"); and (3) percentage of the total volume (for reasoning plus the function of interest) accounted for by the overlap ("similarity"). These measures were obtained for both the forward inference and reverse inference reasoning maps. Overall results for each of the three measures are presented in **Figure 5**.

For the forward inference reasoning map, the feature "numerical" demonstrated the greatest overlap with reasoning across PPC. It overlapped with a large proportion (>50%) of the reasoning activation in most of the parietal ROIs that we examined, except for left IPLd, where the reasoning activation was most extensive. The numerical map also demonstrated the greatest overall similarity to the reasoning map. After numerical, the feature with the second-greatest overlap with reasoning, and also the second-highest similarity score, was calculation. Thus, the math cognition measures were most closely related to reasoning.

In addition to the math cognition features, activation maps from four other features demonstrated notable overlap with the reverse inference reasoning map: attention, visuospatial, phonological, and maintenance. Among these features, the attention and visuospatial maps demonstrated the greatest overlap with reasoning on the right side, particularly in IPLd, IPLe, and SPLc. In contrast, among these four features, the phonological map demonstrated the greatest similarity to the reasoning on the left,

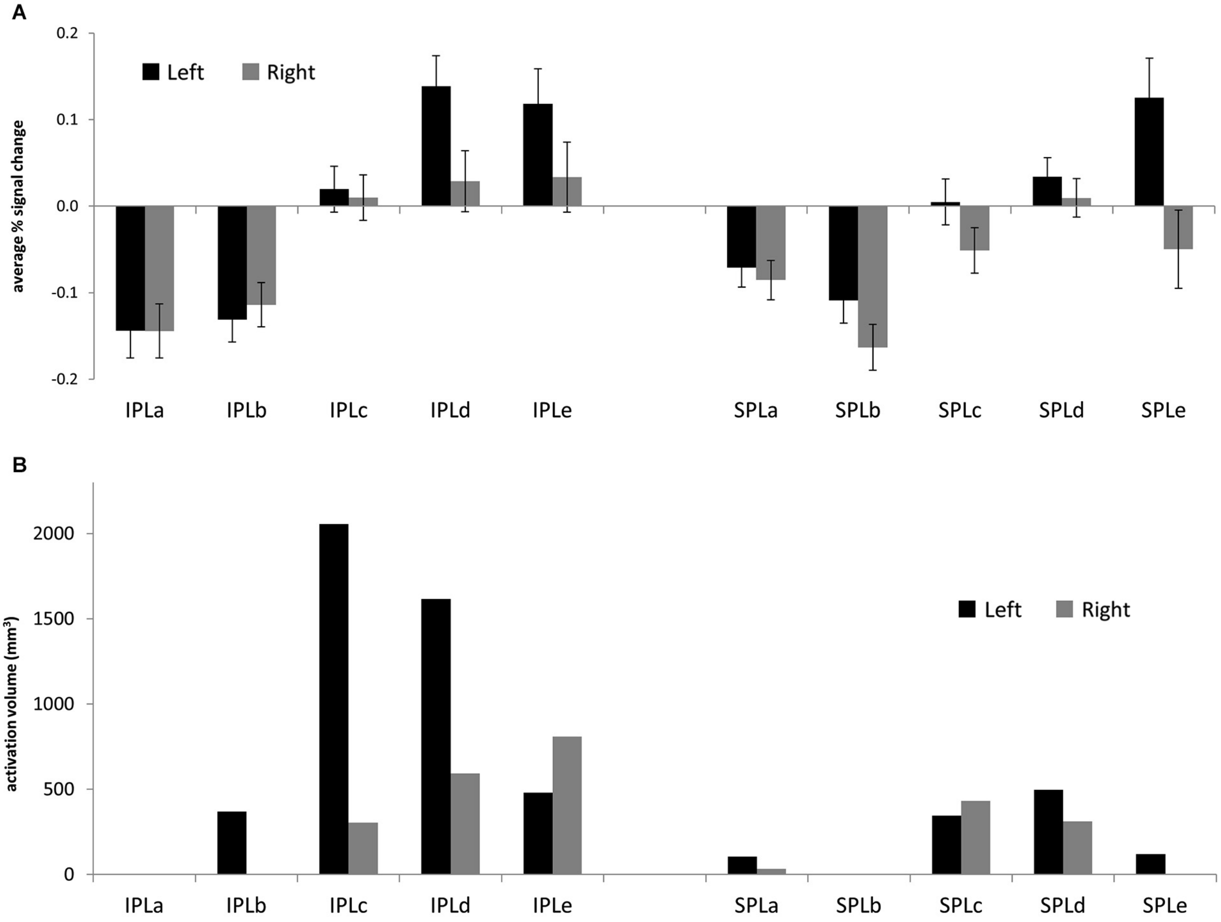


FIGURE 3 | (A) Average % signal change for the contrast between 2nd-order and 1st-order relational reasoning, across four separate studies (matrix reasoning, transitive inference, relational matching, picture matching). Error bars are standard error of the mean across subjects. **(B)** Volume of the Neurosynth forward inference map associated with reasoning, for each posterior parietal ROI.

and overlapped with nearly 50% of the reasoning activation in left SPL. The maintenance map demonstrated a more balanced pattern of similarity to the reasoning map, across the collection of parietal ROIs. For all of the other examined features, the percent-of-reasoning scores were less than 10%.

In addition to examining the forward inference activation map associated with reasoning, we also examined overlaps for the much smaller reverse inference reasoning map. Here again, numerical demonstrated the greatest overlap with reasoning, accounting for 24% of the overall reasoning activation and 25% of its activation within left IPL. The visuospatial map overlapped with 75% of the small reasoning activation within right IPL; however, it accounted for only 3% of the overall reasoning activation. In fact, no feature other than numerical accounted for more than 10% of the reasoning activation. Thus, a large proportion of the activation related to reasoning, particularly within IPLd, appears to be distinct from the activations associated with other parietal functions.

Notably, there was a substantial part of the reasoning activation that did not overlap with that for any other feature. This was particularly true within left IPLd, the region that demonstrated the greatest specificity for second-order relational reasoning in our own studies. The reasoning-specific activation cluster from the Neurosynth analysis is shown in **Figure 6.** Although we did not formally separate dorsal and ventral subdivisions of IPL, or position along the gyrus vs. position in the depth of the IPS, it is clear from the pattern of activations that reasoning-specific activation is concentrated in dorsal IPL, on the border of the IPS but not in the sulcus. By contrast, many other functions appear to overlap more ventrally, and within the depth of the sulcus.

DISCUSSION

The goals of the current study were to (1) better characterize the pattern of posterior parietal engagement during reasoning; and (2) to use this information, along with information about parietal engagement in other domains, to better understand the parietal contribution to reasoning.

PATTERNS OF PARIETAL ENGAGEMENT DURING REASONING

With regard to the first goal, we have obtained complementary evidence from two separate analyses that, within PPC,

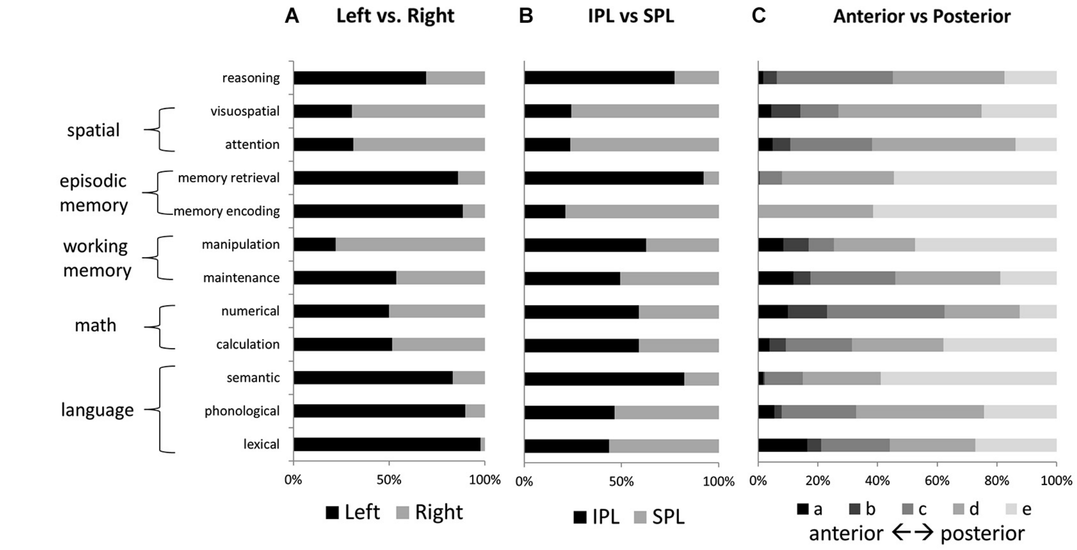


FIGURE 4 | Relative activation volumes associated with each term, for (A) left vs. right parietal cortex, and (B) IPL vs. SPL; and (C) anterior to posterior parietal cortex. All bars add up to 100%. Terms are grouped according to the higher-level category to which they are thought to correspond.

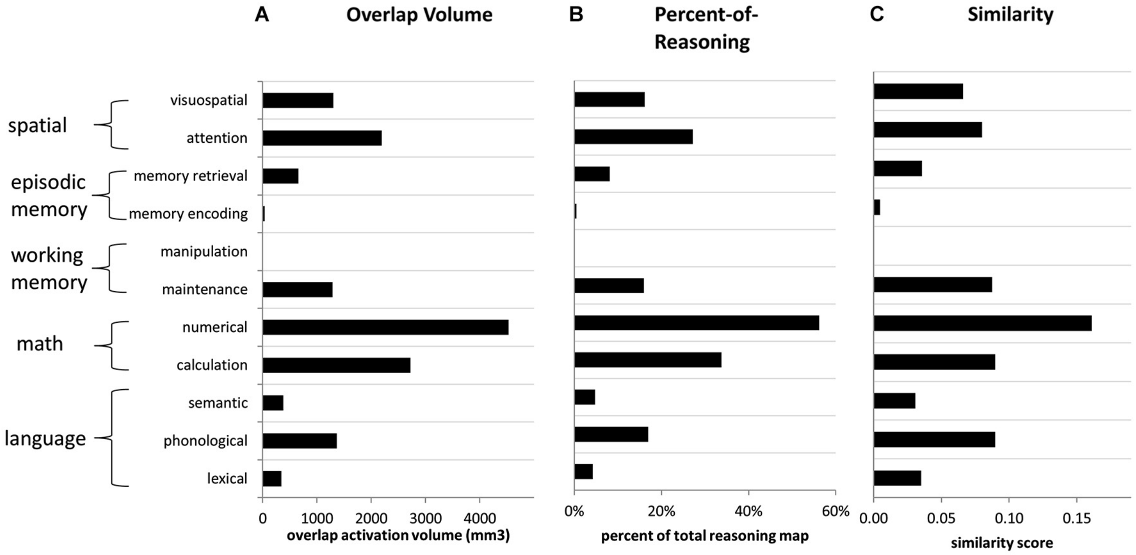


FIGURE 5 | Overlap metrics for reasoning (forward inference map) in relation to other functions (reverse inference maps). (A) Volume of overlap with the reasoning map, for the activation maps associated with each other term. **(B)** Percent of reasoning activation from the reasoning map that overlapped with each other feature, across all parietal ROIs. **(C)** Similarity scores comparing the reasoning map to the map associated with every other feature, across all parietal ROIs. Similarity was computed as the volume of overlap divided by the total volume of activation for both features.

reasoning is most strongly associated with activation of middle to posterior IPL, and to a lesser extent with neighboring regions of middle to posterior SPL. For both analyses that we performed—of average percent signal change across four studies of relational reasoning and of activation volumes associated with reasoning in a large-scale meta-analysis—IPL demonstrated greater involvement in reasoning than did SPL, and left PPC demonstrated greater involvement than right PPC. In both analyses, the anterior-most subdivisions of IPL and SPL demonstrated no involvement in reasoning. There were some differences between the two approaches, with regard to the pattern of involvement across posterior regions: the relational reasoning tasks tended to engage the more posterior regions to a greater extent, whereas activation volumes were greatest within the middle regions for the larger-scale meta-analysis.

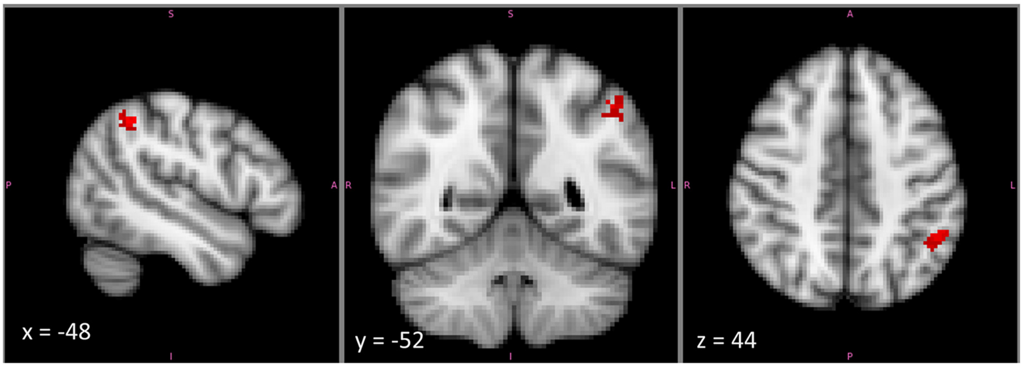


FIGURE 6 | The cluster within left mid-IPL (IPLc and IPLd) that was associated exclusively with reasoning and with no other examined term. This cluster lies on the upper part of the ventral bank of the intraparietal sulcus. Note that the image is displayed in radiological coordinates, with left and right reversed.

Both of our analyses here were focused on uncovering patterns of engagement that are common across reasoning tasks. But in addition to commonalities, we would expect, and indeed have observed, differences among different kinds of reasoning in their patterns of parietal activation. Notably, in our picture matching task, which included both visuospatial and semantic relational reasoning, we observed selectivity for higher-order visuospatial but not semantic reasoning in right PPC (Wendelken et al., 2012). In the transitive inference task, we observed stronger PPC activation for reasoning with inequalities than for reasoning with equalities, and argued that this was due to representation of the more specific inequality relationships in PPC (Wendelken and Bunge, 2010). Moreover, In a meta-analysis that directly examined different kinds of reasoning tasks, Prado et al. (2011) reported bilateral PPC activation during relational reasoning, and left PPC activation during propositional reasoning.

It is notable that the anterior subdivisions of both IPL and SPL, which were not associated with reasoning in the Neurosynth analysis, demonstrated reduced activation for second-order relative to first-order relational reasoning across our four reasoning tasks. These differences were largely driven by larger positive activations for the first-order relational task, and not by deactivation during second-order reasoning. However, this pattern of relatively reduced activation during the generally more difficult second-order reasoning condition in anterior PPC is consistent with participation this region in the default mode network (see Laird et al., 2009). Regions in the default mode network are typically deactivated during a wide spectrum of cognitively demanding tasks; thus, the deactivation in anterior PPC that we observe is likely to be non-specific to reasoning.

THE PARIETAL CONTRIBUTION TO REASONING

With regard to our second goal, evidence from the large-scale meta-analysis indicates clearly that the pattern of activation associated with reasoning is most closely related to that for mathematical cognition. There were also notable similarities between reasoning activations and those associated with visuospatial processing and attention, particularly on the right; between reasoning and phonological processing, particularly on the left; and between reasoning and working memory maintenance, bilaterally. These findings help to clarify the possible contributions of PPC to reasoning.

A key question is the extent to which reasoning is accomplished via mental logic and rule-following, on the one hand, or estimation and probabilistic computation, on the other. The current evidence clearly points towards the latter. Logical rule-following is posited to depend on formal language-like constructs, if not directly on linguistic representations. Although there was some similarity between reasoning and phonological activations, the overlap with mathematical cognition terms was much greater. Moreover, there was practically no parietal overlap between the reasoning map and maps associated with either lexical or semantic processing. In addition to a reliance on language-related processes, manipulation of formal logical rules can also be expected to depend heavily on processes that support manipulation in working memory. But here again, although there was some overlap between reasoning and working memory maintenance, there was practically no overlap between reasoning and manipulation. Thus, the current evidence points away from a logical rule-following as a primary mechanism for reasoning, and is more consistent with accounts that involve estimation and probabilistic computation.

An alternative explanation of the strong overlap between reasoning and math cognition is that, instead of reasoning relying on basic mathematical cognition, some types of mathematical cognition may rely on the capacity for reasoning. Indeed, advanced mathematical operations place a strong demand on reasoning, and math achievement in school is highly dependent on reasoning ability (Taub et al., 2008). It is entirely possible that some part of the overlap between reasoning and math-related activation reflects activation associated with mathematical reasoning. However, reasoning in math tasks is unlikely to fully explain the observed overlap, because the math cognition studies identified by the "numerical" and "calculation" keywords tend to involve simple tasks that put the greatest demand on basic numerical processes (e.g., magnitude estimation) and simple calculations, and put relatively less demand on reasoning.

The overlap between the reasoning map and the map associated with maintenance could reflect the importance of working memory as a component process of reasoning (Kyllonen and Christal, 1990; Salthouse, 1992). But the limited extent of this overlap argues against working memory as the main explanation for parietal engagement during reasoning. Similarly, overlap between the maps for reasoning and attention leaves open the possibility that part of the parietal activation for reasoning reflects attentional processes. Indeed, attentional processes are likely to be involved in many reasoning tasks. But here again, attention does not appear to be the primary explanation for parietal activation during reasoning.

Among potential parietal functions that we did not consider here, social cognition is worthy of mention. One recent meta-analyses highlights the tempo-parietal junction (TPJ), which includes ventral parts of IPL, as a key locus of social cognition, and points to overlap between the social cognitive function of TPJ and other parietal functions including language, memory, and attention (Carter and Huettel, 2013). However, while many of the functions that we examined do activate this ventral IPL/TPJ region, it is notable that reasoning does not, with reasoning activations mostly limited to the more dorsal parts of IPL on the border of the IPS. Notably, one class of social cognition studies—those using false belief stories—have been linked to dorsal IPL (Shurz et al., 2014). False belief studies probe the ability to reason about theory of mind. Thus, dorsal IPL activation in these studies may well be due to the reasoning demand inherent in this social cognitive task.

PARIETAL SPECIALIZATION FOR REASONING?

It is notable that a large part of the activation map for reasoning—particularly in left IPL in the vicinity of the IPS—did not overlap with the maps for any of the other functions that were considered here. Of course, it is possible that some other function of PPC, not considered here, may help to explain the engagement of this region for reasoning. But the current results are at least suggestive of the possibility that this reasoning-related activation represents a fairly narrow specialization of this part of PPC for reasoning processes.

This mid-IPL region that appears as unique for reasoning in our Neurosynth analysis is similar to the IPL activations that we typically observe in studies of relational reasoning, and in particular is consistent with the region for which we reported the strongest contrast activation in our small-scale meta-analysis of relational reasoning studies. We have previously argued that RLPFC, in the frontal lobe, is specialized for second-order relational reasoning. The current results are consistent with the possibility that RLPFC may share this duty with a subregion of mid-IPL. Although direct anatomical connections between RLPFC and mid-IPL have not been reported, it is noteworthy that these two regions demonstrate strong functional connectivity during task execution (Boorman et al., 2009; Wendelken et al., 2012) and even at rest (Vincent et al., 2008).

LIMITATIONS AND FUTURE WORK

It is important to note several limitations in the interpretation of our findings. Our first analysis involved only a small number of studies from our lab. This approach had the advantage, over typical larger-scale meta-analyses, of allowing for extraction of whole-brain contrast images based on complete data from each study. The similarity of the tasks—each involving a contrast between second-order and first-order relational reasoning—was a key advantage that enabled this analysis. However, the fundamental similarity of these tasks, coupled with the small number, limits the generalizability of our initial findings. Beyond the fact of the small number of studies included here, and the similarity of the tasks, all of these studies drew from a similar pool of participants (UC Berkeley undergraduates) and involved similar analytical methods. Moreover, and despite the fundamental similarity of the tasks, there was variation across these studies in terms of parietal activation, and the average activation measure that we examined here only tells part of the story.

While the Neurosynth approach allowed for analysis of a much larger set of studies, individual datapoints within this analysis are much less informative and reliable. There are a number of sources of potential error in the Neurosynth approach: (1) the identification of studies by keyword will lead to both false inclusions and omissions of relevant studies; (2) the identification of coordinates within a study is done without regard to any specific contrast; (3) there is no attempt to distinguish between activations and deactivations; and (4) as with any meta-analysis, there is an inherent confirmation bias, since results that do not fit prior expectations may not be reported. Moreover, while the selection of reasoning tasks examined by Neurosynth is considerably broader than the four relational reasoning tasks examined in our initial analysis, it may still be biased towards certain types of reasoning tasks. Despite these limitations, examinations of Neurosynth results have shown them to be very much in line with those of more traditional meta-analyses. Our side-by-side examination of results from Neurosynth and from our own reasoning studies was intended partly as a validation of the Neurosynth reasoning results, though we could not validate results for other keywords in a similar manner.

Because our focus in the current study was on the contribution of PPC, our results only speak to the PPC role in reasoning, and not to the contribution of other brain regions. Thus, while we interpret the current evidence as supporting the hypothesis that mathematical or probabilistic mechanisms underlie the parietal contribution to reasoning, they do not rule out the possibility that other mechanisms (e.g., linguistic) may support reasoning through the engagement of other brain regions.

The Neurosynth-based analysis does not distinguish between reasoning tasks that are by design deductive (where conclusions follow necessarily from the premises) or inductive (where uncertainty is an explicit part of the task). It is reasonable to suppose that differences in the extent of logical rule following vs. probabilistic calculation would be present for these different kinds of reasoning. But the extent to which human reasoners employ logical rule-following to solve nominally deductive tasks, or probabilistic computation to solve nominally inductive tasks, is unclear. Much of the debate on logical rule-following vs. probabilistic computation focuses specifically on deductive reasoning (Oaksford and Chater, 2009; Khemlani and Johnson-Laird, 2012), though this debate can also apply in the case of inductive reasoning, with tools like fuzzy logic providing a possible rule-based mechanism (Smithson and Oden, 1999). Understanding how the parietal contribution to reasoning might differ as a function of deductive vs. inductive reasoning is an open question and an important follow-up to the current results.

Limitations of the approach notwithstanding, these results demonstrate the value of Neurosynth as a tool. Rigorous meta-analyses have previously characterized patterns of activation associated with reasoning (e.g., Goel, 2007; Prado et al., 2011). But Neurosynth enabled direct comparison of activation maps for reasoning and a wide range of other functions, in a manner and at a scale that would be very difficult to achieve without the automation that it provides. One of the chief ways that neuroimaging work can inform psychological theory is by telling us which functions potentially utilize the same neural circuitry. Thus, the ability to characterize a pattern of activation associated with some function of interest in terms of its overlap with many other functional patterns may emerge as a fundamental analytical tool.

ACKNOWLEDGMENTS

I thank Silvia Bunge and Michael Vendetti for comments on this manuscript. Funding for this work was provided by a James S. McDonnell Foundation Scholar Award to Silvia Bunge.

REFERENCES

Ackerman, C. M., and Courtney, S. M. (2012). Spatial relations and spatial locations are dissociated within prefrontal and parietal cortex. *J. Neurophysiol.* 108, 2419–2429. doi: 10.1152/jn.01024.2011

Amorapanth, P. X., Widick, P., and Chatterjee, A. (2010). The neural basis for spatial relations. *J. Cogn. Neurosci.* 22, 1739–1753. doi: 10.1162/jocn.2009.21322

Awh, E., Jonides, J., Smith, E., Schumacher, E., Koeppe, R., Katz, S., et al. (1996). Dissociation of storage and Rehearsal in verbal working memory: evidence from Positron emission tomography. *Psychol. Sci.* 7, 25–31. doi: 10.1111/j.1467-9280.1996.tb00662.x

Baldo, J. V., Bunge, S. A., Wilson, S. M., and Dronkers, N. F. (2010). Is relational reasoning dependent on language? A voxel-based lesion symptom mapping study. *Brain Lang.* 113, 59–64. doi: 10.1016/j.bandl.2010.01.004

Becker, J. T., MacAndrew, D. K., and Fiez, J. A. (1999). A comment on the functional localization of the phonological storage subsystem of working memory. *Brain Cogn.* 41, 27–38. doi: 10.1006/brcg.1999.1094

Berryhill, M. E., and Olson, I. R. (2012). The right parietal lobe is critical for visual working memory. *Neuropsychologia* 46, 1767–1774. doi: 10.1016/j.neuropsychologia.2008.01.009

Binder, J. R., Desai, R. H., Graves, W. W., and Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* 19, 2767–2796. doi: 10.1093/cercor/bhp055

Boorman, E. D., Behrens, T. E., Woolrich, M. W., and Rushworth, M. F. (2009). How green is the grass on the other side? Frontopolar cortex and the evidnece in favor of alternative courses of action. *Neuron* 62, 733–743. doi: 10.1016/j.neuron.2009.05.014

Braine, M. D. S., and O'Brien, D. P. (eds) (1998). *Mental Logic.* Erlbaum.

Bueti, D., and Walsh, V. (2009). The parietal cortex and the representation of time, space, number and other magnitudes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1831–1840. doi: 10.1098/rstb.2009.0028

Bunge, S. A., Helskog, E. H., and Wendelken, C. (2009). Left, but not right, rostrolateral prefrontal cortex meets a stringent test of the relational integration hypothesis. *Neuroimage* 46, 338–342. doi: 10.1016/j.neuroimage.2009.01.064

Bunge, S. A., Wendelken, C., Badre, D., and Wagner, A. D. (2005). Analogical reasoning and prefrontal cortex: evidence for separable retrieval and integration mechanisms. *Cereb. Cortex* 15, 239–249. doi: 10.1093/cercor/bhh126

Cabeza, R., Ciaramelli, E., and Moscovitch, M. (2012). Cognitive contributions of the ventral parietal cortex: an integrative theoretical account. *Trends Cogn. Sci.* 16, 338–352. doi: 10.1016/j.tics.2012.04.008

Carruthers, P. (2002). The cognitive functions of language. *Behav. Brain Sci.* 25, 657–674; discussion 674–725. doi: 10.1017/S0140525X02000122

Carter, R. M., and Huettel, S. A. (2013). A nexus model of the temporal-parietal junction. *Trends Cogn. Sci.* 17, 328–336. doi: 10.1016/j.tics.2013.05.007

Caspers, S., Geyer, S., Schletcher, A., Mohlberg, H., Amunts, K., and Zilles, K. (2006). The human inferior parietal cortex: cytoarchitectonic parcellation and inter-individual variability. *Neuroimage* 33, 430–448. doi: 10.1016/j.neuroimage.2006.06.054

Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J. K., Holyoak, K. J., et al. (2001). Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *Neuroimage* 14, 1136–1149. doi: 10.1006/nimg.2001.0922

Chuderski, A. (2014). The relational integration task explains fluid reasoning above and beyond other working memory tasks. *Mem. Cognit.* 42, 448–463. doi: 10.3758/s13421-013-0366-x

Cohen Kadosh, R., Henik, A., Rubinsten, O., Mohr, H., Dori, H., van de Ven, V., et al. (2005). Are numbers special? The comparison systems of the human brain investigated by fMRI. *Neuropsychologia* 43, 1238–1248. doi: 10.1016/j.neuropsychologia.2004.12.017

Crone, E. A., Wendelken, C., van Leijenhorst, L., Honomichl, R. D., Christoff, K., Bunge, S. A., et al. (2009). Neurocognitive development of relational reasoning. *Dev. Sci.* 12, 55–66. doi: 10.1111/j.1467-7687.2008.00743.x

Dehaene, S., Piazza, M., Pinel, P., and Cohen, L. (2003). Three parietal circuits for number processing. *Cogn. Neuropsychol.* 20, 487–506. doi: 10.1080/02643290244000239

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., and Tsivkin, S. (1999). Sources of mathematical thinking: behavioral and brain-imaging evidence. *Science* 284, 970–974. doi: 10.1126/science.284.5416.970

Eslinger, P. J., Blair, C., Wang, J., Lipovsky, B., Realmuto, J., Baker, D., et al. (2009). Developmental shifts in fMRI activations during visuospatial relational reasoning. *Brain Cogn.* 69, 1–10. doi: 10.1016/j.bandc.2008.04.010

Gentner, D., and Holyoak, K. J. (1997). Reasoning and learning by analogy. *Am. Psychol.* 52, 32–34. doi: 10.1037//0003-066x.52.1.32

Goel, V. (2007). Anatomy of deductive reasoning. *Trends Cogn. Sci.* 11, 435–441. doi: 10.1016/j.tics.2007.09.003

Goel, V., Buchel, C., Frith, C., and Dolan, R. J. (2000). Dissociation of mechanisms underlying syllogistic reasoning. *Neuroimage* 12, 504–514. doi: 10.1006/nimg.2000.0636

Grefkes, C., and Fink, G. R. (2005). The functional organization of the intraparietal sulcus in humans and monkeys. *J. Anat.* 207, 3–17. doi: 10.1111/j.1469-7580.2005.00426.x

Halford, G. S., Wilson, W. H., and Phillips, S. (1998). Processing capacity defined by relational complexity: implications for comparative, developmental and cognitive psychology. *Behav. Brain Sci.* 21, 803–831; discussion 831–864. doi: 10.1017/s0140525x98001769

Hopfinger, J. B., Woldorff, M. G., Fletcher, E. M., and Mangun, G. R. (2001). Dissociating top-down attentional control from selective perception and action. *Neuropsychologia* 39, 1277–1291. doi: 10.1016/s0028-3932(01)00117-8

Humphreys, G. F., and Lambon Ralph, M. A. (2014). Fusion and Fission of cognitive functions in the human Parietal cortex. *Cereb. Cortex* doi: 10.1093/cercor/bhu198. [Epub ahead of print].

Husain, M., and Nachev, P. (2007). Space and the parietal cortex. *Trends Cogn. Sci.* 11, 30–36. doi: 10.1016/j.tics.2006.10.011

Johnson, J. D., Suzuki, M., and Rugg, M. D. (2013). Recollection, familiarity and content-sensitivity in lateral parietal cortex: a high-resolution fMRI study. *Front. Hum. Neurosci.* 7:219. doi: 10.3389/fnhum.2013.00219

Johnson-Laird, P. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference and Consciousness.* Cambridge, MA: Harvard University Press.

Johnson-Laird, P. N. (2001). Mental models and deduction. *Trends Cogn. Sci.* 5, 434–442. doi: 10.1016/s1364-6613(00)01751-4

Kertesz, A., and McCabe, P. (1975). Intelligence and aphasia: performance of aphasics on Raven's coloured progressive matrices (RCPM). *Brain Lang.* 2, 387–395. doi: 10.1016/s0093-934x(75)80079-4

Khemlani, S., and Johnson-Laird, P. N. (2012). Theories of the syllogism: a meta-analysis. *Psychol. Bull.* 138, 427–457. doi: 10.1037/a0026841

Kyllonen, P., and Christal, R. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence* 14, 389–433. doi: 10.1016/s0160-2896(05)80012-1

Laird, A. R., Eickhoff, S. B., Li, K., Robin, D. A., Glahn, D. C., and Fox, P. T. (2009). Investigating the functional heterogeneity of the default mode network using coordinate-based meta-analytic modeling. *J. Neurosci.* 29, 14496–14505. doi: 10.1523/jneurosci.4004-09.2009

Mars, R. B., Jbabdi, S., Sallet, J., O'Reilly, J. X., Croxson, P. L., Olivier, E., et al. (2011). Diffusion-weighted imaging tractography-based parcellation of the human parietal cortex and comparison with human and macaque resting-state functional connectivity. *J. Neurosci.* 31, 4087–4100. doi: 10.1523/jneurosci.5102-10.2011

Marshall, J. C., and Fink, G. R. (2001). Spatial cognition: where we were and where we are. *Neuroimage* 14(1 Pt. 2), S2–S7. doi: 10.1006/nimg.2001.0834

Marshuetz, C., Smith, E. E., Jonides, J., DeGutis, J., and Chenevert, T. L. (2000). Order information in working memory: fMRI evidence for parietal and prefrontal mechanisms. *J. Cogn. Neurosci.* 12(Suppl. 2), 130–144. doi: 10.1162/08989290051137459

Mesulam, M. M. (1981). A cortical network for directed attention and unilateral neglect. *Ann. Neurol.* 10, 309–325. doi: 10.1002/ana.410100402

Nelson, S. M., Cohen, A. L., Power, J. D., Wig, G. S., Miezin, F. M., Wheeler, M. E., et al. (2010). A parcellation scheme for human left lateral parietal cortex. *Neuron* 67, 156–170. doi: 10.1016/j.neuron.2010.05.025

Nelson, S. M., McDermott, K. B., Wig, G. S., Schlaggar, B. L., and Petersen, S. E. (2013). The critical roles of localization and physiology for understanding parietal contributions to memory retrieval. *Neuroscientist* 19, 578–591. doi: 10.1177/1073858413492389

Nickel, J., and Seitz, R. J. (2005). Functional clusters in the human parietal cortex as revealed by an observer-independent meta-analysis of functional activation studies. *Anat. Embryol.* 210, 463–472. doi: 10.1007/s00429-005-0037-1

Nieder, A., Diester, I., and Tudusciuc, O. (2006). Temporal and spatial enumeration processes in the primate parietal cortex. *Science* 313, 1431–1435. doi: 10.1126/science.1130308

Oaksford, M., and Chater, N. (2009). Précis of bayesian rationality: the probabilistic approach to human reasoning. *Behav. Brain Sci.* 32, 69–84; discussion 85–120. doi: 10.1017/S0140525X09000284

Paulesu, E., Frith, C. D., and Frackowiak, R. S. (1993). The neural correlates of the verbal component of working memory. *Nature* 362, 342–345. doi: 10.1038/362342a0

Penn, D. C., Holyoak, K. J., and Povinelli, D. J. (2008). Darwin's mistake: explaining the discontinuity between human and nonhuman minds. *Behav. Brain Sci.* 31, 109–130; discussion 130–178. doi: 10.1017/s0140525x08003543

Prado, J., Chadha, A., and Booth, J. R. (2011). The brain network for deductive reasoning: a quantitative meta-analysis of 28 neuroimaging studies. *J. Cogn. Neurosci.* 23, 3483–3497. doi: 10.1162/jocn_a_00063

Rips, L. J. (1994). *The Psychology of Proof.* MIT press.

Rosenberg-Lee, M., Chang, T. T., Young, C. B., Wu, S., and Menon, V. (2011). Functional dissociations between four basic arithmetic operations in the human posterior parietal cortex: a cytoarchitectonic mapping study. *Neuropsychologia* 49, 2592–2608. doi: 10.1016/j.neuropsychologia.2011.04.035

Sack, A. T. (2009). Parietal cortex and spatial cognition. *Behav. Brain Res.* 202, 153–161. doi: 10.1016/j.bbr.2009.03.012

Salthouse, T. A. (1992). Working-memory mediation of adult age differences in integrative reasoning. *Mem. Cognit.* 20, 413–423. doi: 10.3758/BF03210925

Seghier, M. L. (2013). The angular gyrus: multiple functions and multiple subdivisions. *Neuroscientist* 19, 43–61. doi: 10.1177/1073858412440596

Shurz, M., Radua, J., Aichhorn, M., Richlan, F., and Perner, J. (2014). Fractionating theory of mind: a meta-analysis of duntional brain imaging studies. *Neurosci. Biobehav. Rev.* 42, 9–34. doi: 10.1016/j.neubiorev.2014.01.009

Smithson, M., and Oden, G. (1999). "Fuzzy set theory and application in psychology," in *International Handbook of Fuzzy Sets and Possibility Theory*, eds D. Dubois and H. Prade (Amsterdam: Kluwer), 557–585.

Taub, G. E., Keith, T. Z., Floyd, R. G., and McGrew, K. S. (2008). Effects of general and broad cognitive abilities on mathematical achievement. *Sch. Psychol. Q.* 23, 187–198. doi: 10.1037/1045-3830.23.2.187

Todd, J. J., and Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature* 428, 751–754. doi: 10.1038/nature02466

Uncapher, M. R., and Wagner, A. D. (2009). Posterior parietal cortex and episodic encoding: insights from fMRI subsequent memory effects and dual-attention theory. *Neurobiol. Learn. Mem.* 91, 139–154. doi: 10.1016/j.nlm.2008.10.011

Vincent, J. L., Kahn, I., Snyder, A. Z., Raichle, M. E., and Buckner, R. L. (2008). Evidence for a frontoparietal control system revealed by intrinsic functional connectivity. *J. Neurophysiol.* 100, 3328–3342. doi: 10.1152/jn.90355.2008

Waechter, R. L., Goel, V., Raymont, V., Kruger, F., and Grafman, J. (2013). Transitive inference reasoning is impaired by focal lesions in parietal cortex rather than rostrolateral prefrontal cortex. *Neuropsychologia* 51, 464–471. doi: 10.1016/j.neuropsychologia.2012.11.026

Wager, T. D., Jonides, J., and Reading, S. (2004). Neuroimaging studies of shifting attention: a meta-analysis. *Neuroimage* 22, 1679–1693. doi: 10.1016/j.neuroimage.2004.03.052

Wager, T. D., Lindquist, M. A., Nichols, T. E., Kober, H., and Van Snellenberg, J. X. (2009). Evaluating the consistency and specificity of neuroimaging data using meta-analysis. *Neuroimage* 45(1 Suppl.), S210–S221. doi: 10.1016/j.neuroimage.2008.10.061

Wager, T. D., and Smith, E. E. (2003). Neuroimaging studies of working memory: a meta-analysis. *Cogn. Affect. Behav. Neurosci.* 3, 255–274. doi: 10.3758/CABN.3.4.255

Wagner, A. D., Shannon, B. J., Kahn, I., and Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. *Trends Cogn. Sci.* 9, 445–453. doi: 10.1016/j.tics.2005.07.001

Watson, C. E., and Chatterjee, A. (2012). A bilateral frontoparietal network underlies visuospatial analogical reasoning. *Neuroimage* 59, 2831–2838. doi: 10.1016/j.neuroimage.2011.09.030

Wendelken, C., and Bunge, S. A. (2010). Transitive inference: distinct contributions of rostrolateral prefrontal cortex and the hippocampus. *J. Cogn. Neurosci.* 22, 837–847. doi: 10.1162/jocn.2009.21226

Wendelken, C., Bunge, S. A., and Carter, C. S. (2008). Maintaining structured information: an investigation into functions of parietal and lateral prefrontal cortices. *Neuropsychologia* 46, 665–678. doi: 10.1016/j.neuropsychologia.2007.09.015

Wendelken, C., Chung, D., and Bunge, S. A. (2012). Rostrolateral prefrontal cortex: domain-general or domain-sensitive? *Hum. Brain Mapp.* 33, 1952–1963. doi: 10.1002/hbm.21336

Wu, C. Y., Ho, M. H. R., and Chen, S. H. A. (2012). A meta-analysis of fMRI studies on Chinese orthographic, phonological and semantic processing. *Neuroimage* 63, 381–391. doi: 10.1016/j.neuroimage.2012.06.047

Xu, Y., and Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature* 440, 91–95. doi: 10.1038/nature04262

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., and Wager, T. D (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods.* 8, 665–670. doi: 10.3389/conf.fninf.2011.08.00058

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: 16 October 2014; accepted: 13 December 2014; published online: 21 January 2015.

Citation: Wendelken C (2015) Meta-analysis: how does posterior parietal cortex contribute to reasoning? Front. Hum. Neurosci. ***8****:1042. doi: 10.3389/fnhum.2014.01042*

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

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 19 January 2015
doi: 10.3389/fnhum.2014.01061

Reasoning with linear orders: differential parietal cortex activation in sub-clinical depression. An fMRI investigation in sub-clinical depression and controls

Elanor C. Hinton[1], Richard G. Wise[2], Krish D. Singh[2] and Ulrich von Hecker[3]*

[1] *Clinical Research and Imaging Centre, University of Bristol, Bristol, UK*
[2] *Cardiff University Brain Research Imaging Centre (CUBRIC), School of Psychology, Cardiff University, Cardiff, UK*
[3] *School of Psychology, Cardiff University, Cardiff, UK*

Edited by:
Vinod Goel, York University, Canada

Reviewed by:
Carter Wendelken, University of California Berkeley, USA
Randall Waechter, St. George's University, Grenada
Thomas Fangmeier, University Clinic Freiburg, Germany

Correspondence:
Ulrich von Hecker, School of Psychology, Tower Building, Park Place, Cardiff CF10 3AT, UK
e-mail: vonheckeru@cardiff.ac.uk

The capacity to learn new information and manipulate it for efficient retrieval has long been studied through reasoning paradigms, which also has applicability to the study of social behavior. Humans can learn about the linear order within groups using reasoning, and the success of such reasoning may vary according to affective state, such as depression. We investigated the neural basis of these latter findings using functional neuroimaging. Using BDI-II criteria, 14 non-depressed (ND) and 12 mildly depressed volunteers took part in a linear-order reasoning task during functional magnetic resonance imaging. The hippocampus, parietal, and prefrontal cortices were activated during the task, in accordance with previous studies. In the learning phase and in the test phase, greater activation of the parietal cortex was found in the depressed group, which may be a compensatory mechanism in order to reach the same behavioral performance as the ND group, or evidence for a different reasoning strategy in the depressed group.

Keywords: fMRI, sub-clinical depression, reasoning

INTRODUCTION

A fundamental ability in both humans and animals is the capacity to flexibly learn new information and to recall and manipulate that information for future use (Simons and Spiers, 2003; Manns and Eichenbaum, 2006). Indeed, both humans and animals can flexibly make novel inferences from the information provided (Dickins, 2005; Vasconcelos, 2008). This process is often studied through linear-order reasoning paradigms (Potts, 1972; Sternberg, 1980), in which participants learn A > B and B > C; evidence of reasoning occurs when they can rearrange the incoming information into a coherent representation, or mental model, in order to infer that A > C. This type of reasoning is not purely an abstract cognitive process, however, but one which has applications in the environment; for example, animals use this type of processing to learn their place in the social order of their groups (Hogue et al., 1996; Paz-Y-Miño et al., 2004). Humans can learn about rank orders within groups of people using linear-order reasoning, and the success of such reasoning may depend on affective state, particularly, sub-clinical depression (Sedek and Von Hecker, 2004). Previous research has found dysfunctions in the frontoparietal network in depressed participants [for an overview see Brzezicka (2013)]. In particular, Thomas and Elliott (2009), as well as Hugdahl et al. (2004) found in their depressed participants that reduced parietal activity was associated with impaired performance in mental arithmetic tasks, as well as hyperactivity was associated with intact performance, leading these authors to conclude that normal performance in depression is associated with enhanced cortical, in particular parietal, function during reasoning. In this study, we use functional MRI to investigate how brain activation during execution of a different reasoning task, that is, linear-order construction, might be altered in the brain, especially in parietal cortical areas, when individuals are in a state of sub-clinical depression.

There is an increasing literature on the neural basis of linear order, or transitive, reasoning [e.g., Christoff et al. (2001), Goel and Dolan (2001, 2003, 2004), Acuna et al. (2002), Knauff et al. (2002), Fangmeier et al. (2006), Greene et al. (2006), Monti et al. (2007), Van Opstal et al. (2008), Wendelken et al. (2008)]. Studies to date have largely taken an abstract form in the tasks employed to reveal the underlying brain activation of making inferences. A review of the above literature demonstrates that a "network" of brain regions subserve reasoning, including the hippocampus, parietal, and prefrontal cortices. Knauff et al. (2002) found that an occipital–parietal–frontal network was activated during relational reasoning, which includes areas in the visuospatial system. In line with this research, we suggest that spatial processing of relations is paramount to processing orders or hierarchies in order to solve reasoning problems (Leth-Steensen and Marley, 2000). Specifically, the present study will look into the areas of intra-parietal sulcus, inferior parietal lobe (BA 40), and posterior parietal lobe (BA 7) as earlier work has suggested that these regions might be involved in tasks involving spatial and numerical operations, as well as working memory [e.g., D'Esposito et al. (1998), Sakai et al. (1998), Pinel et al. (2001)], and, more specifically, in the spatial operations during transitive inference (Goel and Dolan, 2001; Acuna et al., 2002; Knauff et al., 2002). Furthermore, the role of the prefrontal cortex (PFC) in reasoning has been highlighted in studies of relational complexity and integration (Christoff et al., 2001; Acuna et al., 2002; Kroger et al., 2002; Wendelken et al., 2008).

In the present study, rather than employing abstract symbols in the task, we focus on more naturalistic linear-orders regarding relationships within small sets of people. During functional magnetic resonance imaging (fMRI), participants learned a series of pairwise information, such as "Andrew is taller than Brian," "Brian is taller than Colin," and "Colin is taller than David." Evidence suggests that people spontaneously rearrange the three presented pairs of information and integrate them into a coherent mental model (≥"taller"): A > B > C > D, most likely involving spatial representations (Huttenlocher, 1968; Waltz et al., 1999). After the learning phase, test queries were asked about all possible pairs of names, such as the three presented ones, i.e., A/B, B/C, C/D, and also queries about those relations that were not presented during learning, such as A/C, B/D (an inference spanning two distance steps along the assumed mental model), and A/D (involving two inferences, and corresponding to three distance steps along the model).

There is some evidence to suggest that transitive reasoning is affected by sub-clinical depression (Sedek and Von Hecker, 2004). Such reasoning deficits may lie at the heart of some cognitive problems found in those with depression, such as loss of creativity and inferior ability to solve problems in the social domain (Gotlib and Hammen, 1992; Marx et al., 1992; von Hecker and Sedek, 1999). Depressed participants showed inferior performance as compared to non-depressed (ND) controls in the linear-order task as described above, especially concerning the inferred pairs (Sedek and Von Hecker, 2004, Exp. 1, 3, and 4). The authors suggested that while ND individuals might create the comprehensive model A > B > C > D spontaneously during learning, depressed individuals might not do so (or not be successful in doing so), but engage in reasoning more upon particular queries during the test phase, this resulting in a less efficient processing overall. The present hypothesis, therefore, is that compared to those without depression, individuals in depressed states may show higher indices of brain activation in the spatial areas supporting transitive reasoning as described above, when tested on queries of any pair distance across the linear-order A > B > C > D.

MATERIALS AND METHODS

PARTICIPANTS

Female participants were recruited into this study on the basis of their score on the Beck depression inventory-II (Beck et al., 1996). Only females were recruited for this study, as there is a greater prevalence of depression in females (Nolen-Hoeksema, 2002). Participants attended one or two sessions. In the first session, participants were given the BDI and CED depression scales, and the operation span (OSPAN) and digital symbol substitution test (DSST) tasks (see below for details). Participants who fitted the BDI criteria for the ND or D groups in the first session were asked to attend a second session 1 week later. In the second session, participants were given both depression scales again. If their scores allowed them to remain in their original group classification, they immediately took part in the imaging phase. If not, the reasons for them not continuing onto the imaging session were given, and they were thanked and debriefed. For the ND group, those with a score of 5 or below, on two occasions 1 week apart, were chosen ($n = 17$). Those with a score of 13 or above, on two occasions 1 week apart, were included in the mildly depressed (D) group ($n = 15$). Participants were given a second depression scale (Center for Epidemilogic Studies Depression scale, CES-D, Radloff, 1977) in the second session, on which participants had to get a score of 16 or above to remain in the D group.

Data from three participants from the D group and three from the ND group had to be excluded from the analysis either due to excessive movement in the scanner or misunderstanding the task instructions. Twenty-six participants remained in the analysis: 14 in the ND group and 12 in the D group. **Table 1** summarizes the group demographics. All participants indicated that they were right-handed, none had any history of psychiatric or neurological disorders, and none were currently taking psychotrophic medications. All participants gave informed consent. This study was approved by the Cardiff School of Psychology Ethics Committee.

BEHAVIORAL TASKS AND DESIGN

During the fMRI, a mixed block/sparse event-related design was used to present the linear-order reasoning task (**Figure 1**). As described above, participants were shown information regarding the relationships between four people (A > B > C > D), upon which they were then tested. In the initial learning phase, presented as a block, participants were sequentially shown three sentences for 10 s each, followed by 30 s of fixation to a cross (X). Participants were asked to remember the names and the relationships between them, e.g., of one set (1) Andrew is taller than Brian (A > B) (2) Brian is taller than Colin (B > C) (3) Colin is taller than David (C > D). Other relational terms included "older," "richer," "smarter," "braver," and "faster" (18 in total). Relational pairs were presented in equal numbers of one of two order types: (i) where the pairs are presented in the order in which they appear in the putative model (e.g., A > B, B > C, C > D), or (ii) where the relations appear in a different order to the model (e.g., B > C, A > B, C > D), in order to assess whether the latter required differential brain activity to support the greater cognitive demands to support the integration of pairs. A test phase followed in which a query sentence was presented for 4.5 s followed by 10 s of fixation to allow the BOLD response to return to baseline between events. Three query sentences were presented in each test phase: One sentence was randomly chosen from those presented in the learning phase

Table 1 | Participant information.

	Non-depressed group (ND)	Depressed group (D)
N	14	12
Age (years)	22.6 (3.9)	22.7 (5.7)
BDI-II at time 1	2.3 (1.6)	16.6 (2.9)*
BDI-II at time 2 (imaging)	0.9 (1.1)	20.6 (7.0)*
CES-D at time 2 (imaging)	1.7 (2.2)	23.9 (6.9)*
DSST score	48.4 (6.8)	48.7 (10.1)
OSPAN words score (WM)	11.9 (6.3)	12.0 (7.8)
OSPAN maths score	37.2 (5.2)	37.4 (2.7)

*(SD given in brackets) *indicates a significant difference at the level of p < 0.05 using an independent groups t test.*

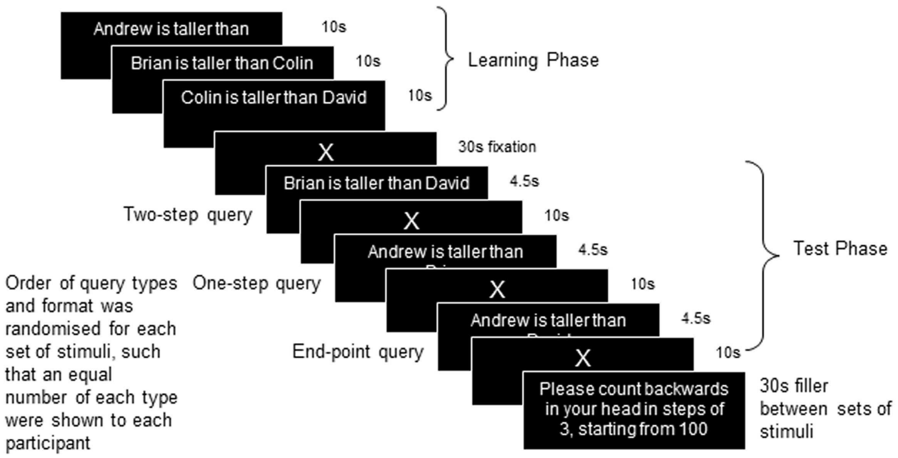

Andrew is taller than
10s
Brian is taller than Colin
10s
Colin is taller than David
10s
Learning Phase
X
30s fixation
Brian is taller than David
4.5s
Two-step query
X
10s
Andrew is taller than
4.5s
Test Phase
Order of query types and format was randomised for each set of stimuli, such that an equal number of each type were shown to each participant
One-step query
X
10s
Andrew is taller than
4.5s
End-point query
X
10s
Please count backwards in your head in steps of 3, starting from 100
30s filler between sets of stimuli

FIGURE 1 | Design of the study.

("one step" queries, A > B, B > C, or C > D, equivalent to one step (A to B) on the hypothetical mental model), one sentence was randomly chosen from "two-step" queries (A > C, B > D), and the end-point query was presented (A > D). The queries were either presented in correct or incorrect format (e.g., Andrew is taller than Brian, or Brian is taller than Andrew). Participants had to respond whether or not the query content was correct on the basis of the information learned about the group of people in the learning phase. Twelve sets of stimuli were presented in three imaging runs of four sets. The format of the test phase queries were pseudo-randomized such that over the course of the three runs there were an equal number (12) of each type of query (one step, two step, and end point), and an equal number of correctly (6),and incorrectly presented trials (6). Total scan time was approximately 30 min, followed by an anatomical brain scan for a further 10 min.

Data from the reasoning task were analyzed using ANOVA. The dependent measures were the percentage of correct responses and response time (within the 4.5 s window) to each query type (one step, two step, and end point). Following the imaging session, participants were given a post-imaging questionnaire, designed to ascertain how participants reported doing the task. Participants were also given an OSPAN task (Turner and Engle, 1989) as a measure of working memory capacity, and the DSST, a subset of the WAIS-R (Wechsler, 1981), as a measure of processing speed. There were no significant differences between the groups on these two control measures (see **Table 1** for means; OSPAN $t = -0.031$, $p > 0.05$; DSST $t = -0.071$, $p > 0.05$), so any differences found on the reasoning task cannot be attributed to differences in processing speed or working memory capacity.

IMAGE ACQUISITION

Anatomical and functional images were acquired at the Cardiff University Brain Research Imaging Center (CUBRIC), using a General Electric Excite-HDx 3 T MRI scanner. Functional images were collected using a gradient-echo echo-planar pulse sequence (TE = 35 ms; TR = 2500 ms; flip angle = 90°; acquisition matrix = 64 × 64; field of view (FOV) 64 × 64; in plane resolution 3.75 mm). The volumes covered the whole brain in 37 slices (thickness 3.8 mm) and were acquired in line with the anterior commissure/posterior commissure line. A total of 684 volumes were acquired for each participant in 3 sessions of 228 volumes each. In each run of 228 volumes, 3 sets of stimuli were presented. For each set (as described above, see **Figure 1** for presentation timing of one set of stimuli), a learning phase of 3 premise pair sentences (e.g., A > B, B > C, C > D), each presented for 10 s, was followed by a fixation cross for 30 s. The test phase then immediately followed with 3 test queries (4.5 s each), each followed by 10 s fixation. A filler task (counting backwards for 30 s) was given to participants between each set in order to reduce possible interference between sets of relations. This results in 12 block scans for analysis of the learning phase, and 12 one-step, 12 two-step, and 12 end-point test queries for analysis of the test phase. The timing of the program in presentation was designed such that the test queries were not presented until a pulse had been received by the scanner. This ensured that the task was always in synchrony with the scanner. Finally, a high-resolution T1-weighted FSPGR anatomical image was acquired (TR = 7.9 s; TE = 3 ms; inversion time = 450 ms; flip angle = 20°; acquisition matrix 256 × 256 × 176; FOV 256 × 256 × 176, resulting in 1 mm isotropic voxels).

IMAGE ANALYSIS

Data was analyzed using the FSL package from FMRIB, University of Oxford (http://www.fmrib.ox.ac.uk/fsl/). For each participant, data were acquired in three runs. At the first level, each run was pre-processed and analyzed separately, using the following stages: motion correction using MCFLIRT (Jenkinson et al., 2002), non-brain removal using BET (Smith, 2002), spatial smoothing using a Gaussian kernel of FWHM 5 mm, mean-based intensity normalization of all volumes, and high-pass temporal filtering. Time-series statistical analysis was carried out using FILM with local autocorrelation correction (Woolrich et al., 2001). The first level modeled nine explanatory variables (EVs) for learning phase order 1 and 2, the filler task between sets, one-step, two-step, and end-point test queries presented in the correct or incorrect format. Contrasts compared: (1) learning phase to baseline, (2) the two different order of premises in the learning phase, (3) each test query type to baseline, (4) one-step to two-step queries, and (5) presented (one step) vs. inferred queries (two step and end point). At the second level, the separate runs were combined into a fixed analysis for each person, and then finally data from all participants was combined in a third level analysis for each contrast. Higher-level group analysis was carried out using a mixed effects group analysis – FLAME (stage 1 only) (Beckmann et al., 2003; Woolrich et al., 2004). Z statistic images were thresholded using Gaussian random field (GRF)-theory based maximum cluster thresholding with a corrected significance threshold of $p = 0.05$ (Worsley et al., 1992). Registration to high resolution and standard images was carried out using FLIRT (Jenkinson and Smith, 2001; Jenkinson et al., 2002).

The study was designed to examine differences between groups (ND and D) and between test relation types (one step, two step, and end point). For the learning phase data, contrasts examined (i) activation during the learning phase compared to baseline (fixation cross) between the two groups, and (ii) activation during the learning phase for each order of presented relations, using a whole-brain corrected cluster-based threshold ($z > 2.3$, $p < 0.05$). A subsequent analysis repeated (i), but for both groups together. When reporting data for both groups together a stricter threshold ($z > 5$) was chosen due to the large extent of activation found when simply comparing task to fixation baseline.

For the test phase data, only correctly answered trials were included in the analysis. This resulted in 5.3% of the total number of trials being excluded from the analysis. Contrasts examined (iii) each test query type compared to baseline across groups, (iv) previously presented queries compared to queries requiring inference, and most importantly (v) between group differences for each test query type (one step, two step, and end point). The MNI coordinate system is used in the results section when reporting the activation peaks.

RESULTS

BEHAVIORAL DATA

Reasoning accuracy

The percentage of correct responses to the test queries is shown in **Figure 2A**. The main effect of pair distance (step) was significant ($F_{2,48} = 5.061$, $p = 0.01$), with accuracy increasing from one step queries to end-point queries. However, there were no significant differences in task accuracy between mood groups, between neighboring distances (one step/two step or two step/end point), or any interaction between group and pair distance.

Response times data

A significant stepwise decrease in reaction time was found across query types of increasing pair distance ($F_{2,48} = 11.30, p < 0.001$) – see **Figure 2B**. Pairwise comparisons showed that end-point queries needed significantly less time than two-step queries ($p = 0.005$), while the difference between one-step and two-step queries was not significant. There was also no significant difference between mood groups or any interaction between group and pair distance.

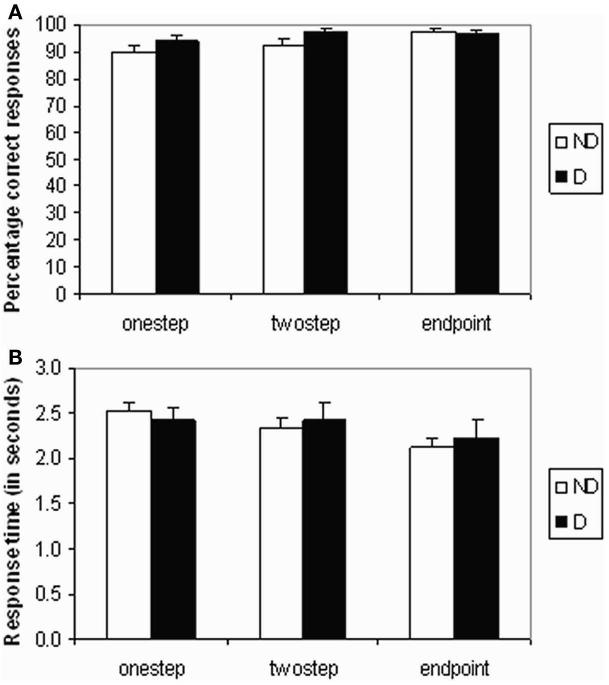


FIGURE 2 | Behavioral data. **(A)** Mean accuracy scores (percentage of correct responses) and **(B)** mean response time in seconds for each test query type (one step, two step, or end point) and for each group (non-depressed or depressed).

Questionnaire responses

All 26 participants reported, without prompting, that they had ordered the people in each set according to the relation specified between them, during the learning phase. Twenty-four out of 26 participants reported verbally rehearsing the correct order of the people in each set during the fixation between the learning and test phases; of the remaining participants, 1 reported using a purely visual strategy, and the other reported simply fixating on the cross.

NEUROIMAGING DATA

Learning phase

No significant differences were found in the whole-brain analyses brain activation between D and ND groups, while the participants were learning the relations between the people in each group. Moreover, no significant differences were found according to the order of presenting the relational pairs. As such, the following results are reported including all 26 participants and both order types using a whole-brain corrected cluster-based threshold ($z > 5$, $p < 0.05$). A distributed network of areas was activated in association with the learning phase of the task relative to fixation (see Table S1 in Supplementary Material), including prefrontal and parietal cortex, hippocampus, as well as occipital cortex and cerebellum. (NB. *Post hoc* ROI analyses of the learning phase are presented below).

Test phase

First, to investigate the basic pattern of activation associated with the test phase queries, an average map of the activation found in association with each type of test query relative to fixation (one step, two step, end point) for both D and ND groups together is reported, which revealed a similar pattern across the query types. For summary purposes, Table S2 in Supplementary Material contains the results from all 26 participants together, using a whole-brain corrected cluster-based threshold ($z > 5$, $p < 0.05$).

In order to examine, which areas were involved in making inferences, a further comparison was made between the response to test relations involving making an inference (two-step and end-point relations) and those involving one-step relations that would require recalling the previously presented information from the learning phase. A significant difference in brain activation between inferred and presented queries was found in the ND group only, using a whole-brain corrected cluster-based threshold ($z > 2.3$, $p < 0.05$). As shown in **Figure 3**, greater activation was found in the superior and medial frontal cortex in association with inferred queries (e.g., A > C, A > D) compared to the previously presented queries (e.g., A > B). The same regions were not significantly differentially activated in the depressed group for the same contrast. However, the direct comparison between groups did not reach significance [ND(inferred-presented) − D(inferred-presented)]. It is possible that the frontal cortex was activated more to inferred queries than presented queries in the depressed group as well, but that this difference in activation did not reach significance[1].

One of the key contrasts of interest in this study was to investigate differences in activation in response to the different test relations, relative to fixation, between the D and ND groups [D(test-fixation) − ND(test-fixation)]. Activation associated with each test query was analyzed between groups, using a whole-brain corrected cluster-based threshold ($z > 2.3$, $p < 0.05$). A significantly different pattern of activation was found in the parietal lobe/post-central gyrus for end-point and one-step queries between groups (D–ND), as shown in **Figure 4A** (end point), **Figure 4C** (one step). For end-point queries, foci were found in superior parietal cortex (26, −46, 60, $z = 3.72$), supramarginal gyrus/post-central gyrus ($x = 44$, $y = -26$, $z = 40$, $Z = 3.76$). For one-step queries, foci were found in the parietal lobe (post-central gyrus $x = 60$, $y = -12$, $z = 20$, $Z = 3.95$; 58, −1, 46, $z = 3.87$). **Figure 4B** (end point) and **Figure 4D** (one step) show how activity in these regions varies as a function of BDI-II score. These scatter-plots show that on average the ND group shows relative deactivation in these regions, whereas the D group show activation. A similar pattern of activation was found in the two-step contrast as for the other test query types (as shown in Table S2 in Supplementary Material above), and a D–ND difference was found for two-step queries in the same areas as for end-point and one-step queries after lowering the threshold slightly, suggesting that any difference between the groups did not quite survive the cluster threshold for two-step queries.

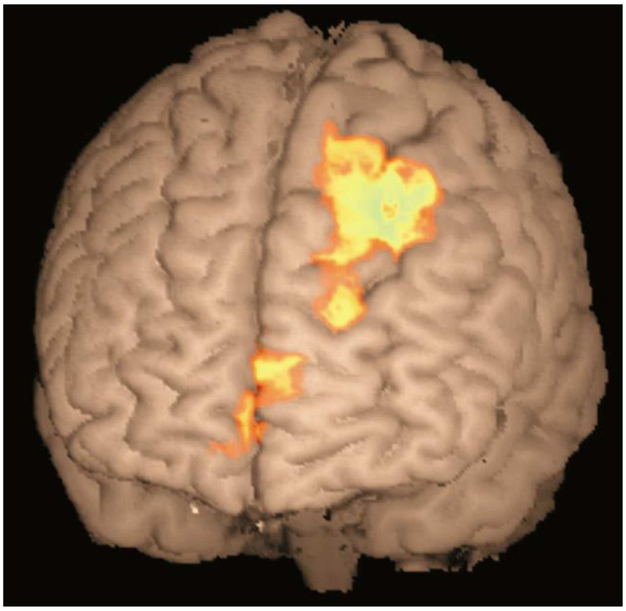


FIGURE 3 | Activation map for inferred queries compared to previously presented queries in ND group. Activation shown in medial frontopolar cortex (BA 10, peak −4, 64, 14, $z = 3.26$) and superior frontal cortex (BA8, peak −26, 32, 50, $z = 3.63$) in the contrast between inferred queries (two step and end point) compared to previously presented queries (one step), in the non-depressed group only. Cluster-based threshold: $z > 2.3$, $p < 0.05$.

To attempt to further understand the nature of these differences, correlations were performed between activity during end-point and one-step queries in the regions showing a significant difference between groups, and performance on the task (**Figure 5**). A significant negative correlation was found in the D group between activity and response times to end-point queries ($r = -0.579$, $p = 0.048$). The longer the response time, the less activity was found in the parietal regions showing a difference between groups. As **Figure 5** shows, this correlation was only found in the D group, with no such relationship in the ND group ($r = -0.009$, $p = 0.975$).

Given the difference in the parietal cortex response between groups during the test phase, further *post hoc* analyses were conducted in order to test for differences during the learning phase in this parietal region. Two separate masks of the parietal activation showing differences between the D and ND groups in response to one-step and end-point queries were created. In two separate analyses, these were inputted into the learning phase group feat analysis using pre-threshold masking. For both types of queries (one step and end point), the D group do show significant activation in the corresponding parietal region during the learning phase, whereas the ND group do not. The contrast between the two groups (D–ND) does show a significant difference in this region of the parietal cortex during the learning phase ($x = 48$, $y = -28$, $z = 40$, $Z = 3.5$). The same activation peak during the learning phase was seen using the one-step and end-point parietal cortex masks, as these masks almost entirely overlap.

[1] We thank the reviewer for pointing out that the inferred vs. presented contrast can be difficult to interpret since the recall can interfere with the model creation and the inference process. Maybe the participants hold the premises in mind and rehearse them more or less extensively, which may interfere with later stages of model creation.

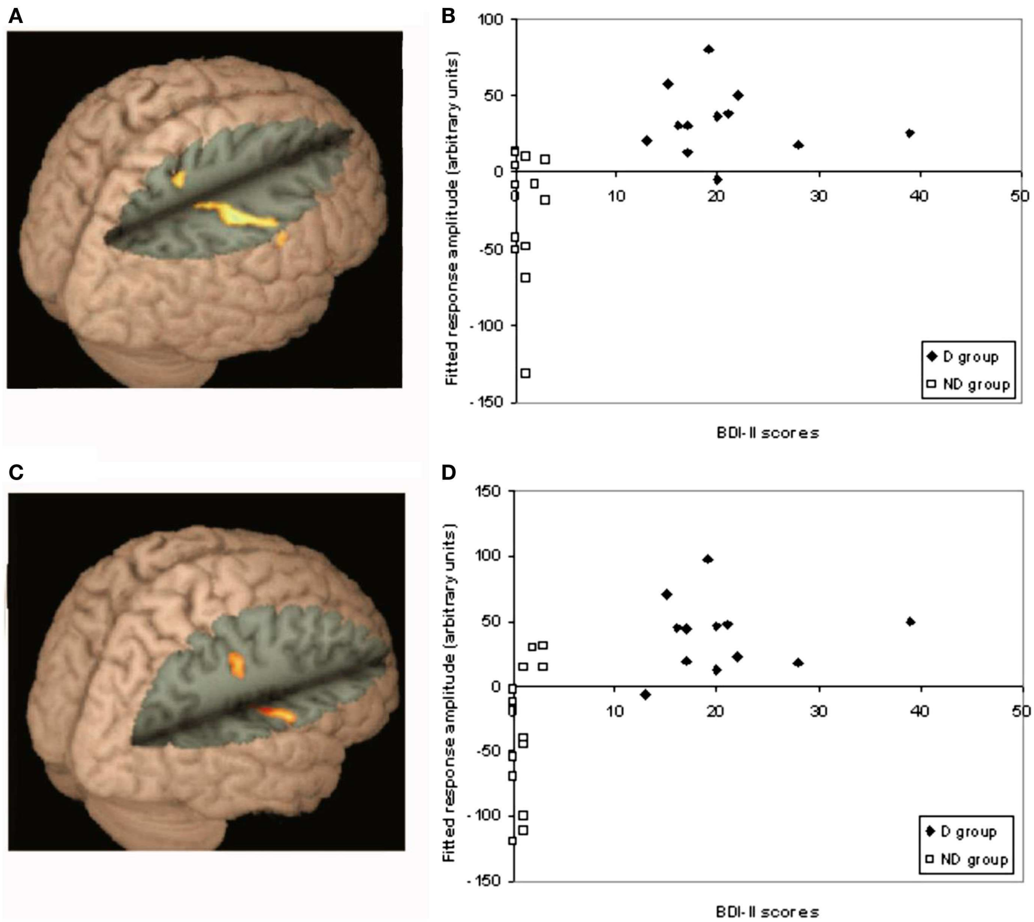


FIGURE 4 | Differences between D–ND groups for end-point and one-step queries. This figure shows the significantly different pattern of activation found in the parietal lobe/post-central gyrus for end-point and one-step queries between groups (D–ND): **(A)** for end-point queries, foci were found in superior parietal cortex (26, −46, 60, $z = 3.72$), supramarginal gyrus/post-central gyrus (44, −26, 40, $z = 3.76$); **(B)** shows how activation in each group during end-point queries varied according to BDI-II score; **(C)** for one-step queries, foci were found in the parietal lobe (post-central gyrus 60, −12, 20, $z = 3.95$; 58, −1, 46, $z = 3.87$); **(D)** shows how activation in each group during one-step queries varied according to BDI-II score.

DISCUSSION

Our results indicate that linear-order reasoning is an effective strategy when learning about, and reasoning with, naturalistic orders in humans. Moreover, hippocampal, parietal, and prefrontal cortical activations during the task provide corroborative evidence for a network of regions associated with reasoning found in previous studies (Christoff et al., 2001; Acuna et al., 2002; Knauff et al., 2002; Goel and Dolan, 2004; Schubotz et al., 2004; Fangmeier et al., 2006; Greene et al., 2006; Wendelken et al., 2008). In accordance with our hypotheses, greater activation was shown by the mildly depressed group compared to the ND group in spatial areas supporting transitive reasoning, namely the parietal cortex, during the spatial-like operations of solving the reasoning queries. This may be a compensatory mechanism in order to reach the same behavioral performance as the ND group [see Thomas and Elliott (2009), Brzezicka (2013)], or evidence for a different reasoning strategy in the depressed group. In *post hoc* analyses, corresponding differences in parietal activation between the two groups were also found for the learning phase.

DEPRESSED GROUP SHOW RELATIVELY GREATER PARIETAL ACTIVATION DURING REASONING

When solving the test queries, the depressed group showed relatively greater activation in the superior parietal lobe and in the region of the supramarginal gyrus and post-central gyrus compared to the ND group who showed relative deactivation during the task (relative to baseline and the depressed group). Activation in the somatosensory cortices (post-central gyrus) is assumed to reflect movement or non-task-related sensory feedback from pressing the response button, in line with suggestions by Acuna et al. (2002).

The greater activation in the parietal cortex during the test phase in the depressed group may be more task-related. The parietal lobe has been shown to be involved during mental operations that require spatial manipulation of internal representations, such as transitive inference (Goel and Dolan, 2001; Acuna et al., 2002; Knauff et al., 2002; Monti et al., 2007). Recently, Waechter et al. (2012) showed that patients with focal lesions in the parietal cortex were significantly impaired on transitive reasoning tasks, as

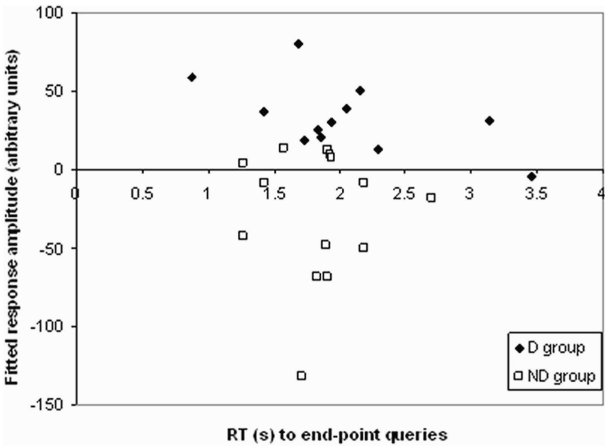


FIGURE 5 | Correlation in D group only between RT to end-point queries and activation in regions in the D–ND contrast. A scattergram plotting the reaction time during end-point queries against activation in regions in the D–ND contrast in both groups. A significant negative correlation between response time, after checking for outliers (none were found) using the Tukey criterion (Clark-Carter, 2004, Chapter 9), and activation during end-point trials was found in the D group ($r = -0.579$, $p = 0.048$; filled diamonds), but not ND group ($r = -0.009$, $p = 0.975$; clear squares).

compared to normal controls. It appears that the depressed group required more activation than the ND group at test to make the spatial aspects of the task sufficiently salient to arrive at the same behavioral outcome. It should be noted, however, that by using the contrast with fixation to examine the between group differences, this interpretation is not the only one possible. Greater parietal lobe activation between depressed and ND in the particular contrast D(test-fixation) − ND(test-fixation) could either reflect greater parietal lobe activation during the task (as stated), but alternatively could reflect no change in task activation but a greater deactivation during fixation in the depressed relative to the ND. Future research should further examine these possibilities.

The longer the time the depressed group took to respond to the test queries, the less activation was found in the parietal cortex. In other words, the quicker the depressed individuals responded, the more effort was indicated by brain activation. Given that this correlation is based only on correct responses, it appears that depressed participants needed to spend more effort to achieve quicker, correct responses, a correlation not found in the ND participants. These results are in accordance with earlier behavioral findings of Sedek and Von Hecker (2004). These authors suggested that depressed individuals are not as successful or efficient in constructing a linear order during the learning stage, and so engage in a different, compensatory style of reasoning when prompted by a test query. By compensation we mean that the same region in the brain may have to work harder in the depressed group than in the ND control group, in order to achieve the same performance level. This may be expected if depression is associated with more difficulties in the early deployment of suitable strategies of task execution and information integration (Hertel and Rude, 1991; Sedek and Von Hecker, 2004).

Our argument follows the general logic that processing disadvantages can be indicated by the observation that in order to achieve the same level of performance in a cognitive task, the disadvantaged group (in our case, depressed individuals) has to exert relatively more mental effort than the non-disadvantaged group (ND individuals). As such, this reasoning has previously been applied to other domains within the literature on behavioral correlates of cortical activation. For example, Fangmeier et al. (2006) (Ruff et al., 2003) suggested that for individuals with high spatial ability, the reasoning problems may have required less demand for visuospatial processing such that less activity in the parietal cortex was required to solve the problems, as compared to individuals with low spatial ability. In our case, the relative deactivation shown in the ND group in this study may take this argument one step further. A number of explanations for decreases in the BOLD signal have been put forward, including suppression of task irrelevant activity or reallocation of resources [e.g., McKiernan et al. (2003), Tomasi et al. (2006)], the default mode network (Raichle et al., 2001; Singh and Fawcett, 2008), greater activity in the baseline task than the task of interest (Gusnard et al., 2001; Stark and Squire, 2001), or optimizing activity to focus task performance (Astur and Constable, 2004; Rekkas et al., 2005). It is possible that the deactivation seen in the ND group could be explained as optimization of the activity in the parietal cortex, along the lines of that suggested for hippocampal deactivation during a similar relational task (Astur and Constable, 2004), in which it was suggested that inhibition was used to dampen irrelevant relations while the representation of important relations remained. This would be in line with the behavioral data, which suggests that retrieval of the correct response is made easier through the use of an organized mental array [see also Leth-Steensen and Marley (2000), Sedek and Von Hecker (2004)]. It is possible that the ND group, after successful construction of a mental array, tend to inhibit any additional (i.e., unnecessary) spatial processing that could interfere with retrieval from the already existing representation.

The fact that in the *post hoc* analyses, the depressed, unlike the ND, group displayed significant activation levels in the target parietal region during the learning phase may be due to the characteristics of the assumed process of mental model construction. As argued earlier (Sedek and Von Hecker, 2004), depressed individuals may find such construction more difficult to do than ND individuals. If it is further assumed that construction takes place in the learning phase, and that spatial functions are involved in this type of construction (Leth-Steensen and Marley, 2000), the more intense recruitment of parietal regions in the depressed group during learning appears plausible. It is further plausible to speculate that depressed individuals, more so than ND participants for whom construction would be easier (and already accomplished at the time of testing), would again recruit parietal regions more, even at test, in their attempts to arrive at clear mental models of the rankings[2].

[2] We thank the reviewer who drew our attention to the possibility that part of the reason why such parietal recruitment may be particularly required in depressed individuals may be the fact that the premises for model construction, i.e., the one-step pairs, are still rehearsed at test in the depressed, which would potentially entail ongoing constructive effort during test, and as such would interfere with their quick

While differential brain activity was found between groups during the test phase and the learning phase, the behavioral results, and the debriefing following scanning, did not show significant differences in performance between the depressed and the ND group, in contrast to earlier findings (Sedek and Von Hecker, 2004). The difference in the results between this study and these earlier findings could be due to differences in the paradigm arising from changes needed to prepare the task for fMRI; for example, participants were given extensive practice up to a criterion before being admitted to the task, unlike in Sedek and Von Hecker (2004), so the lack of performance differences may be due to a ceiling effect. Also, the timing in the fMRI task provided participants with a fixed study time of 10 s when learning the relations as opposed to response-driven timing, thereby providing more structure to the task, and possibly helping to focus attention. Indeed, Hertel and Rude (1991) showed that depressed participants exhibited performance deficits only in task conditions where their attention remained unfocused during task execution, but had normal performance when their attention was focused by task constraints.

This discrepancy between the group differences showing the neuroimaging results but not the behavioral data is not unprecedented. There is evidence to suggest that there are cognitive impairments in depression that are only demonstrable using neuroimaging techniques. Several studies have shown comparable performance on working memory and Stroop interference tasks in depressed and control participants, but in association with increased activation of the PFC in the depressed group (Wagner et al., 2006; Matsuo et al., 2007; Walter et al., 2007). Explanations for this differential brain activation include compensatory recruitment of PFC resources to complete the task successfully (Walter et al., 2007) and cortical inefficiency due to hyperactivity of key brain regions (Wagner et al., 2006). Smith et al. (2014) induced effect in a within-participant design by having participants view positive, negative, and neutral picture stimuli. They found that emotion did not impair logical reasoning, but that the neural systems underlying such reasoning differed in activation from those in the neutral condition. This dovetails with our finding that equivalent levels of reasoning between depressed and ND participants were associated with different activation levels in brain areas known as underlying performance in the particular task.

GREATER PREFRONTAL ACTIVATION DURING INFERENCE

Several studies now suggest that the rostral PFC is important for integration of relations into an internal representation (Christoff et al., 2001; Kroger et al., 2002; Fangmeier et al., 2006; Van Opstal et al., 2008; Wendelken et al., 2008). The results from the ND group in this study clarify this further by suggesting that rostral medial PFC (BA 8 and 10) activity is required when making novel inferences by manipulating information within an integrated mental model compared to recalling the answer to queries on previously presented relations. While some studies have found lateral RPFC activity to be associated with relational integration (Christoff et al., 2001; Wendelken et al., 2008), others have found medial RPFC activation, including the present one (Fangmeier et al., 2006; Van Opstal et al., 2008). In a review of models into the functions of the anterior PFC (BA 10), Ramnani and Owen (2004) suggest that the role of this region overall is "in integrating outcomes of two or more separate cognitive operations in the pursuit of a higher behavioral goal" (p. 1). The exact location of the activation found could be a function of the particular task employed, the specific cognitive processes required, sample recruited, stimuli used, and so on.

LIMITATIONS AND CONCLUSION

These results should be considered in light of the limitations of the study. The study was designed to compare directly activation between test queries or the learning phase, as well as between groups, as such a fixation baseline was deemed adequate. More specific findings relating to the learning phase, in particular, may have been possible with a baseline that provided greater control over the non-reasoning task processes, such as reading or making a response. Also we were unable to differentiate between activation associated with maintaining the structure of the array (ABCD) when presented in correct order type (i), as compared to the shuffled order type (ii) which should pose greater integration demands. These cognitive demands appeared not to require differential brain activity within this design. However, this investigation may have been improved if the design had allowed a greater number of examples of each type.

In conclusion, we have shown that reasoning with naturalistic linear orders in humans is subserved by a similar network of brain regions, including hippocampus, parietal, and prefrontal cortices, as compared to reasoning with purely abstract information found in previous studies. As predicted, sub-clinically depressed participants demonstrated higher activation of parietal areas during a test, and the learning, of presented and inferred relations, possibly reflecting a different strategy of task execution.

ACKNOWLEDGMENTS

This work was supported by a grant from the Economic and Social Research Council (RES-000-22-1788). We thank the volunteers who took part in this study. We also thank C. John Evans for help with imaging data collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at http://www.frontiersin.org/Journal/10.3389/fnhum.2014.01061/abstract

and efficient use of the mental model as a retrieval device. We agree. This possibility is in line with earlier research showing that in non-depressed individuals, premises of transitive mental models tend to be forgotten after successful construction (Mayberry et al., 1986), and that sad and depressed individuals tend to process detail information meticulously, i.e., preserve behavioral information more than individuals in neutral mood, when inferences from that information can be drawn (Gannon et al., 1994; Yost and Weary, 1996).

REFERENCES

Acuna, B. D., Eliassen, J. C., Donoghue, J. P., and Sanes, J. N. (2002). Frontal and parietal lobe activation during transitive inference in humans. *Cereb. Cortex* 12, 1312–1321. doi:10.1093/cercor/12.12.1312

Astur, R. S., and Constable, R. T. (2004). Hippocampal dampening during a relational memory task. *Behav. Neurosci.* 118, 667–675. doi:10.1037/0735-7044.118.4.667

Beck, A. T., Steer, R. A., Ball, R., and Ranieri, W. (1996). Comparison of beck depression inventories-IA and -II in psychiatric outpatients. *J. Pers. Assess.* 67, 588–597. doi:10.1207/s15327752jpa6703_13

Beckmann, C., Jenkinson, M., and Smith, S. (2003). General multi-level linear modelling for group analysis in fMRI. *Neuroimage* 20, 1052–1063. doi:10.1016/S1053-8119(03)00435-X

Brzezicka, A. (2013). Integrative deficits in depression and in negative mood states as a result of fronto-parietal network dysfunctions. *Acta Neurobiol. Exp.* 73, 313–325.

Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J. K., Holyoak, K. J., et al. (2001). Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *Neuroimage* 14, 1136–1149. doi:10.1006/nimg.2001.0922

Clark-Carter, D. (2004). *Quantitative Psychological Research: A Student's Handbook.* New York, NY: Psychology Press.

D'Esposito, M., Aguirre, G. K., Zarahn, E., Ballard, D., Shin, R. K., and Lease, J. (1998). Functional MRI studies of spatial and nonspatial working memory. *Brain Res. Cogn. Brain Res.* 7, 1–13. doi:10.1016/S0926-6410(98)00004-4

Dickins, D. W. (2005). On aims and methods in the neuroimaging of derived relations. *J. Exp. Anal. Behav.* 84, 453–483. doi:10.1901/jeab.2005.92-04

Fangmeier, T., Knauff, M., Ruff, C. C., and Sloutsky, V. (2006). fMRI evidence for a three-stage model of deductive reasoning. *J. Cogn. Neurosci.* 18, 320–334. doi:10.1162/jocn.2006.18.3.320

Gannon, K. M., Skowronski, J. J., and Betz, A. L. (1994). Depressive diligence in social information processing: implications for order effects in impressions and for social memory. *Soc. Cogn.* 12, 263–280. doi:10.1521/soco.1994.12.4.263

Goel, V., and Dolan, R. J. (2001). Functional neuroanatomy of three-term relational reasoning. *Neuropsychologia* 39, 901–909. doi:10.1016/S0028-3932(01)00024-0

Goel, V., and Dolan, R. J. (2003). Reciprocal neural response within lateral and ventral medial prefrontal cortex during hot and cold reasoning. *Neuroimage* 20, 2314–2321. doi:10.1016/j.neuroimage.2003.07.027

Goel, V., and Dolan, R. J. (2004). Differential involvement of left prefrontal cortex in inductive and deductive reasoning. *Cognition* 93, B109–B121. doi:10.1016/j.cognition.2004.03.001

Gotlib, I. H., and Hammen, C. (1992). *Psychological Aspects of Depression: Towards an Interpersonal Integration.* New York, NY: Wiley.

Greene, A. J., Gross, W. L., Elsinger, C. L., and Rao, S. M. (2006). An fMRI analysis of the human hippocampus: inference, context, and task awareness. *J. Cogn. Neurosci.* 18, 1156–1173. doi:10.1162/jocn.2006.18.7.1156

Gusnard, D. A., Raichle, M. E., and Raichle, M. E. (2001). Searching for a baseline: functional imaging and the resting human brain. *Nat. Rev. Neurosci.* 2, 685–694. doi:10.1038/35094500

Hertel, P. T., and Rude, S. S. (1991). Depressive deficits in memory: focusing attention improves subsequent recall. *J. Exp. Psychol. Gen.* 120, 301–309. doi:10.1037/0096-3445.120.3.301

Hogue, M., Beaugrand, J., and Lague, P. (1996). Coherent use of information by hens observing their former dominant defeating or being defeated by a stranger. *Behav. Processes* 38, 241–252. doi:10.1016/S0376-6357(96)00035-6

Hugdahl, K., Rund, B. R., Lund, A., Asbjørnsen, A., Egeland, J., Ersland, L., et al. (2004). Brain activation measured with fMRI during a mental arithmetic task in schizophrenia and major depression. *Am. J. Psychiatry* 161, 286–293. doi:10.1176/appi.ajp.161.2.286

Huttenlocher, J. (1968). Constructing spatial images: a strategy in reasoning. *Psychol. Rev.* 75, 550–560. doi:10.1037/h0026748

Jenkinson, M., Bannister, P., Brady, J., and Smith, S. (2002). Improved optimisation for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17, 825–841. doi:10.1006/nimg.2002.1132

Jenkinson, M., and Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Med. Image Anal.* 5, 143–156. doi:10.1016/S1361-8415(01)00036-6

Knauff, M., Mulack, T., Kassubek, J., Salih, H. R., and Greenlee, M. W. (2002). Spatial imagery in deductive reasoning: a functional MRI study. *Brain Res. Cogn. Brain Res.* 13, 203–212. doi:10.1016/S0926-6410(01)00116-1

Kroger, J. K., Sabb, F. W., Fales, C. L., Bookheimer, S. Y., Cohen, M. S., and Holyoak, K. J. (2002). Recruitment of anterior dorsolateral prefrontal cortex in human reasoning: a parametric study of relational complexity. *Cereb. Cortex* 12, 477–485. doi:10.1093/cercor/12.5.477

Leth-Steensen, C., and Marley, A. A. J. (2000). A model of response time effects in symbolic comparison. *Psychol. Rev.* 107, 62–100. doi:10.1037/0033-295X.107.1.162

Manns, J. R., and Eichenbaum, H. (2006). Evolution of declarative memory. *Hippocampus* 16, 795–808. doi:10.1002/hipo.20205

Marx, E. M., Williams, J. M. G., and Claridge, G. C. (1992). Depression and social problem solving. *J. Abnorm. Psychol.* 101, 78–86. doi:10.1037/0021-843X.101.1.78

Matsuo, K., Glahn, D. C., Peluso, M. A., Hatch, J. P., Monkul, E. S., Najt, P., et al. (2007). Prefrontal hyperactivation during working memory task in untreated individuals with major depressive disorder. *Mol. Psychiatry* 12, 158–166. doi:10.1038/sj.mp.4001894

Mayberry, M. T., Bain, J. D., and Halford, G. S. (1986). Information processing demands of transitive inference. *J. Exp. Psychol. Learn. Mem. Cogn.* 12, 600–613.

McKiernan, K. A., Kaufman, J. N., Kucera-Thompson, J., and Binder, J. R. (2003). A parametric manipulation of factors affecting task-induced deactivation in functional neuroimaging. *J. Cogn. Neurosci.* 15, 394–408. doi:10.1162/089892903321593117

Monti, M. M., Osherson, D. N., Martinez, M. J., and Parsons, L. M. (2007). Functional neuroanatomy of deductive inference: a language-independent distributed network. *Neuroimage* 37, 1005–1016. doi:10.1016/j.neuroimage.2007.04.069

Nolen-Hoeksema, S. (2002). "Gender differences in depression," in *Handbook of Depression*, eds I. Gotlib and C. Hammen (New York, NY: Guilford Press), 492–509.

Paz-Y-Miño, C. G., Bond, A. B., Kamil, A. C., and Balda, R. P. (2004). Pinyon jays use transitive inference to predict social dominance. *Nature* 430, 778–781. doi:10.1038/nature02723

Pinel, P., Dehaene, S., Riviere, D., and LeBihan, D. (2001). Modulation of parietal activation by semantic distance in a number comparison task. *Neuroimage* 14, 1013–1026. doi:10.1006/nimg.2001.0913

Potts, G. R. (1972). Information processing strategies used in the encoding of linear orderings. *J. Verbal. Learn. Verbal. Behav.* 11, 727–740. doi:10.1016/S0022-5371(72)80007-0

Radloff, L. S. (1977). The CES-D scale: a self-report depression scale for research in the general population. *Appl. Psychol. Meas.* 1, 385–401. doi:10.1177/014662167700100306

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., and Shulman, G. L. (2001). A default mode of brain function. *Proc. Natl. Acad. Sci. U.S.A.* 98, 676–682. doi:10.1073/pnas.98.2.676

Ramnani, N., and Owen, A. M. (2004). Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. *Nat. Rev. Neurosci.* 5, 184–194. doi:10.1038/nrn1343

Rekkas, P. V., Westerveld, M., Skudlarski, P., Zumer, J., Pugh, K., Spencer, D. D., et al. (2005). Neural correlates of temporal-order judgments versus those of spatial-location: deactivation of hippocampus may facilitate spatial performance. *Brain Cogn.* 59, 103–113. doi:10.1016/j.bandc.2005.05.013

Ruff, C. C., Knauff, M., and Fangmeier, T. (2003). "A neuro-cognitive account of individual differences in reasoing," in *The Cognitive Neuroscience of Individual Differences*, eds I. Reidvar, M. Greenlee, and M. Hermann (Oldenburg: Bibliotheks- und Informationssystem der Universität Oldenburg). p. 157–176.

Sakai, K., Hikosaka, O., Miyauchi, S., Takino, R., Sasaki, Y., and Putz, B. (1998). Transition of brain activation from frontal to parietal areas in visuomotor sequence learning. *J. Neurosci.* 18, 1827–1840.

Schubotz, R. I., Sakreida, K., Tittgemeyer, M., and von Cramon, D. Y. (2004). Motor areas beyond motor performance: deficits in serial prediction following ventrolateral premotor lesions. *Neuropsychology* 18, 638–645. doi:10.1037/0894-4105.18.4.638

Sedek, G., and Von Hecker, U. (2004). Effects of subclinical depression and aging on generative reasoning about linear orders: same or different processing limitations? *J. Exp. Psychol. Gen.* 133, 237–260. doi:10.1037/0096-3445.133.2.237

Simons, J. S., and Spiers, H. J. (2003). Prefrontal and medial temporal lobe interactions in long-term memory. *Nat. Rev. Neurosci.* 4, 637–648. doi:10.1038/nrn1178

Singh, K. D., and Fawcett, I. P. (2008). Transient and linearly graded deactivation of the human default-mode network by a visual detection task. *Neuroimage* 41, 100–112. doi:10.1016/j.neuroimage.2008.01.051

Smith, K. W., Vartanian, O., and Goel, V. (2014). Dissociable neural systems underwrite logical reasoning in the context of induced emotions with positive and negative valence. *Front. Hum. Neurosci.* 8:736. doi:10.3389/fnhum.2014.00736

Smith, S. (2002). Fast robust automated brain extraction. *Hum. Brain Mapp.* 17, 143–155. doi:10.1002/hbm.10062

Stark, C. E., and Squire, L. R. (2001). When zero is not zero: the problem of ambiguous baseline conditions in fMRI. *Proc. Natl. Acad. Sci. U.S.A.* 98, 12760–12766. doi:10.1073/pnas.221462998

Sternberg, R. J. (1980). Representation and process in linear syllogistic reasoning. *J. Exp. Psychol. Gen.* 109, 119–159. doi:10.1037/0096-3445.109.2.119

Thomas, E. J., and Elliott, R. (2009). Brain imaging correlates of cognitive impairment in depression. *Front. Hum. Neurosci.* 3:30. doi:10.3389/neuro.09.030.2009

Tomasi, D., Ernst, T., Caparelli, E. C., and Chang, L. (2006). Common deactivation patterns during working memory and visual attention tasks: an intra-subject fMRI study at 4 Tesla. *Hum. Brain Mapp.* 27, 694–705. doi:10.1002/hbm.20211

Turner, M., and Engle, R. (1989). Is working memory capacity task dependent? *J. Mem. Lang.* 28, 127–154. doi:10.1016/0749-596X(89)90040-5

Van Opstal, F., Verguts, T., Orban, G. A., and Fias, W. (2008). A hippocampal-parietal network for learning an ordered sequence. *Neuroimage* 40, 333–341. doi:10.1016/j.neuroimage.2007.11.027

Vasconcelos, M. (2008). Transitive inference in non-human animals: an empirical and theoretical analysis. *Behav. Processes* 78, 313–334. doi:10.1016/j.beproc.2008.02.017

von Hecker, U., and Sedek, G. (1999). Uncontrollability, depression, and the construction of mental models. *J. Pers. Soc. Psychol.* 77, 833–850. doi:10.1037/0022-3514.77.4.833

Waechter, R. L., Goel, V., Raymont, V., Kruger, F., and Grafman, J. (2012). Transitive inference reasoning is impaired by focal lesions in parietal cortex rather than rostrolateral prefrontal cortex. *Neuropsychologia* 51, 464–471. doi:10.1016/j.neuropsychologia.2012.11.026

Wagner, G., Sinsel, E., Sobanski, T., Kohler, S., Marinou, V., Mentzel, H. J., et al. (2006). Cortical inefficiency in patients with unipolar depression: an event-related fMRI study with the Stroop task. *Biol. Psychiatry* 59, 958–965. doi:10.1016/j.biopsych.2005.10.025

Walter, H., Wolf, R. C., Spitzer, M., and Vasic, N. (2007). Increased left prefrontal activation in patients with unipolar depression: an event-related, parametric, performance-controlled fMRI study. *J. Affect. Disord.* 101, 175–185. doi:10.1016/j.jad.2006.11.017

Waltz, J., Knowlton, B., Holyoak, K., Boone, K., Mishkin, F., de Menezes Santos, M., et al. (1999). A system for relational reasoning in human prefrontal cortex. *Psychol. Sci.* 10, 119–125. doi:10.1111/1467-9280.00118

Wechsler, D. (1981). *Wehsler Adult Inelligence Scale-Revised.* San Antonio, TX: The Psychological Corporation.

Wendelken, C., Nakhabenko, D., Donohue, S. E., Carter, C. S., and Bunge, S. A. (2008). "Brain is to thought as stomach is to ??": investigating the role of rostrolateral prefrontal cortex in relational reasoning. *J. Cogn. Neurosci.* 20, 682–693. doi:10.1162/jocn.2008.20055

Woolrich, M., Behrens, T., Beckmann, C., Jenkinson, M., and Smith, S. (2004). Multi-level linear modelling for fMRI group analysis using Bayesian inference. *Neuroimage* 21, 1732–1747. doi:10.1016/j.neuroimage.2003.12.023

Woolrich, M., Ripley, B., Brady, J., and Smith, S. (2001). Temporal autocorrelation in univariate linear modelling of fMRI data. *Neuroimage* 14, 1370–1386. doi:10.1006/nimg.2001.0931

Worsley, K., Evans, A., Marrett, S., and Neelin, P. (1992). A three-dimensional statistical analysis for CBF activation studies in human brain. *J. Cereb. Blood Flow Metab.* 12, 900–918. doi:10.1038/jcbfm.1992.127

Yost, J. H., and Weary, G. (1996). Depression and the correspondent inference bias: evidence for more effortful cognitive processing. *Pers. Soc. Psychol. Bull.* 22, 192–200. doi:10.1177/0146167296222008

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: 09 October 2014; accepted: 19 December 2014; published online: 19 January 2015.

*Citation: Hinton EC, Wise RG, Singh KD and von Hecker U (2015) Reasoning with linear orders: differential parietal cortex activation in sub-clinical depression. An fMRI investigation in sub-clinical depression and controls. Front. Hum. Neurosci. **8**:1061. doi: 10.3389/fnhum.2014.01061*

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

frontiers in
HUMAN NEUROSCIENCE

REVIEW ARTICLE
published: 27 February 2015
doi: 10.3389/fnhum.2015.00101

Imaging deductive reasoning and the new paradigm

Mike Oaksford *

Department of Psychological Sciences, Birkbeck College, University of London, London, UK

Edited by:
Jérôme Prado, Centre National de la Recherche Scientifique, France

Reviewed by:
Oshin Vartanian, Defence Research and Development Canada; Toronto Research Centre, Canada
Kinga Morsanyi, Queen's University Belfast, UK

***Correspondence:**
Mike Oaksford, Department of Psychological Sciences, Birkbeck College, University of London, Malet Street, London, WC1E 7HX, UK
e-mail: mike.oaksford@bbk.ac.uk

There has been a great expansion of research into human reasoning at all of Marr's explanatory levels. There is a tendency for this work to progress within a level largely ignoring the others which can lead to slippage between levels (Chater et al., 2003). It is argued that recent brain imaging research on deductive reasoning—implementational level—has largely ignored the new paradigm in reasoning—computational level (Over, 2009). Consequently, recent imaging results are reviewed with the focus on how they relate to the new paradigm. The imaging results are drawn primarily from a recent meta-analysis by Prado et al. (2011) but further imaging results are also reviewed where relevant. Three main observations are made. First, the main function of the core brain region identified is most likely elaborative, defeasible reasoning not deductive reasoning. Second, the subtraction methodology and the meta-analytic approach may remove all traces of content specific System 1 processes thought to underpin much human reasoning. Third, interpreting the function of the brain regions activated by a task depends on theories of the function that a task engages. When there are multiple interpretations of that function, interpreting what an active brain region is doing is not clear cut. It is concluded that there is a need to more tightly connect brain activation to function, which could be achieved using formalized computational level models and a parametric variation approach.

Keywords: Marr's levels, Bayesian inference, brain imaging, new paradigm

This paper presents a focused review of the brain imaging results on deductive reasoning. The focus is given by the new paradigm in reasoning (Over, 2009; also see Elqayam and Over, 2013, which is an introduction to a special issue in the new paradigm), which is based on Bayesian probability and dual processes. This new paradigm offers an alternative theoretical framework to those typically assumed in imaging research on deductive reasoning. In providing such a review, it is fortuitous that there has been a recent detailed meta-analysis of this area (Prado et al., 2011). I therefore concentrate on the findings of this meta-analysis, bringing in other relevant imaging results as they bear on the line of argument.

I first discuss why we might expect slippage between different levels of explanation in reasoning research in terms of Marr's levels. Brain imaging is concerned with the implementational level whereas the new paradigm is a computational level theory. I then summarize the results of Prado et al.'s (2011) meta-analysis of 28 imaging studies. I then introduce the new paradigm and trace the consequences of its two critical features—(i) it is probabilistic and (ii) it invokes dual processes—for the interpretation of these brain imaging results. In doing so, I make several proposals. First, the main function of the core brain region identified by Prado et al. (2011) is most likely elaborative, defeasible reasoning not deductive reasoning. Second, the subtraction methodology and the meta-analytic approach may remove all traces of content specific System 1 processes thought by many to underpin much if not most human reasoning. Third, interpreting the function of brain regions activated by a task depends on our theories of the function that a task engages. When there are multiple interpretations of that function, interpreting what an active brain region is doing is not clear cut. Moreover, this issue is not resolvable at the implementational level. I conclude that imaging research may need to catch up with the computational level where there has been much recent progress.

COMPUTATIONAL LEVELS

The multilevel nature of computational explanation in the cognitive sciences leads to multiple research strategies for investigating the cognitive processes that underlie any human behavior. At Marr's (1982) computational level, the function that the mind/brain is believed to be computing in the performance of some task is specified. At the algorithmic level, the sequence of processing steps that compute this function is specified. At this level, various processing limitations need to be taken in to account, which may serve a critical explanatory role, e.g., working memory limitations. Finally, at the implementational level, the actual physical hardware in which the cognitive algorithm is instantiated in the brain is specified. At this level, the limitations of the physical components implementing the cognitive algorithm are taken into account, e.g., the time course of neural responses. As Marr envisaged these levels, addressing the computational level was the priority, i.e., the "function first" approach, because only this strategy was likely to prove successful. For example, little progress was made in understanding the operation of the heart until it was realized that its function was to circulate blood around the body. This multilevel nature of computational explanation means that researchers often pursue different research strategies that focus on only one level, usually determined by their own particular technical competences. This is usually unproblematic but it can create slippage between levels whereby research may proceed at different paces for a period of time, i.e., one level may move ahead while our understanding at the other levels lags behind (Chater et al., 2003).

In this paper, I argue that there has been slippage between the computational and implementational levels in the study human reasoning. Brain imaging research has largely appealed to theoretical frameworks at the computational level that over the last 20 years have been strongly challenged by the new probabilistic paradigm in human reasoning (Oaksford and Chater, 1994, 2001, 2007; Over, 2009; Elqayam and Over, 2013). In this paper, I examine what may be involved in re-aligning these levels of explanation in reasoning research.

IMAGING RESULTS: PRADO ET AL.'S (2011) META-ANALYSIS

In describing the existing research on the brain imaging of deductive reasoning, a good starting point is to briefly summarize Prado et al.'s (2011) meta-analysis. These studies initially presented a confusing set of results, which led (Goel, 2007, p. 440), to suggest that there may not be a unitary neural system for deductive reasoning, but rather "a fractionated system that is dynamically configured in response to certain task and environmental cues". Prado et al.'s (2011) meta-analysis seems to reveal more consistency amongst these studies. They appear to show a core, mainly left lateralized, system being active in deductive reasoning with other subsystems being recruited dependent on the nature of the task, be it propositional, categorical, or relational reasoning. The core system involved the left lateralized inferior frontal gyrus (IFG), middle frontal gyrus (MFG), precentral gyrus (PG), posterior parietal cortex (PPC), and the basal ganglia (BG); it also included one medial structure, the medial frontal gyrus (MeFG). Prado et al. (2011) interpret this finding as consistent with the "left brain interpreter" hypothesis (Roser and Gazzaniga, 2006). The left hemisphere is primarily engaged in interpreting incoming information and filling in the missing information via inferential processes. The primary involvement of left lateralized brain systems seems to run counter to some accounts of human reasoning that place special emphasis on visual-spatial representations and processes, i.e., mental models (Johnson-Laird, 1983), which are primarily right lateralized.

Additional systems seem to be recruited for specific deductive tasks. Propositional reasoning involves relations between propositions like *if the key is turned, the car starts*, *the key is turned*, therefore, *the car starts*. This is the classical propositional inference of modus ponens and it depends purely on the connectives (*if. . .then* here but also *and*, *or*, *not*) and not on any deeper analysis of the propositions involved. Relational and categorical reasoning rely on going deeper in to the subject/predicate structure of a proposition. Categorical reasoning involves categorical statements like *All artists are beekeepers*, where "artists" is the subject and "beekeepers" is the predicate. This mode of reasoning is typically investigated using two premise quantified syllogisms such as *All artist are beekeepers*, *Some artists are smokers*, therefore, *Some beekeepers are smokers*. Relational reasoning moves from unary predicates, involving one variable, to relations, usually only binary, e.g., *John is taller than Fred*. These are typically investigated using the transitive inference paradigm—*John is taller than Fred*, *Fred is taller than Jane*, *is Jane taller than John*?—and spatial reasoning, e.g., *John is to the left of Fred*, *Fred is to the right of Jane*, *is Jane to the right of John*?

Relational arguments activate bilateral PPC and right MFG. Bilateral activation of the PPC is commonly seen in studies of visuospatial tasks and the reliable activation of right PPC in relational arguments seems consistent with theories like mental models. Categorical arguments only show strong activation of left lateralized IFG and BG and this activation is more consistent than for relational or propositional reasoning. These regions seem to be most consistently associated with processing syntax and grammar (e.g., Goel et al., 2000; Ullman, 2006; Grodzinsky and Santi, 2008). Propositional arguments are also left lateralized and most strongly activate PPC, PG, and MeFG. PPC and MeFG have been associated with non-syntactic verbal processing and maintaining abstract rules in memory respectively (Bunge et al., 2003; Booth et al., 2007).

Prado et al. (2011) draw an important conclusion from the finding that there is no one neural system apparently involved in all three domains of deductive reasoning investigated in these studies. No theory that suggests that these different domains all rely on a unitary underlying cognitive process is likely to be able to explain these results. Only some types of reasoning, apparently relational reasoning, seem to invoke visuospatial processing, propositional and categorical reasoning do not. They suggest that this tends to rule out unitary theories like mental logic (e.g., Rips, 1994) and mental models (Johnson-Laird, 1983) which propose that either formal rules or visuospatial representations underlie all deductive reasoning. Indeed, mental models theory makes the broader claim that such unitary visuospatial representations underlie all reasoning, deductive or inductive.

In most of the studies in Prado et al.'s (2011) meta-analysis, the theoretical rationale was to compare just two computational and implementational level theories of human reasoning. At the computational level, both mental models and mental logic theories take standard binary truth functional logic as defining the function the cognitive system is trying to compute.[1] They diverge only on the nature of the representations and processes that implement this logic in the human mind i.e., they disagree primarily at the algorithmic level. Framing these investigations

[1] This can be disputed (Schroyens, 2010). It is possible that mental models has introduced slippage between the computational and algorithmic levels. That is, mental models has been making advances by proposing a particular representation/process pair which can mimic logic under certain circumstances but the actual full computational level theory of mental models, i.e., the actual logic it implements at the algorithmic level, remains to be defined. This is a coherent proposal and there may be candidate logics that might make good on this claim. However, I have never heard this argument put forward by any other mental models theorist.

as deciding between these two theories also suggests that investigating deductive reasoning means to only study reasoning which can be captured by standard logic. However, it is arguable that over the last 15–20 years the most notable progress in the study of human reasoning has been at the computational level where alternative probabilistic theories of what people are doing in deductive reasoning tasks have been proposed (Hahn, 2014). These probabilistic accounts have become known as the "new paradigm" (Over, 2009; Manktelow, 2012). I now trace the origins of the new paradigm and its consequences for the interpretation of neuroimaging data.

THE NEW PARADIGM

There are two strands to the new paradigm. First, it is probabilistic. Second, it is a dual process theory that invokes both System 1 and System 2 processes (Evans, 2010; Stanovich, 2011). System 1 is Kahneman's (2011) fast system and System 2 is his slow system. I look first at the probabilistic strand and its motivations and relate these directly to some of the results discussed in Prado et al. (2011).

PROBABILITIES

In motivating the probabilistic strand of the new paradigm, I begin with a quote from Dennett:

> "But it is obviously true that most people never engage in explicit non-enthymematic formal reasoning" (Dennett, 1998, p. 289).

Enthymematic reasoning, for example, Tweety is a bird therefore Tweety flies, explicitly involves the use of world knowledge in order to fill in information not explicitly stated, i.e., that *all birds fly*, *normally birds fly* or *the probability that birds fly is high*. We make these inferences automatically with little conscious thought. As Dennett's remark implies, this is the kind of inference that underpins our everyday lives and interactions with others. It also implies that the kind of "non-enthymematic formal" reasoning required in most of the reasoning tasks investigated in Prado et al. (2011) and in most deductive reasoning tasks used in the lab, are not commonly engaged in by the man or woman in the street. Consequently, attempting to derive a general theory of human reasoning by investigating these kinds of tasks is perhaps to step off on the wrong foot.

Concerns could be assuaged if this kind enthymematic reasoning could be captured by standard logic. However, one of the primary motivations for moving to probabilistic theories in the new paradigm has been the fact that enthymematic reasoning is defeasible (Oaksford and Chater, 1991, 2007). That is, learning that Tweety is an ostrich *defeats* the inference that Tweety can fly on learning that Tweety is a bird. We have rehearsed the problems of attempting to reconstruct such reasoning in standard logic many times before and do not do so again here (Oaksford and Chater, 1991, 1993, 1995, 2007). The probabilistic approach characterizes these inferences as being underpinned by probabilistic relations such as being a bird makes the probability that something flies high. That is, the world knowledge that underpins the enthymematic inference above is something like, *if x is a bird then x can fly*, where Pr(*if x is a bird then x can fly*) = Pr(*x can fly*|*x is a bird*) and this probability is high.

Another important aspect of this kind of reasoning, which Fodor (1983) calls non-demonstrative inference, is that it is the prototypical central cognitive process (Fodor, 1983; Oaksford and Chater, 1991). The contrast between modular and central cognitive processes is drawn along the lines of those that require large amounts of world knowledge and those that do not. Fodor (1983) argued that central cognitive processes are *Quinean*.[2] A process is Quinean when it apparently invokes the whole of our belief system. So the reason we draw the inference that Tweety can fly is that this is the most *plausible* inference to draw. But plausibility is only definable against the backdrop of everything else we know or believe. Moreover, any Bayesian probabilistic account is going to be Quinean. Our best bet about how we determine someone's subjective probability Pr(*x can fly*|*x is a bird*) is given by the Ramsey test. This test involves assuming Tweety is a bird, i.e., adding this proposition to our stock of beliefs while making minimum adjustments to our other beliefs, and reading off our new degree of belief that Tweety flies. This is a philosophical prescription but its implications for psychological processes are clear: defeasible reasoning, probabilistically construed or not, must invoke central cognitive processes.

Imaging, inference and central cognitive processes

This brief account of the underlying motivations for the probabilistic strand of the new paradigm (see also, Oaksford and Chater, 2007, Chapters 1–4) leads to two conclusions that appear to be supported by the imaging results discussed by Prado et al. (2011). First, Prado et al. (2011) identify their left lateralized core system with Gazzaniga's "left brain interpreter" hypothesis (Roser and Gazzaniga, 2006). *It is important to be clear on the nature of the inferences that underpin this hypothesis.* A main source of evidence for the left brain interpreter hypothesis is the *elaborative* inferences that some patients and normal participants make in interpreting pictures. These elaborative inferences seem to be responsible for false recognition of novel pictures as being previously viewed. Of course, our enthymematic inference that Tweety can fly is an elaborative inference of precisely this sort. It could only be construed deductively if the enthymematically provided premise was ***all** birds can fly* but then it would not be defeasible. But all elaborative and enthymematic inferences are defeasible and people may not even be aware of the fact that they have drawn one until it is overturned, e.g., on being told Tweety is an ostrich, and the mild sense of surprise that they then experience. In sum, if the left brain interpreter hypothesis is correct as an interpretation of the brain imaging results, then its primary function is probably not in deductive reasoning but rather elaborative, defeasible, and probabilistic reasoning. At least this is the kind of reasoning that has provided the principal evidence for the left brain interpreter hypothesis in the past.

[2] The philosopher, Willard Van Ormond Quine, famously commented that a belief can always be saved from refutation by making adjustments elsewhere in our belief system, i.e., the mechanisms of belief fixation and revision are holistic, depending on everything else that we know or believe (Quine, 1953).

Second, such defeasible, probabilistic reasoning, as we have just discussed, is perhaps our best candidate for a central cognitive process. That is, it is one of the processes that is least likely to be subserved by a unitary cognitive module. And this would appear to be exactly what the brain imaging data reveals, reasoning is not subserved by a unitary cognitive process, be it formal rules or visual spatial representations, in a single isolable module. It is also worth noting that, given the defeasible, probabilistic nature of the inferences that underpin the left brain interpreter hypothesis, when deployed in deductive tasks this brain system is probably not being used to perform functions for which it originally evolved. That is, at best, deductive reasoning is a limiting case of this system's primary function, for example, when the probabilities go to 0 or 1.

Deductive tasks

A possible objection to the line of argument in the last section is that the imaging results reviewed in Prado et al. (2011) specifically focused on deduction, i.e., the tasks were very specifically deductive tasks, which could not form the evidential basis for generalizing to defeasible non-demonstrative reasoning. However, in the reasoning literature mental models theory *has* taken these tasks to provide the basis for a wholly general theory of reasoning subsuming deduction (Johnson-Laird and Byrne, 1991), probabilistic inductive reasoning (Johnson-Laird et al., 1999), causal reasoning (Goldvarg and Johnson-Laird, 2001) and much else besides. Moreover, mental logic and mental models are the theoretical frameworks on which the imaging research has primarily concentrated. The new paradigm argues that because everyday, defeasible reasoning is the ubiquitous phenomena people apply sensible reasoning strategies for dealing with the everyday world to laboratory deductive reasoning tasks. This strategy can explain away many of the so called biases observed in human deductive reasoning (Oaksford and Chater, 2007).

Could it nonetheless be argued that the specific tasks used in the imaging studies review by Prado et al. (2011) are uniquely deductive and consequently they genuinely investigate just this very narrow domain of human reasoning? A point I elaborate on further below, is that we require a computational level theory to define the function that a task engages (*Functions, Tasks, and Active Regions*). In imaging research, "deduction" is taken to refer to binary truth functional logic as it is in mental logic and mental model theory. But there are a range of alternative logics especially for the conditional (see, e.g., Haack, 1975; Bennett, 2003) and there are well specified probabilistic accounts of categorical reasoning (Chater and Oaksford, 1999). Moreover, there are varieties of probability logic (Adams, 1998) in which coherent probability intervals are *deduced* from probability assignments to the premises (Pfeifer and Kleiter, 2010; Pfeifer, 2013). Such logics are just as deductive as binary truth functional logic.

Perhaps it could be argued that at least tasks like relational and spatial reasoning have deterministic binary logical solutions and as such are genuinely "deductive" tasks in the sense intended in mental logic and mental models theory. However, phenomena like perspectival relativity (Barwise and Perry, 1983) question this view. Take, for example, the premises *John is to the left of Fred*, *Fred is to the right of Jane* which is assumed to lead to the deterministic logical conclusion that *Jane is to the right of John*. If *Jane* and *John* are both facing each other with Fred in the middle facing neither then the question of whether *Jane is to the right of John* has no deterministic answer, they are neither to the left nor to the right of each other, despite the truth of the premises. Left and right depend on our subjective frame of reference in personal space. Another example is if *Fred* is standing at the North pole and *Jane* and *John* at the South pole. In this case, *Jane* and *John* would appear to be simultaneously to *Fred's* left and to his right. Such counterexamples suggest that there are certain orientations that make the conclusion more likely but it does not follow deterministically. Even relations like *taller*, which rely on being able to measure the world, may require a probabilistic theory. Measurement error suggests that our representations of items on a scale use distributions which may overlap. Such representations can explain the symbolic distance effect where for a long transitive chain, e.g., $a > b > c > d > e$ ("$>$" = is taller than), people find it harder to discriminate whether $c > d$ than $a > e$ (Cohen Kadosh et al., 2005). In summary, tasks are not deductive in and of themselves. What function a task engages is determined by the empirically most adequate computational level theory of that task.

Imaging: deduction vs. induction

We have argued that the core system identified by Prado et al. (2011) is concerned with defeasible, non-demonstrative reasoning. The new paradigm has been characterized as "imperialistic" (Rips, 2002) in that it attempts to assimilate deduction to probabilistic inductive reasoning. However, there is behavioral data suggesting that these processes dissociate (Rips, 2001; Heit and Rotello, 2010). Although recently Lassiter and Goodman (2015) have shown these differences may have more to do with the semantics of the terms used to elicit people's responses, i.e., is the conclusion "necessary" (deduction) or "plausible" (induction), than with fundamental differences in the reasoning process which remains probabilistic. A suggestion originally made by Oaksford and Hahn (2007). There is also imaging data relevant to this question.

Goel and Dolan (2004) found that some structures were more active in deduction (left IFG) than in induction and that some were more active in induction (primarily left MFG) than in deduction. They argue that their findings are more consistent with other studies, particularly lesion studies, than previous work apparently showing that these modes of reasoning were lateralized with induction associated with the left hemisphere and deduction with the right (Parsons and Osherson, 2001). Goel and Dolan's (2004) studies were included in Prado et al.'s (2011) meta-analysis and both these structures are part of the core system they identified. Goel and Dolan argue that left IFG is associated with Broca's area and hence language, working memory and perhaps syntactic processing. Left MFG activation, they hypothesize is associated with the recruitment of general knowledge required for induction.

Induction and deduction activate much the same brain system. Moreover, given the nature of these inferences even what differential patterns of activation there were are understandable.

The new paradigm does not deny that deduction and induction are distinct (Evans and Over, 2013). Deduction involves inferences over the syncategorematic or logical terms of a language (*if. . .then*, *and*, *or*, *not*, *all* etc.), i.e., the inference follows from the meaning of these terms. This is not the case for the inductive inferences that Rips (2001), Goel and Dolan (2004), and Heit and Rotello (2010) investigated which involved categorical induction. In deduction processing of the structure of premises is important but it is less so for the premises of an inductive inference which may simply present a string of facts (e.g., domestic cats have 32 teeth, lions have 32 teeth). Moreover, we learn about the world by observation in a similar way, i.e., inductive inferences do not have to be mediated by language in the way deductive inferences are. In probability logic, the meaning of the conditional is given by the conditional probability. The assertion of a conditional means that that the conditional probability is high. So while both inferences types are probabilistic, and both rely to a degree on world knowledge, there is an important structural difference between induction and deduction, which is what Goel and Dolan's result are presumably picking up. A final observation is that we can find no lesion study showing a full double dissociation between induction and deduction. Although Goel and Dolan cite one case study involving a single dissociation using a theory of mind task (Varley and Siegal, 2000), no classical deductive or inductive reasoning tasks were used.

DUAL PROCESSES

In the new paradigm, it is agreed that a dual process theory is required (Evans and Over, 2004; Evans, 2010; Oaksford and Chater, 2010, 2011; Stanovich, 2011). System 1 is implicit, probabilistic, and based on world knowledge. System 2 is explicit, involves working memory, and is based on "analytic" processes. These analytic processes have been argued to be either also probabilistic (Evans and Over, 2004; Oaksford and Chater, 2009, 2010, 2011; Evans, 2010; Pfeifer and Kleiter, 2010) or based on standard binary logic (Rips, 1994, 2001; Stanovich and West, 2000; Heit and Rotello, 2010; Klauer et al., 2010; Stanovich, 2011). Whatever view one takes, it is generally agreed that deductive reasoning behavior is a product of an interaction between both these systems.

Kahneman (2011) uses some instructive examples to illustrate the nature of System 1 and System 2 processes. To illustrate System 1, he simply presents the juxtaposition of two words:

Banana Vomit

As he observes, a whole panoply of responses are triggered automatically by this juxtaposition. A whole causal story is probably constructed connecting the ingestion of bananas and vomiting. Moreover, a mild sense of surprise is invoked by this unusual juxtaposition. Unpleasant visual and auditory images will also be briefly triggered. The processes that produce these reactions happen unconsciously and very rapidly, all we are aware of is a reaction. He illustrates System 2, by tasks like counting back in threes from say 1037. This task is effortful, fully conscious, difficult to keep going, and involves applying the rules of arithmetic. Tasks illustrating the interaction of these systems are those like the bat and ball problem. In this task participants are told that the bat costs a dollar more than the ball and that together they cost $1.10 and they are asked how much does the ball cost? A spontaneous System 1 response is ten cents, which must be wrong because this would make the total cost of the bat and ball $1.20. In such tasks, the automatic System 1 response may need to be overridden and the actual cost consciously calculated in System 2.

In deductive reasoning tasks, it may be that a spontaneous System 1 response needs to be overridden but it seems unlikely that lay participants are capable of then engaging the correct logical rules in System 2 as they can the rules arithmetic for the bat and ball problem. Except for the logically trained these rules are simply not consciously available (of course for the bat and ball problem to be solvable, the rules of arithmetic also had to be learned). Consequently in deductive reasoning performance, it is probably best not to consider System 2 as conscious. This seems consistent with recent work on *logical intuitions* which shows that people appear to unconsciously detect the conflict between the intuitive System 1 response and the correct response even if they make the apparently biased System 1 response (De Neys, 2012, 2014). What people will be conscious of is a response, initially triggered by System 1, accompanied by a *feeling of rightness* (Thompson et al., 2011). This feeling may well depend on how the intuitive System 1 response agrees or conflicts with the output of System 2.

A great deal of work in the new paradigm is on showing that apparently irrational performance on many tasks is actually rational from a probabilistic perspective. Moreover, much of this behavior is hypothesized to be the responsibility of System 1. Kahneman's illustrative example of System 1 in action suggests that much of the information required by a rational theory of inference and decision is automatically computed at this level. For example, to understand the juxtaposition of just these two words people seem to generate a causal model relating the ingestion of bananas to vomiting. Moreover, a surprising event is one that is improbable, which suggests that relevant probabilities are automatically computed. Furthermore, people have a spontaneous emotional reaction to this juxtaposition expressing relevant hedonic or experienced utilities. The almost immediate availability of all this information may suggest that System 1 is indeed capable of some complex inferential processes, consistent with logical intuitions (De Neys, 2012, 2014).

Recently, it has been suggested that System 1 uses this information in inference in a similar way to the unconscious inferences involved in perception and action hypothesized by Helmholtz (Oaksford, 2014, Submitted). Again most progress on unconscious inference is being made at the computational level by computational biologists. These unconscious inferential processes are being understood in probabilistic terms in the Bayesian brain hypothesis (Dayan and Hinton, 1996; Friston, 2005, 2008; Clark, 2013). In brief, perception is viewed as the process of using alternative generative models of the current context to generate hypotheses about the causes of the pertubations of our sensory surfaces. These hypotheses, e.g., it is a dog or it is a cat, are at the top level of a hierarchical Bayesian model and these cascade down making lower level predictions ultimately for the responses of center surround units in our sensory receptors. Prediction errors, e.g., the hypothesis says the unit should be on when it is off, are then fed back up the hierarchy minimizing expected surprise or entropy concerning the cause of the proximal stimulus, i.e., the least surprising interpretation is adopted. It has also been shown how these cascaded inferential processes can be implemented in cortex.

In sum, most reasoning is largely unconscious, it occurs automatically based on the rich information generated by System 1 which also seems directly implicated in unconscious inferences in perception and action. Our theories of System 1 in the psychology of explicit verbal reasoning and our theories of unconscious inference in perception and action also converge on a Bayesian account.[3] This means that content, which fixes the relevant probabilities, is central to the reasoning process. But most imaging studies have framed their investigations in term of mental logic and mental models in which content is largely irrelevant. As I now argue, this fact may have important consequences for the interpretation of imaging results in the psychology of verbal reasoning.

Imaging system 1

Most brain imaging studies use the subtraction methodology to isolate brain regions that are specific to deduction and this usually involves contrasting materials with relevant content. So for example, in Goel and Dolan (2003) experiments on belief bias in categorical reasoning, materials like:

(A) No reptiles can grow hair
Some elephants can grow hair

So, No elephants are reptiles (true conclusion, invalid inference) were contrasted with a baseline:

(B) No reptiles can grow hair
Some elephants can grow hair
No fried foods have cholesterol

Subtracting out activation due to this baseline may remove any traces of the automatically activated content based processes like those involved in Kahneman's System 1 example. These processes are automatically activated by the content of the words which are also present in the contrast. But if most of the inferential action is at the System 1 level this means that the subtraction methodology may be removing most activations of interest (see also, Monti and Osherson, 2012, for a similar line of argument). Other contrasts that have been used, e.g., a simple fixation location, may seem to avoid this problem. However, even if such contrasts retained activations associated with content, the goal of Prado et al.'s (2011) meta-analysis was to detect active regions *across* studies. Consequently, these content based activations will be removed in the meta-analysis because content varied between studies (and indeed between tasks).

Content-based System 1 activations may be subject to a great deal of variation not only across studies but also across individuals. Would one expect, for example, there to be much spatial overlap between two people's representations of the concept "horse"? When one thinks of horses, regions associated with their shapes, movements, smells, and locations where they have been encountered are activated and binding these disparate responses together is the crucial step in having the concept "horse". Given what is likely to be a diffuse pattern of activation, presumably involving different sensory centers and memories, it seems unlikely that there will be much spatial overlap in regions activated across individuals, especially given the good spatial resolution of fMRI. Presumably this information is lost as a result of aggregating across individuals: even though each individual is doing the same thing slightly different brain regions are active.

Some studies support this contention. Having people think of a particular concept, e.g., "horse," leads to diffuse activation of many regions across the whole brain (Pereira et al., 2011). Pereira et al. (2011) also showed that at a certain level of abstraction these activation patterns could predict the topic being thought about and words associated with those topics. This was achieved by extracting a latent topic model from Wikipedia articles. Using machine learning technique a mapping was learnt between the latent factors that summarized the articles and patterns of distributed brain activity. This mapping could then be inverted to use the pattern of brain activity to predict the topic being thought about and hence words associated with that topic. Consequently, at quite a high level of abstraction there may be some consistency between topics being thought about and the spatial distribution of activation in the brain. However, we know of no work that relates individual concepts, such as "horse" to consistent patterns of activation across individuals. Moreover, the simple fact that these activations do not survive the subtraction methodology used in the reasoning studies summarized by Prado et al. (2011) suggests that across individuals there is little consistency in the brain regions activated.

The notion that for many different concepts and events people's own unique experience may fail to lead to patterns of brain activity that generalize fully across individuals is consistent with the subjective nature of probabilities in the new Bayesian paradigm. Our own unique experiences mean we may assign quite different probabilities to the same events. Indeed, if we did not differ in our beliefs in this way then there would be nothing to argue about at the social level where, it has been argued, most reasoning goes on (Hahn and Oaksford, 2007; Mercier and Sperber, 2011).

In summary, these imaging studies are not recording System 1 in action.

Functions, tasks, and active regions

I have concentrated so far on what imaging studies may miss in investigating System 1 processes. Before moving on to look at the difficulties in interpreting the activations that remain, I pause briefly to consider the relationship between cognitive tasks, the functions they engage and the interpretation of active brain regions. I argue that (i) function comes first, and two (ii) the function a task engages may be in dispute. In the next section, I trace the consequences of (i) and (ii) for the interpretations of the regions identified by Prado et al. (2011).

[3] This is also important because it suggests a unified account of System 1 and unconscious inference in perception and action (Oaksford, 2014, Submitted).

Function is assigned partly historically. For example, in investigating belief bias, Goel and Dolan (2003) contrasted correct and incorrect performance on trials that show a conflict between the validity of an inference and the truth of the conclusion (see (A) and (B)). One contrast revealed activation of right inferior prefrontal cortex (rIPFC) and the other of ventromedial prefrontal cortex (VMPFC). How do we interpret such findings? This question is answered partly in terms of the nature of the current task but also in terms of past history. So rIPFC is active when correct responses are made to conflict problems implicating inhibitory processes consistent with previous results. VMPFC is active when incorrect responses are made to conflict problems implicating intuitive, emotional processes, again consistent with previous results. The functions assigned to these regions are partly based on computational level assumptions. These determine the "correct" response and the assumption that "inhibition" is required to identify the correct response. But it is also based on history, what tasks (with assumed functions) have activated the region in the past. While this is all perfectly reasonable, there are potential problems.

First, there is the problem of a general historical bias. Just because a certain type of task, t_1, with a certain presumed function, f_1, was first found to activate a region, r_1, then this is the function associated with that region. But this is simply a historical artifact. If the current task, t_2, with presumed function, f_2, had been investigated first and found to activate region r_1 then f_2 would be the function presumed to be engaged when this region is activated and t_1 may be assumed to engage f_2 as well as f_1.

Second, this line of argument suggests that interpreting imaging results requires us to be very clear on the functions that cognitive tasks engage. Moreover, if this is clear then function drives interpretation. If region r_1 is activated by t_2, even though it has been previously associated with f_1, it must now be regarded as also computing f_2. At least there is no reason, other than history, to argue that instead t_2 engages f_1. Moreover, in cognitive science, and in particular deductive reasoning, the task/function relationship may be in dispute. So called deductive tasks, say t_1, are being interpreted as not engaging deduction, f_1, but rather probabilistic reasoning, f_2. We can only interpret the function of a brain region in terms of the tasks that engage those functions and activate that region. If our theory of the function engaged by a task changes, then so does our interpretation of what active brain regions are doing. For example, later on I argue that the computational level assumptions underlying the interpretation of belief bias results (Goel and Dolan, 2003) may be wrong (*NIRS, TMS and Belief Bias*). Imaging studies are only informative against the backdrop of a computational level theory of the tasks used in these studies. Consequently, whatever one's preferred research strategy, i.e., whether you concentrate on the implementational, algorithmic, or computational level, function comes first.[4]

Imaging beyond system 1

Against the backdrop of these last two arguments, I now consider the other patterns of activation that Prado et al. (2011) found with relational, categorical and propositional reasoning. With relational arguments, in particular in transitive inference, e.g., A is taller than B, B is taller than C. . .etc, is C taller than A?, Prado et al. (2011) found activation of bilateral PPC and right MFG consistent with the use of visual representations. Although this finding has recently been qualified by results showing that when the transitive chain involves quantifiers, all A are B, all B are C. . .etc, only left hemisphere activation is found (Prado et al., 2013). These findings suggest, what many researchers have suspected, that relational and spatial reasoning are not part of our core reasoning system. Rather when such arguments can be easily represented visually the mind/brain exploits this fact but this is a specific strategy. Moreover, as Prado et al. (2013) have shown, when this strategy is difficult, i.e., when the transitive chain involves whole sets and not individuals, the system reverts to the left brain interpreter.

Prado et al.'s (2013) results also argue against the mental model theory of quantified syllogistic reasoning. In this account, categorical reasoning proceeds over an imagistic representation of a small number of arbitrary exemplars of the sets described by the quantifiers. So according to mental model theory both categorical reasoning and relational reasoning should engage right lateralized systems. In contrast, the main probabilistic account of categorical reasoning, the probability heuristics model (Chater and Oaksford, 1999; Oaksford et al., 2002), suggests that a simple set of probabilistically motivated heuristics operate over linguistic representations of the premise and conclusion. Prado et al.'s (2011) results for categorical reasoning are consistent with this account. They show strong activation of left lateralized IFG and BG, regions most consistently associated with processing syntax and grammar. The heuristics in PHM select a syntactic conclusion frame using probabilistically motivated heuristics and then use other heuristics to determine the order of end terms in this syntactic frame (Oaksford et al., 2002). These heuristics depend on an ordering over the informativeness (the inverse of probability) of the premises. Specific content has the potential to alter this informativeness ordering leading the heuristics to make different predictions. While this possibility has never been experimentally tested, it shows that even this relatively abstractly defined probabilistic theory still relies on System 1, i.e., on content.

Prado et al. (2011) found that propositional arguments most strongly activate PPC and MeFG which have been associated with non-syntactic verbal processing and maintaining abstract rules in memory respectively. Perhaps the most researched and important area in propositional reasoning is conditional reasoning, i.e., reasoning using what is rendered in English as *if. . .then*. Most recent research has involved causal conditional reasoning, where it is clear that the specific contents are important. However, conditional reasoning has also been extensively researched using

[4] Clearly the weight of evidence matters here. For example, if across a broad range of different tasks, $t_1 \ldots t_n$, thought to engage probabilistic reasoning, r_1 is consistently activated but it is not in say t_{n+1}, i.e., a nominally deductive task, then we might be begin to be persuaded that probabilistic reasoning is not involved in deductive tasks. However, (i) this question has not been investigated with a broad range of different tasks, and (ii) as we have argued that the bulk of probabilistic reasoning is a System 1 process, i.e., a central process unlikely to be associated with a single isolable brain region.

abstract materials, which seemingly could not engage content. The fact that regions associated with maintaining abstract rules in memory are activated suggests that perhaps formal syntactic processes are directly involved. There are good arguments against this interpretation.

First, as I have argued, the functions engaged by a brain region may well be in dispute. Whether we need to use abstract rules in language processing or reasoning is contentious. In language processing the debate has raged since the advent of neural networks in the 1980s (Rumelhart, 1986). The issues hinge on whether generalization is achieved by abstract general rules or by similarity and analogy to pre-existing knowledge. Thus, as we discussed above, whether MeFG co-ordinates the processes involved in computing similarity and analogy or storing abstract rules is contentious from this perspective. An interesting prediction is that if computing similarity and analogy is involved in reasoning with abstract material one might expect more rather than less general knowledge to be activated. As materials become more abstract they will be similar to more of what we know, e.g., to all domains we tend to describe using conditionals. We may find an answer to this question once appropriate methods to image System 1in action are used.

Second, it seems doubtful that humans have evolved a specific module for handling abstract logical rules of inference that are the product of the last two millennia of logico-philosophical labor. Formal logic is a cultural product, a tool, for reasoning with pencil paper or computer. It is not the workings of the human mind made concrete in symbols. The *if...then* construction is used ubiquitously because it can be used to describe the various relationships or dependencies in the world, like causes, dispositions, intentions, regulations and so on, which allow us to predict what will happen next and to explain why what happened happened. The reasoning mind is likely to be very concrete constructing specific small scale models of reality in System 1, like Kahnemen's banana-vomit example or using specific relations, and reasoning over these (Oaksford and Chater, 2013, 2014; Oaksford, 2014, Submitted). These last two points make the argument that there are functions, f_2 and f_3, that are in contention to account for the tasks that engage MeFG. Consequently, there is reasonable doubt about whether it engages abstract rules.

I finish this section by looking again at the function of the core brain system identified by Prado et al. (2011). As I argued above, it seems unlikely that either System 1 or System 2 processes in most "deductive" reasoning tasks are like consciously performing mental arithmetic like that required to solve the bat and ball problem. However, in all reasoning tasks the results of these processes must become conscious and be turned into a verbal response to be delivered verbally (production task) or to match to a range of possible response options (selection task). What becomes conscious may also be a feeling of wrongness when the outputs of System 1 and System 2 conflict.[5] This would seem to be the shared common core of most reasoning tasks. But of course it is the final stage not the actual core of the reasoning process.

FURTHER IMAGING STUDIES

So far I have only discussed the fMRI localization studies included in Prado et al.'s (2011) meta-analysis. However, there are other imaging studies using fMRI and other imaging techniques, such as EEG using ERPs, Infra-red Spectroscopy (NIRS) and Transcranial Magnetic Stimulation (TMS), which are relevant to the dual-process aspect of the new paradigm. In this section, I deal with these further studies by the imaging technique used and then by the task/functions investigated.

fMRI studies

Here I look at further fMRI studies used to investigate (i) component process of deductive reasoning and (ii) the matching effect (Evans and Lynch, 1973; Oaksford and Stenning, 1992).

Component processes. Some fMRI (Fangmeier et al., 2006) and lesion studies (e.g., Reverberi et al., 2009) have concentrated on the component processes of deductive reasoning. Reverberi et al.'s (2009) lesion study was broadly consistent with the conclusion of Prado et al. (2011) that the right hemisphere and imagistic processing are not part of the core reasoning system. Right frontal lesions did not impair deductive reasoning. Patients with left frontal regions and impaired working memory did show deficits. More revealing evidence distinguishing the fast System 1 from the slow System 2 would be expected from studies investigating the time course of reasoning. Fangmeier et al. (2006) investigated the component processes of deductive reasoning separating out premise presentation, premise integration, and validation. These stages were defined by the timing of the presentation of two premises in visually presented spatial linear syllogisms, e.g., premises: V X (after 2 s), X W (after 6 s), conclusion: V W? (after 10 s). Perhaps unsurprisingly, given the visual presentation of premises, the premise presentation phase activated left and right occipital lobes. Premise integration and validation phases shifted activation toward frontal structures. As I have remarked, these purely visuospatial tasks are unlikely to invoke the same reasoning processes that underlie human *verbal*, reasoning. Moreover, the lack of content and the artificial pacing of the stimulus presentation to allow data collection using the relatively poor temporal resolution of fMRI are unlikely to be very revealing of the rapid System 1 in action.

Matching effects. There have been studies looking at phenomena that have provided evidence for dual processes, in particular, the matching effect (Evans and Lynch, 1973). Matching occurs when negations are included in the sentences used in a reasoning task. Usually these are in conditionals, e.g., *if there is an H then there is not a circle*. If asked to construct a falsifying instance of this rule people find it relatively easy because the falsifying instance, H and circle (a True/False instance TF), perceptually matches the named items in the rule. However, if they are asked the same question with the rule, *if there is an H then there is a circle*, then they find it more difficult. The TF instance is, e.g., H and square (or any non-circle), which does not completely match the named items. In a PET study, Houdé et al. (2000) showed that prior to perceptual inhibition training, this task primarily activated occipital visual regions, consistent with perceptual

[5] De Neys et al. (2008) have shown that the anterior cingulate cortex, associated with conflict detection, is active when these two systems conflict.

matching, but post inhibition training activation shifted to more frontal areas. More recently, Prado and Noveck (2006, 2007) have used fMRI to investigate the matching phenomenon. Prado and Noveck (2007) used a novel parametric variation approach identifying brain regions whose activation varied with the number of mismatches or negations in a rule. They also showed that frontal regions, which became more active with more mismatches, showed decreases in their interactions with visual cortex, consistent with inhibiting matching. Perceptual matching can be regarded as one of the perhaps many subsystems of System 1 (Stanovich, 2011) and the frontal systems that inhibit this system is System 2.

The new paradigm is an evolving body of theory and there is active disagreement over the interpretations of some phenomena. Evans (2003) cites Houdé et al. (2000) as support for the dual process theory. However, there are many reasons to doubt that these PET and fMRI studies are recording System 1 in action. First, there is a very close overlap in the regions activated in Houdé et al. (2000) pre-intervention phase and in Fangmeier et al.'s (2006) premise presentation phase. Of course, it is not surprising that presenting premises activates visual areas as written language is still a visual stimulus. There is no immediate reason to think that activity in these regions should be a source of reasoning bias. Second, matching is a far more nuanced phenomenon than described in Houdé et al. (2000) and in Prado and Noveck (2006, 2007). For example, in the original studies (Evans, 1972; Oaksford and Stenning, 1992) it occurs only for falsifying trials, like the example in the last paragraph. However, verifying trials (constructing True/True instances) show a similar pattern of mismatches as for falsifying trials. So, if the matching phenomenon were a simple perceptual matching effect then both types of trial should reveal the bias. Third, much simpler manipulations than inhibition training remove this bias. For example, using real world thematic content rather than abstract alphanumeric stimuli or shapes removes the bias (Oaksford and Stenning, 1992). This simple fact suggests that matching is not a major factor in biasing everyday reasoning. Moreover, making it easier to identify the "contrast class" for a negated constituent removes the bias. Logically the contrast class for *there is not a circle* can be anything, literally, that is not a circle (e.g., a coal scuttle). But in context it is clear that another shape is intended. If there were only two shapes and participants knew this, then matching is likely to disappear, as it does when using rules like *if there is a vowel, then there is not an even number* (Oaksford and Stenning, 1992). A number that is not even is obviously odd. Prado and Noveck (2007) did detect areas that were differentially active depending on the number of negations, i.e., right anterior pre-frontal cortex, and suggest that this may be involved in computing contrast classes.

Oaksford and Stenning (1992), (see also Oaksford and Moussakowski, 2004) argued that the matching phenomenon is part of the normal process of computing contrast classes which is made difficult by the use of abstract material. They also show how this account combines with the probabilistic component of the new paradigm to explain matching effects both in the Wason selection task (Oaksford and Chater, 1994, 2003, 2007) and in the conditional inference task (Oaksford et al., 2000). Constructing contrast classes is part of the System 1 processes involved in generating probabilities.

Why do these imaging studies show the effects they do, i.e., mismatches correlated with regions that are inhibiting visual areas? I suspect that this is part of the much more general phenomenon of suppressing distracting information in attentional control. If shown a picture of a white bear (Wegner, 1994) and told not to think about it, all you can think about is white bears. Similar patterns of activation are likely to occur on many tasks requiring the suppression of distractors regardless of whether they are reasoning tasks. Moreover, suppression in the visual modality can be made more difficult in the presence of noise in the auditory modality. There is also work on the neural basis of these effects (Smucny et al., 2013) which reveals similar interactions between brain regions as shown by Prado and Noveck (2007). fMRI scanners are very noisy places and PET scanners are also quite noisy. Consequently, while being scanned these attentional effects would be expected to be even more pronounced and to dominate the normal processes of contrast class construction. In normal discourse, a whole range of phonetic, syntactic, semantic and pragmatic factors contribute to making contrast class construction easy (Oaksford and Stenning, 1992). It is only in abstract tasks where these supports are removed that matching is observed.

In sum, there is good reason to doubt that these studies of matching bias tap into the fast System 1 responsible for the effects in Kahneman (2011) anecdotal example and in contrast class construction (although Prado and Noveck (2007) show some evidence for the localization of these latter processes). Rather the primary effects observed seem to be concerned with the general suppression of distractors observed in many tasks which are exacerbated by the noisy environment of the scanner.

NIRS, TMS and belief bias

The studies we looked at in the last section all used fMRI which has limited temporal resolution and so is perhaps unlikely to reveal much about fast System 1 processes. Where they have been revealing on System 2 processes this has primarily involved the function of dorsolateral pre-frontal cortex in inhibiting distracting information emanating from visual areas not of the analytic processes thought to require working memory. Perhaps a better insight into the neural processes involved at the interface between System 1 and System 2 might be found using imaging methods with greater temporal resolution. In this section, I briefly look and working using near infra-red spectroscopy and TMS.

A series of four studies using NIRS by Tsujii et al. investigated the role of inferior frontal cortex (IFC, which includes the IFG) in the belief bias effect (Tsujii and Watanabe, 2009, 2010; Tsujii et al., 2010, 2011). This effect has also been assumed to provide evidence for dual processes. The effect is usually investigated using quantified syllogisms which can be systematically varied along the binary dimensions of validity (valid, invalid) and believability of the conclusion (believable, unbelievable). For example, *No mammals are birds*, *All dogs are birds*, therefore, *No dogs are birds* is valid and believable, whereas *No pigeons are mammals*, *All pigeons are birds*, therefore, *No birds are mammals* is invalid and believable. The belief bias effect is an interaction effect (Evans et al., 1983) such that people endorse invalid believable conclusions as much as valid believable conclusions (92% in both cases), whereas they endorse valid unbelievable conclusions (46%) far more than invalid unbelievable conclusions (8%). Accuracy is far greater for congruent trials (valid/believable and invalid/unbelievable, 92%) than for incongruent trials (valid/unbelievable and invalid/believable, 37%). In these imaging studies accuracy on congruent and incongruent trials was the behaviorial dependent variable. Incongruent trials require the System 1 belief based response to be inhibited to allow the System 2 analytic response to be made.

In Tsujii et al.' studies they used manipulations to impair working memory performance either by using a dual task (Tsujii and Watanabe, 2009), time restrictions (Tsujii and Watanabe, 2010), or by using repetitive TMS on the IFC region (Tsujii et al., 2010) thought to be involved in working memory. High dual task load, short time restriction, and right IFC rTMS stimulation led to less accurate performance but only on incongruent trials. High dual task load and a short time restriction also reduced IFC/IFG activation but only in the right hemisphere. These findings suggest that right IFG is required to inhibit the System 1 heuristic or belief based response. In a further study, Tsujii et al. (2011) also used rTMS on the superior parietal lobule (SPL) as well as IFG using the belief bias paradigm. Stimulation in this region impaired performance on abstract syllogisms and incongruent trials, which they suggest require analytic System 2 processes. Tsujii et al. conclude that the function of right IFG is in inhibiting belief biased responding, the function of left IFG is a language area responsible for semantic processing and belief bias, while the function of bilateral SPL is analytic reasoning.

There are several points to make about these NIRS studies. First, the activations were integrated over a period lasting over a minute and so are not looking at rapid processes of the type that underlie Kahneman's System 1. Second, the results are not consistent with previous fMRI studies. For example, the seat of inhibitory processing has moved from DLPFC (BA 46) in Prado and Noveck (2007) to right IFG (BA 44, 45, 47). Moreover, there seems to be little evidence of Prado et al.'s (2011) core left lateralized deductive reasoning system. Further problems of interpretation arise from the interactional nature of the belief bias phenomenon.

Recently, Dube et al. (2010) showed that the belief bias interaction has been misinterpreted. They show that the interaction effect observed in belief bias is consistent with curvilinear ROC curves. Properly analyzed, accuracy remains the same between conditions, and believability effects are pure response biases. They argue that their modeling results, "provide support for processing theories of deduction that assume responses are driven by a graded argument-strength variable, such as the probability heuristic model proposed by Chater and Oaksford (1999)." Their results are also consistent with probabilistic single function dual process theory (Oaksford and Chater, 2012, 2014). There is a clear distinction between processes based on long term memory for our beliefs about the world and processes that require working memory. However, the single function approach argues that these processes, where they concern reasoning, are both probabilistic.

Dube et al.'s (2010) analysis shows that the belief bias phenomenon that underpins the theoretical framework (logical analytic System 2 and belief based/heuristic System 1) used to interpret Tsujii et al. results, may not actually exist. A similar state of affairs exists in the study of optimism bias (Weinstein and Klein, 1996) where proper statistical analysis (Harris and Hahn, 2011) has shown that this phenomenon, apparently investigated in many imaging studies (e.g., Sharot et al., 2011), may not actually exist. These re-analyses of these phenomena are at the computational level, i.e., they show that the actual functions being computed in these tasks may not be what they first seemed. As we argued in the section *Imaging beyond System 1*, theories of function drive the interpretation of these imaging results, i.e., its function first. Consequently the interpretation of Tsujii et al. results may need to be re-thought.

A paper aimed at making general theoretical points about the current state of imaging research into deductive reasoning is not the place to offer such a re-interpretation of these results. However, it is worth observing that the interpretation is going to be further complicated by the fact the that people seem to unconsciously process both the nominally analytic and heuristic responses as evidenced by the activation of brain regions associated with conflict detection, i.e., the anterior cingulate cortex, whether people make the supposedly biased response or not (De Neys et al., 2008). That is, both possible responses seem to be computed in System 1. Such findings tend to suggest that System 2 doesn't so much do analytic reasoning as adjudicate between possibilities and form a response (Oaksford, 2014, Submitted).

In sum, a major problem for imaging research is that there seem to be no onus to explore all the possible computational level interpretations of any set of results. Moreover, there is only a very loose connection between function and the activity of brain regions assumed to compute it. For example, the inference to SPL being the seat of analytic reasoning is based on a statistical tendency for rTMS stimulation of that region to impair abstract and incongruent tasks. In the light of Dube et al.'s analysis, it is very difficult to know what to make of this result. However, it most certainly does not tie this region to making deductive inference in a mental logic.

ERP and conditional inference

To explore the brain systems involved in the rapid System 1 processes, event related potentials recorded using EEG would seem to be the most promising route. The temporal resolution is excellent and many of the evoked waveforms have a well understood interpretation developed over many years of research. The studies I review here have all focused on the conditional reasoning paradigm. Inexplicably, some studies on conditional reasoning using ERPs have focused on contentless, abstract material (Bonnefond and Van der Henst, 2009). This is despite the fact that in the psychology of conditional inference, the dominant paradigm since Cummins et al. (1991) ground breaking paper has been the causal conditional inference task, which has arguably completely altered the theoretical landscape of research into the conditional.

The failure to consider the full theoretical possibilities is repeated in Bonnefond and Van der Henst (2013), who introduce the paper using the theoretical framework of mental logic, which has not been applied to any of the major results in conditional inference over the last twenty years of research. They argue that a sustained late positive component to the EEG waveform suggests that "participants consider logical arguments as a rule-governed sequence." The absence of an N400 (a negative going waveform at around 400 ms) associated with semantic processing is not consistent with apparent inconsistencies being semantic in origin rather than formal. The implication of their results is that, even though their materials introduced content, the main effect was to facilitate activation of terms expected as a matter of logical inference.

However, even more recently Bonnefond et al. (2014) investigated the correlates of defeaters in conditional inference. A defeater in a causal conditional reasoning task is an event that could prevent the cause from producing its effect. For example, *if you turn the key the car starts*, is defeated by the *petrol tank being empty* or the *battery being flat*. In the Cummins paradigm, causal conditionals are pretested for the number of defeaters they allow. The primary behavioral observation is that the more defeaters a conditional allows, the less willing participants are to endorse the MP inference. Bonnefond et al. (2014) replicated these results and found specific effects on EEG waveforms. Their main finding was that presenting the conclusion of an MP inference led to,

> "...a more pronounced N2 and less pronounced P3b for many disabler conditionals. In the ERP literature this specific N2/P3b pattern has been linked to the violation and satisfaction of expectations, respectively...Thereby, the present ERP findings support the idea that disabler retrieval specifically modulates our expectations that the standard MP conclusion will follow." (Bonnefond et al. (2014), p. 258).

It is suggested that these results are consistent with conditional inference not being mediated by formal logical rules. Indeed the first demonstration of defeater effects (Byrne, 1989) was interpreted as refuting mental logicians' explanation of why introducing alternative causes leads to reduced levels of the affirming the consequent fallacy (Braine et al., 1984). Such pragmatic factors may influence fallacies but they would not be expected to affect logical rules of inference such as MP if they play a role in real human inference. That is, Bonnefond et al.'s (2014) results showing the brain correlates of defeater information supplies just the evidence required to refute their interpretation of their own previous results (Bonnefond and Van der Henst, 2013). These results are also consistent with the probabilistic approach adopted in the new paradigm (Oaksford et al., 2000; Oaksford and Chater, 2007).

Another recent ERP study of conditional reasoning using the MP inference has shown a strong N400 component, which Bonnefond and Van der Henst (2013) did not observe (Blanchette and El-Deredy, 2014). This component of time locked EEG signals is strongly related to the processing of semantic content (Kutas and Hilyard, 1980). This early response to the premises of an argument is consistent with Kahneman's banana-vomit example: the content of the premises is processed very rapidly. Blanchette and El-Deredy (2014) conclude that "conditional reasoning is not a purely formal process but that it importantly implicates semantic processing." This conclusion is consistent with rapid System 1 processes which generate the kinds of information we discussed earlier and perhaps build an initial concrete model of the described situation. Of course this interpretation does not preclude System 2 involvement at some later point in the process.

In summary, the last two ERP studies reviewed are the closest to seeing System 1 in action. Bonnefond et al. (2014) also very commendably concede that their results question their earlier interpretation of their findings using abstract materials. Nonetheless, it is concerning that imaging results are published which do not consider the current state of theoretical development a topic has achieved in other areas of cognitive science. I can but agree with Bonnefond et al.'s (2014, p. 260) conclusion:

> "Behavioral studies have also focused on the impact of different types of conditionals (e.g., tips, warnings, promises, and causal statements)...We belief [*sic*] that the present study will pave the way for a further exploration of the neural basis of these content factors in future studies."

Such studies are a pressing need in this area but also required are methods that allow a much tighter integration between formal computational level theories of function and the brain.

CONCLUSION

In this paper, I have discussed the interpretation of what is currently known about the brain systems involved in human deductive reasoning mainly using different imaging techniques to localize function to specific brain regions. In doing so, I have dealt with the results of Prado et al.'s (2011) meta-analysis and a range of other results from the perspective of the new paradigm in human reasoning. Prado et al. (2011) identified a relatively restricted group of brain regions consistently activated in deductive reasoning tasks. Like the studies in the meta-analysis, Prado et al. (2011) interpret their results largely in terms of mental logic and mental models theories. In this paper, I have reinterpreted most of these findings in terms of the new paradigm in reasoning which is a probabilistic dual process theory.

The first substantive issue to emerge was that Prado et al. identify their core left lateralized system with Gazzaniga's left brain interpreter hypothesis. This identification is not consistent with this system being dedicated to deductive reasoning. The kinds of inferences that motivates Gazzaniga's hypothesis are elaborative, defeasible inferences of the type that motivated the introduction of the probabilistic approach to human reasoning (Oaksford and Chater, 1991). Moreover, this is exactly the mode of inference, i.e., non-demonstrative inference involving world knowledge, which Fodor (1983) identified with central cognitive processes, i.e., those processes least likely to be subserved by an isolable cognitive module.

The second substantive issue concerned the apparent inability of the studies used in the meta-analysis to uncover the brain regions involved in System 1 processes. These are highly content dependent and are responsible for the automatic computation of a range of information used in inference. As I argued, the subtraction methodology and meta-analytic approach meant the whole brain diffuse activations caused by specific contents (Pereira et al., 2011) must have been subtracted out. Thus the current methodology would appear to leave us largely ignorant of the brain systems involved in System 1. I also explored a range of other results using different imaging techniques and two recent ERP studies (Blanchette and El-Deredy, 2014; Bonnefond et al., 2014) seem to show results capable of illuminating the nature of System 1.

Many of the studies using other techniques also seemed to have problems related to the third substantive issue concerning the interpretation of active brain regions. The interpretation of these findings depends on the computational level theory of the function engaged by a cognitive task. In general, either the attribution of function provided by Prado et al. (2011) was broadly consistent with the new paradigm, e.g., categorical reasoning, or it was clear that there were multiple interpretations of the function a region computed, e.g., abstract rules vs. similarity and analogy or small scale models of specific relations. Similar problems arose for the interpretation of studies of matching bias using fMRI (Prado and Noveck, 2006, 2007) and the NIRS studies of Tsujii et al. There is a failure to consider the full range of computational level interpretations available in the area.

While the localization approach has provided useful information about the brain systems involved in deductive reasoning, and its extension to looking a functional connectivity may be even more revealing, the interpretation of these results remains problematic. Certainly, following Goel (2007), I doubt that any single isolable region "does" deductive or inductive reasoning. Reasoning and inference are not special purpose add-ons to the cognitive system. Unconscious inference in perception and action, elaborative inference in language understanding and explicit verbal reasoning are major functions of the brain. These processes allow us to act adaptively and comprehend an uncertain world the state of which at any point in time we are mostly ignorant. Inference allows us to make the best guess about what will happen next, what someone means, and whether what they said is a good argument. One would imagine that a large amount of cortex would be dedicated to these processes.

System 1 automatically generates a large range of information and if the results using simple stimuli, e.g., thinking of horses, is anything to go by many diffuse brain regions will be activated by the materials in a reasoning problem. It is a reasonable hypothesis that this is the source of information for the left brain interpreter. The nature of System 2 is less clear. Results on logical intuitions suggest that people are unconsciously generating the logically correct answer even as they give the biased response. A radical possibility is that analytic (putatively System 2) and heuristic/probabilistic process (putatively System 1) are both computed by the one System, i.e., System 1 (Oaksford, 2014, Submitted). That is, in spontaneous human reasoning, without logical training, pencil and paper, computer, or friends, there is no conscious analytic process akin to the mental arithmetic required to solve the bat and ball problem. That is, all spontaneous reasoning is unconscious (Lakoff and Johnson, 1999). System 2 is where the products of these processes are posted and decisions made about which response to go with and which response to inhibit (Oaksford, 2014, Submitted). This is the core system most likely identified in Prado et al.'s (2011) meta-analysis, and thus interpreted, it seems that fMRI and lesions studies have been most revealing of these slow System 2 processes.

An approach is required that can reveal how the interactions between Systems unfolds over time and how these different systems communicate with the process of forming a response. System 1 responds rapidly and as we have seen, two very recent EEG studies (Blanchette and El-Deredy, 2014; Bonnefond et al., 2014), with good temporal resolution seem to provide the most informative studies of System 1 in action. Perhaps the most important innovation would be to conduct studies that had the potential to tightly correlate formal computational models of reasoning to brain activation be it using ERPs or fMRI. Many models of reasoning are formally well specified. These tend to be mostly emanating from the probabilistic side of the new paradigm (Oaksford and Chater, 1994; Chater and Oaksford, 1999; Oaksford et al., 2000). Formal models of dual processes are less in evidence, although Klauer et al. (2010), for example, present a formal model with a specific parameter that indexes System 1 vs. System 2 involvement. The value of such formal models is that model based imaging can reveal correlations between specific parameters of the formal model and brain activation providing much tighter integration between imaging results and the computational level. Pursing this line, I would argue, could provide a more integrated approach bringing the computational level and the implementational level into closer alignment.

REFERENCES

Adams, E. W. (1998). *A Primer of Probability Logic.* Stanford, CA: CSLI Publications.

Barwise, J., and Perry, J. (1983). *Situations and Attitudes.* Cambridge, MA: MIT Press.

Bennett, J. (2003). *A Philosophical Guide to Conditionals.* Oxford England: Oxford University Press.

Blanchette, I., and El-Deredy, W. (2014). A ERP investigation of conditional reasoning with emotional and neutral contents. *Brain Cogn.* 91, 45–53. doi: 10.1016/j.bandc.2014.08.001

Bonnefond, M., Kaliuzhna, M., Van der Henst, J., and De Neys, W. (2014). Disabling conditional inferences: an EEG study. *Neuropsychologia* 56, 255–262. doi: 10.1016/j.neuropsychologia.2014.01.022

Bonnefond, M., and Van der Henst, J. (2009). What's behind an inference? An EEG study with conditional arguments. *Neuropsychologia* 47, 3125–3133. doi: 10.1016/j.neuropsychologia.2009.07.014

Bonnefond, M., and Van der Henst, J. B. (2013). Deduction electrified: ERPs elicited by the processing of words in conditional arguments. *Brain Lang.* 124, 244–256. doi: 10.1016/j.bandl.2012.12.011

Booth, J., Coch, D., Fischer, K., and Dawson, G. (2007). "Brain bases of learning and development of language and reading," in *Human Behavior, Learning and the Developing Brain*, eds D. Coch, G. Dawson and K. W. Fischer (New York: Guilford Press), 279–300.

Braine, M. D. S., Reiser, B. J., and Rumain, B. (1984). "Some empirical justification for a theory of natural propositional logic," in *The Psychology of Learning and Motivation*, ed G. H. Bower (New York: Academic Press), 317–371.

Bunge, S. A., Kahn, I., Wallis, J. D., Miller, E. K., and Wagner, A. D. (2003). Neural circuits subserving the retrieval and maintenance of abstract rules. *J. Neurophysiol.* 90, 3419–3428. doi: 10.1152/jn.00910.2002

Byrne, R. M. J. (1989). Suppressing valid inferences with conditionals. *Cognition* 31, 61–83. doi: 10.1016/0010-0277(89)90018-8

Chater, N., and Oaksford, M. (1999). The probability heuristics model of syllogistic reasoning. *Cogn. Psychol.* 38, 191–258. doi: 10.1006/cogp.1998.0696

Chater, N., Oaksford, M., Nakisa, R., and Redington, M. (2003). Fast, frugal and rational: how rational norms explain behavior. *Organ. Behav. Hum. Decis. Process.* 90, 63–86. doi: 10.1016/s0749-5978(02)00508-3

Clark, A. (2013). Whatever next? Predictive brains, situated agents and the future of cognitive science. *Behav. Brain Sci.* 36, 181–253. doi: 10.1017/S0140525X12000477

Cohen Kadosh, R., Henik, A., Rubinsten, O., Mohr, H., Dori, H., van de ven, V., et al. (2005). Are numbers special? The comparison systems of the human brain investigated by fMRI. *Neuropsychologia* 43, 1238–1248. doi: 10.1016/j.neuropsychologia.2004.12.017

Cummins, D. D., Lubart, T., Alksnis, O., and Rist, R. (1991). Conditional reasoning and causation. *Mem. Cognit.* 19, 274–282. doi: 10.3758/bf03211151

Dayan, P., and Hinton, G. (1996). Varieties of Helmholtz machine. *Neural Netw.* 9, 1385–1403. doi: 10.1016/s0893-6080(96)00009-3

De Neys, W. (2012). Bias and conflict: a case for logical intuitions. *Perspect. Psychol. Sci.* 7, 28–38. doi: 10.1177/1745691611429354

De Neys, W. (2014). Conflict detection, dual processes and logical intuitions: some clarifications. *Think. Reason.* 20, 169–187. doi: 10.1080/13546783.2013.854725

De Neys, W., Vartanian, O., and Goel, V. (2008). Smarter than we think: when our brains detect that we are biased. *Psychol. Sci.* 19, 483–489. doi: 10.1111/j.1467-9280.2008.02113.x

Dennett, D. (1998). "Reflections on language and mind," in *Language and Thought: Interdisciplinary Themes*, eds P. Carruthers and J. Boucher (Cambridge: Cambridge University Press), 284–294.

Dube, C., Rotello, C. M., and Heit, E. (2010). Assessing the belief bias effect with ROCs: it's a response bias effect. *Psychol. Rev.* 117, 831–863. doi: 10.1037/a0019634

Elqayam, S., and Over, D. E. (2013). New paradigm psychology of reasoning: an introduction to the special issue edited by Elqayam, Bonnefon and over. *Think. Reason.* 19, 249–265. doi: 10.1080/13546783.2013.841591

Evans, J. St. B. T. (1972). Reasoning with negatives. *Br. J. Psychol.* 63, 213–219. doi: 10.1111/j.2044-8295.1972.tb02102.x

Evans, J. St. B. T. (2003). In two minds: dual-process accounts of reasoning. *Trends Cogn. Sci.* 7, 454–459. doi: 10.1016/j.tics.2003.08.012

Evans, J. St. B. T. (2010). *Thinking Twice: Two Minds in One Brain.* New York, NY: Oxford University Press.

Evans, J. St. B. T., Barston, J. L., and Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Mem. Cognit.* 11, 295–306. doi: 10.3758/bf03196976

Evans, J. B., and Lynch, J. S. (1973). Matching bias in the selection task. *Br. J. Psychol.* 64, 391–397. doi: 10.1111/j.2044-8295.1973.tb01365.x

Evans, J. St. B. T., and Over, D. E. (2004). *If.* Oxford: Oxford University Press.

Evans, J. St. B. T., and Over, D. E. (2013). Reasoning to and from belief: deduction and induction are still distinct. *Think. Reason.* 19, 267–283. doi: 10.1080/13546783.2012.745450

Fangmeier, T., Knauff, M., Ruff, C. C., and Sloutsky, V. (2006). fMRI evidence for a three-stage model of deductive reasoning. *J. Cogn. Neurosci.* 18, 320–334. doi: 10.1162/jocn.2006.18.3.320

Fodor, J. A. (1983). *The Modularity of Mind.* Cambridge, MA: MIT Press.

Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 815–836. doi: 10.1098/rstb.2005.1622

Friston, K. (2008). Hierarchical models in the brain. *PLoS Comput. Biol.* 4:e1000211. doi: 10.1371/journal.pcbi.1000211

Goel, V. (2007). Anatomy of deductive reasoning. *Trends Cogn. Sci.* 11, 435–441. doi: 10.1016/j.tics.2007.09.003

Goel, V., Buchel, C., Frith, C., and Dolan, R. (2000). Dissociation of mechanisms underlying syllogistic reasoning. *Neuroimage* 12, 504–514. doi: 10.1006/nimg.2000.0636

Goel, V., and Dolan, R. J. (2003). Explaining modulation of reasoning by belief. *Cognition* 87, B11–B22. doi: 10.1016/s0010-0277(02)00185-3

Goel, V., and Dolan, R. J. (2004). Differential involvement of left prefrontal cortex in inductive and deductive reasoning. *Cognition* 93, B109–B121. doi: 10.1016/j.cognition.2004.03.001

Goldvarg, E., and Johnson-Laird, P. N. (2001). Naive causality: a mental model theory of causal meaning and reasoning. *Cogn. Sci.* 25, 565–610. doi: 10.1207/s15516709cog2504_3

Grodzinsky, Y., and Santi, A. (2008). The battle for Broca's region. *Trends Cogn. Sci.* 12, 474–480. doi: 10.1016/j.tics.2008.09.001

Haack, S. (1975). *Deviant Logic.* Cambridge: Cambrdige University Press.

Hahn, U. (2014). The Bayesian boom: good thing or bad?. *Front. Psychol.* 5:765. doi: 10.3389/fpsyg.2014.00765

Hahn, U., and Oaksford, M. (2007). The rationality of informal argumentation: a Bayesian approach to reasoning fallacies. *Psychol. Rev.* 114, 704–732. doi: 10.1037/0033-295x.114.3.704

Harris, A. J. L., and Hahn, U. (2011). Unrealistic optimism about future life events: a cautionary note. *Psychol. Rev.* 118, 135–154. doi: 10.1037/a0020997

Heit, E., and Rotello, C. M. (2010). Relations between inductive reasoning and deductive reasoning. *J. Exp. Psychol. Learn. Mem. Cogn.* 36, 805–812. doi: 10.1037/a0018784

Houdé, O., Zago, L., Mellet, E., Moutier, S., Pineau, A., Mazoyer, B., et al. (2000). Shifting from the perceptual brain to the logical brain: the neural impact of cognitive inhibition training. *J. Cogn. Neurosci.* 12, 721–728. doi: 10.1162/089892900562525

Johnson-Laird, P. N. (1983). *Mental Models.* Cambridge: Cambridge University Press.

Johnson-Laird, P. N., and Byrne, R. M. J. (1991). *Deduction.* Hillsdale, N. J.: Lawrence Erlbaum Press.

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M. S., and Caverni, J. P. (1999). Naïve probability: a mental model theory of extensional reasoning. *Psychol. Rev.* 106, 62–88. doi: 10.1037//0033-295x.106.1.62

Kahneman, D. (2011). *Thinking, Fast and Slow.* London: Penguin Books.

Klauer, K. C., Beller, S., and Hütter, M. (2010). Conditional reasoning in context: a dual-source model of probabilistic inference. *J. Exp. Psychol. Learn. Mem. Cogn.* 36, 298–323. doi: 10.1037/a0018705

Kutas, M., and Hilyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207, 203–205. doi: 10.1126/science.7350657

Lakoff, G., and Johnson, M. (1999). *Philosophy in the Flesh.* New York: Basic Books.

Lassiter, D., and Goodman, N. D. (2015). How many kinds of reasoning? Inference, probability and natural language semantics. *Cognition* 136, 123–134. doi: 10.1016/j.cognition.2014.10.016

Manktelow, K. I. (2012). *Thinking and Reasoning.* Hove: Psychology Press.

Marr, D. (1982). *Vision.* San Francisco: W. H. Freeman.

Mercier, H., and Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behav. Brain Sci.* 34, 57–74. doi: 10.1017/s0140525x10000968

Monti, M. M., and Osherson, D. N. (2012). Logic, language and the brain. *Brain Res.* 1428, 33–42. doi: 10.1016/j.brainres.2011.05.061

Oaksford, M., and Chater, N. (1991). Against logicist cognitive science. *Mind and Language* 6, 1–38. doi: 10.1111/j.1468-0017.1991.tb00173.x

Oaksford, M., and Chater, N. (1993). "Reasoning theories and bounded rationality," in *Rationality*, eds K. I. Manktelow and D. E. Over (London: Routledge), 31–60.

Oaksford, M., and Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychol. Rev.* 101, 608–631. doi: 10.1037//0033-295x.101.4.608

Oaksford, M., and Chater, N. (1995). Theories of reasoning and the computational explanation of everyday inference. *Think. Reason.* 1, 121–152. doi: 10.1080/13546789508251501

Oaksford, M., and Chater, N. (2001). The probabilistic approach to human reasoning. *Trends Cogn. Sci.* 5, 349–357. doi: 10.1016/s1364-6613(00)01699-5

Oaksford, M., and Chater, N. (2003). Optimal data selection: revision, review and reevaluation. *Psychon. Bull. Rev.* 10, 289–318. doi: 10.3758/bf03196492

Oaksford, M., and Chater, N. (2007). *Bayesian Rationality: The Probabilistic Approach to Human Reasoning.* Oxford: Oxford University Press.

Oaksford, M., and Chater, N. (2009). Precis of bayesian rationality: the probabilistic approach to human reasoning. *Behav. Brain Sci.* 32, 69–120. doi: 10.1017/S0140525X09000284

Oaksford, M., and Chater, N. (2010). "Cognition and conditionals: an introduction," in *Cognition and Conditionals: Probability and Logic in Human Thinking*, eds M. Oaksford and N. Chater (Oxford: Oxford University Press), 3–36.

Oaksford, M., and Chater, N. (2011). "Dual systems and dual processes but a single function," in *The Science of Reason: A Festschrift for Jonathan St. B. T. Evans*, eds K. I. Manktelow, D. E. Over and S. Elqayam (Hove: Psychology Press), 339–351.

Oaksford, M., and Chater, N. (2012). Dual processes, probabilities and cognitive architecture. *Mind Soc.* 11, 15–26. doi: 10.1007/s11299-011-0096-3

Oaksford, M., and Chater, N. (2013). Dynamic inference and everyday conditional reasoning in the new paradigm. *Think. Reason.* 19, 346–379. doi: 10.1080/13546783.2013.808163

Oaksford, M., and Chater, N. (2014). Probabilistic single function dual process theory and logic programming as approaches to non-monotonicity in human vs. artificial reasoning. *Think. Reason.* 20, 269–295. doi: 10.1080/13546783.2013.877401

Oaksford, M., Chater, N., and Larkin, J. (2000). Probabilities and polarity biases in conditional inference. *J. Exp. Psychol. Learn. Mem. Cogn.* 26, 883–899. doi: 10.1037//0278-7393.26.4.883

Oaksford, M., and Hahn, U. (2007). "Induction, deduction and argument strength in human reasoning and argumentation," in *Inductive Reasoning: Experimental, Developmental and Computational Approaches*, eds A. Feeney and E. Heit (Cambridge: Cambridge University Press), 269–301.

Oaksford, M., and Moussakowski, M. (2004). Negations and natural sampling in data selection: ecological versus heuristic explanations of matching bias. *Mem. Cognit.* 32, 570–581. doi: 10.3758/bf03195848

Oaksford, M., Roberts, L., and Chater, N. (2002). Relative informativeness of quantifiers used in syllogistic reasoning. *Mem. Cognit.* 30, 138–149. doi: 10.3758/bf03195273

Oaksford, M., and Stenning, K. (1992). Reasoning with conditionals containing negated constituents. *J. Exp. Psychol. Learn. Mem. Cogn.* 18, 835–854. doi: 10.1037//0278-7393.18.4.835

Over, D. E. (2009). New paradigm psychology of reasoning. *Think. Reason.* 15, 431–438. doi: 10.1080/13546780903266188

Parsons, L. M., and Osherson, D. (2001). New evidence for distinct right and left brain systems for deductive versus probabilistic reasoning. *Cereb. Cortex* 11, 954–965. doi: 10.1093/cercor/11.10.954

Pereira, F., Detre, G., and Botvinick, M. (2011). Generating text from functional brain images. *Front. Hum. Neurosci.* 5:72. doi: 10.3389/fnhum.2011.00072

Pfeifer, N. (2013). The new psychology of reasoning: a mental probability logical perspective. *Think. Reason.* 19, 329–345. doi: 10.1080/13546783.2013.838189

Pfeifer, N., and Kleiter, G. (2010). "Mental probability logic," in *Cognition and Conditionals: Probability and Logic in Human Thinking*, eds M. Oaksford and N. Chater (Oxford, UK: Oxford University Press), 153–173.

Prado, J., Chadha, A., and Booth, J. R. (2011). The brain network for deductive reasoning: a quantitative meta-analysis of 28 neuroimaging studies. *J. Cogn. Neurosci.* 23, 3483–3497. doi: 10.1162/jocn_a_00063

Prado, J., Mutreja, R., and Booth, J. R. (2013). Fractionating the neural substrates of transitive reasoning: task-dependent contributions of spatial and verbal representations. *Cereb. Cortex* 23, 499–507. doi: 10.1093/cercor/bhr389

Prado, J., and Noveck, I. A. (2006). How reaction time measures elucidate the matching bias and the way negations are processed. *Think. Reason.* 12, 309–328. doi: 10.1080/13546780500371241

Prado, J., and Noveck, I. A. (2007). Overcoming perceptual features in logical reasoning: a parametric functional magnetic resonance imaging study. *J. Cogn. Neurosci.* 19, 642–657. doi: 10.1162/jocn.2007.19.4.642

Quine, W. V. O. (1953). "Two dogmas of empiricism," in *From a Logical Point of View*, ed W. V. O. Quine (Cambridge, MA: Harvard University Press), 20–46.

Reverberi, C., Shallice, T., D'Agostini, S., Skrap, M., and Bonatti, L. L. (2009). Cortical bases of elementary deductive reasoning: inference, memory and metadeduction. *Neuropsychologia* 47, 1107–1116. doi: 10.1016/j.neuropsychologia.2009.01.004

Rips, L. J. (1994). *The Psychology of Proof: Deductive Reasoning in Human Thinking.* Cambridge, MA: The MIT Press.

Rips, L. J. (2001). Two kinds of reasoning. *Psychol. Sci.* 12, 129–134. doi: 10.1111/1467-9280.00322

Rips, L. J. (2002). "Reasoning," in *Stevens' Handbook of Experimental Psychology: Vol. 2. Cognition*, 3rd Edn. eds H. F. Pashler and D. L. Medin (New York: Wiley), 317–362.

Roser, M. E., and Gazzaniga, M. S. (2006). "The interpreter in human psychology," in *The Evolution of Primate Nervous Systems*, eds T. M. Preuss and J. H. Kaas (Oxford, UK: Academic Press), 503–508.

Rumelhart, D. E., and J.L. McClelland and the PDP Research Group. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations.* Cambridge, MA: MIT Press.

Schroyens, W. (2010). "Logic and/in psychology: the paradoxes of material implication and psychologism in the cognitive science of human reasoning," in *Cognition and Conditionals: Probability and Logic in Human Thinking*, eds M. Oaksford and N. Chater (New York, NY: Oxford University Press), 69–84.

Sharot, T., Korn, C., and Dolan, R. (2011). How unrealistic optimism is maintained in the face of reality. *Nat. Neurosci.* 14, 1475–1479. doi: 10.1038/nn.2949

Smucny, J., Rojas, D. C., Eichman, L. C., and Tregallas, J. R. (2013). Neuronal effects of auditory distraction on visual attention. *Brain Cogn.* 81, 263–270. doi: 10.1016/j.bandc.2012.11.008

Stanovich, K. E. (2011). *Rationality and the Reflective Mind.* Oxford: Oxford University Press.

Stanovich, K. E., and West, R. F. (2000). Individual differences in reasoning: implications for the rationality debate. *Behav. Brain Sci.* 23, 645–726. doi: 10.1017/s0140525x00003435

Thompson, V. A., Prowse Turner, J. A., and Pennycook, G. (2011). Intuition, reason and metacognition. *Cogn. Psychol.* 63, 107–140. doi: 10.1016/j.cogpsych.2011.06.001

Tsujii, T., Masuda, S., Akiyama, T., and Watanabe, S. (2010). The role of inferior frontal cortex in belief-bias reasoning: an rTMS study. *Neuropsychologia* 48, 2005–2008. doi: 10.1016/j.neuropsychologia.2010.03.021

Tsujii, T., Sakatani, K., Masuda, S., Akiyama, T., and Watanabe, S. (2011). Evaluating the roles of the inferior frontal gyrus and superior parietal lobule in deductive reasoning: an rTMS study. *Neuroimage* 58, 640–646. doi: 10.1016/j.neuroimage.2011.06.076

Tsujii, T., and Watanabe, S. (2009). Neural correlates of dual-task effect on belief-bias syllogistic reasoning: a near-infrared spectroscopy study. *Brain Res.* 1287, 118–125. doi: 10.1016/j.brainres.2009.06.080

Tsujii, T., and Watanabe, S. (2010). Neural correlates of belief-bias reasoning under time pressure: a near-infrared spectroscopy study. *Neuroimage* 50, 1320–1326. doi: 10.1016/j.neuroimage.2010.01.026

Ullman, M. T. (2006). Is Broca's area part of a basal ganglia thalamocortical circuit? *Cortex* 42, 480–485. doi: 10.1016/s0010-9452(08)70382-4

Varley, R., and Siegal, M. (2000). Evidence for cognition without grammar from causal reasoning and theory of mind in an agrammatic aphasic patient. *Curr. Biol.* 10, 723–726. doi: 10.1016/s0960-9822(00)00538-8

Wegner, D. M. (1994). *White Bears and other Unwanted Thoughts: Suppression, Obsession and the Psychology of Mental Control.* New York, NY: The Guilford Press.

Weinstein, N. D., and Klein, W. M. (1996). Unrealistic optimism: present and future. *J. Soc. Clin. Psychol.* 15, 1–8. doi: 10.1521/jscp.1996.15.1.1

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: 08 October 2014; accepted: 10 February 2015; published online: 27 February 2015.

Citation: Oaksford M (2015) Imaging deductive reasoning and the new paradigm. Front. Hum. Neurosci. ***9****:101. doi: 10.3389/fnhum.2015.00101*

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

frontiers in Human Neuroscience

OPINION
published: 03 November 2015
doi: 10.3389/fnhum.2015.00588

Nothing new under the sun, or the moon, or both

Luca L. Bonatti[1], *Paolo Cherubini*[2,3] *and Carlo Reverberi*[2,3]*

[1] Institución Catalana de Investigación y Estudios Avanzados and Universitat Pompeu Fabra, Barcelona, Spain, [2] Department of Psychology, University of Milano-Bicocca, Milan, Italy, [3] NeuroMi - Milan Center for Neuroscience, Milan, Italy

Keywords: reasoning, deduction, probability, dual system theory, mental models, mental logic, neuroimaging

OPEN ACCESS

Edited by:
Jérôme Prado,
Centre National de la Recherche Scientifique, France

Reviewed by:
Evan Heit,
University of California, Merced, USA

Correspondence:
Carlo Reverberi
carlo.reverberi@unimib.it

Received: *27 June 2015*
Accepted: *09 October 2015*
Published: *03 November 2015*

Citation:
Bonatti LL, Cherubini P and Reverberi C (2015) Nothing new under the sun, or the moon, or both. Front. Hum. Neurosci. 9:588. doi: 10.3389/fnhum.2015.00588

The investigation of the mechanisms and principles of human reasoning is as ancient as the history of philosophy. It has always been clear that there is something special that allows humans, to a greater degree than other animals, to think about future states, make plans, have rational discussions, handle complex social situations, and invent marvelous things such as science. What this "something" was, however, has remained buried in mystery, and it still partially is. At the same time, demonstrations of human rationality have always been countered by staggering examples of bad reasoning, in history, in psychology, and, as many people (not us) will admit, in personal experience. The camp of psychologists and philosophers has thus been divided among those who were more impressed by the successes of humans against nature (Aristotle, Bacon, Descartes, Kant, or closer to us, the neopositivists; in psychology, Johnson-Laird, Holyoak, Newell and Simon, the Mental Logic camp) and those who were more impressed by their miserable failures (Bacon, Schoepnhauer, Kierkegaard, the nichilists, or the deconstructivists; in psychology, Tversky, Kahnemann, Evans, etc.). The latter group has argued that developing a theory of rational/logical reasoning is doomed because there is no object to study. The former group has tried to explain the (admittedly limited) rationality of the mind by developing theories of the mental representations and processes involved in deductive, causal, or probabilistic reasoning (O'Brien, 1995; Braine and O'Brien, 1998; Goldvarg and Johnson-Laird, 2001; Johnson-Laird, 2010): call this approach the Not-So-New Paradigm.

Recently, a way to reconcile the angelic and the demoniac aspects of human reasoning has taken the form of a single theory, the Dual System theory. As its name says, it replaces two alternative theories with one single theory which postulates two alternative subsystems. One may get the impression that the Dual System theory amounts to a mere reshuffling of the problems it was supposed to address, however, some of its claims may make it more than a simple trick of cards. The theory holds that one of the two systems is evolutionarily ancient, implicit, fast, mostly geared to track statistical regularities, whereas the second system is explicit, slow, effortful, error-prone, evolutionarily more recent, and perform abstract and logical reasoning. It is the characteristics of this second system that explain human errors with logical or complex probabilistic problems. Merge Bayesianism to this theory and you get what Oaksford calls the "New Paradigm," which, he writes, is "based on Bayesian probability and dual processes" (Oaksford, 2015). Not only does the New Paradigm offer a novel theoretical framework to advance our knowledge of human reasoning, but it also offer "an alternative theoretical framework to those typically assumed in imaging research on deductive reasoning."

We cannot feel the same enthusiasm. First, it seems to us that explaining human reasoning by constraining it within the dual system theory is overly optimistic. Even within the narrow realm of deductive reasoning, many systems are likely involved. Certainly beyond deduction a whole constellation of inferential systems exist, and the interaction between them is neither simple nor predictable along the very rough boundaries provided by the dual system theory. Infants seem to be able to draw correct probabilistic inferences, both before and after being able to verbalize their reasoning (Téglás et al., 2007, 2011, 2015), but it is not clear if these abilities are implicit or explicit. So, does probabilistic reasoning belong to System 1 or 2?

There is also strong evidence that rational problem solving is deeply entrenched in the human mind at its earliest stages. Infants understand goals and the optimality of actions in a variety of situations difficult to capture by the postulation of a single, non-rational, system (Gergely et al., 1995, 2002; Csibra, 2008; Csibra and Gergely, 2009; Southgate and Csibra, 2009); they explore unknown situations making very specific hypotheses and testing them (Gweon and Schulz, 2011; Stahl and Feigenson, 2015); and they know how to interpret simple probabilistic situations and how event probabilities change in many different contexts (Téglás et al., 2011, 2015). What system do these abilities belong to, and, is it useful to even ask this question? With the little we know about basic reasoning abilities and their development, it is hard to see how jumping from paradigm to paradigm can help in developing the necessary knowledge. Finally, as Oaksford himself recalls, the Dual System theory cuts the pie in the wrong way. For example, it is an assumption of the theory that errors in deductive reasoning depend on it being a System-2-kind of phenomenon. However, we now know that an important part of deduction is implicit (De Neys and Schaeken, 2007; De Neys, 2012; Reverberi et al., 2012b), and that many easy deductive inferences are fast, spontaneous, and make no use of working memory to hold intermediate conclusions (Braine and O'Brien, 1998; Johnson-Laird, 2010), something that would make them a System-1-like process. Again, does deduction belong to System 1 or System 2? We believe that the best way to address this question is to refuse to answer in terms of a theory that is too coarse to provide any substantial answer. In short, we fail to see what is new in the New Paradigm, insofar as its novelty depends on the adoption of the Dual System theory.

Second, besides the Dual Theory, the novelty of the New Paradigm entirely consists of its probabilistic claim, mostly spelled out in a Bayesian framework. We agree with Oaskford that Bayesianism has made substantial new progress in the understanding of human reasoning, although the framework is so powerful that it is difficult to find its limits (Endress, 2013). However, it is an illusion to think that such progress is reason to dismiss the very same questions with which the Not-So-New paradigm struggles. Bayesianism is a theory about how hypotheses change in the face of experience. There is no Bayesian Theory to begin with, if one does not specify the language with which the very same hypotheses whose degree of confidence should change are framed. This language is going to involve a logic, because it has to incorporate logical connectives, quantifiers, modal operators, epistemic operators, and the like—precisely the kind of objects that the Not-So-New paradigm aims at studying (Tenenbaum et al., 2006; Stuhlmüller and Goodman, 2014). In short, the New Paradigm holds that most knowledge is probabilistic, but that probabilstic knowledge must lie on a bed of logical representations and of logical inference. So if you want a new paradigm, you'd better develop the Not-so-New paradigm along.

Given all the above, understanding how the human brain implements the elementary building blocks of human deductive competence is a fundamental goal. Neuroimaging can and has been used to inform/constrain psychological theories of deduction (see also Henson, 2005; Heit, 2015). However, Oaksford argues that many studies mistakenly understood as imaging deduction concern "elaborative, defeasible, and probabilistic reasoning", thus suggesting that imaging data do not support the existence of deduction mechanisms. We believe these criticisms underestimate the methodological and experimental progress that the neuropsychology of reasoning, inspired by the Not-So-New paradigm, has made in these last 15 years.

First, many studies already factor in the methodological criticisms raised by Oaksford. For example, it has been pointed out that specific task demands may greatly modify how participants solve deductive problems, e.g., by using analytic or heuristic processing (Reverberi et al., 2009a). The importance of choosing an adequate baseline has also been emphasized (Monti et al., 2007; Reverberi et al., 2007), or appropriate behavioral indices (Rotello and Heit, 2014). Also, recent studies consider between subject variability and try to identify fine-grained functional specializations within the network involved in deduction (e.g., Reverberi et al., 2010).

Second, recent convergent findings "deductive tasks" can be naturally interpreted within the framework of the Not-So-New paradigm:

1. The left ventro-lateral prefrontal cortex (left VLPFC, Brodmann Area 47/10) is active when participants are either evaluating or generating new deductive conclusions, both when the problems are abstract, and when they contain thematic information (Monti et al., 2007, 2009; Reverberi et al., 2010; Prado et al., 2014). Furthermore, activity in the left VLPFC predicts whether individuals tend to generate valid answers to deductive problems (Reverberi et al., 2012a), and is modulated in tasks requiring to evaluate compatibility of simple propositional sentences with evidence (Baggio et al., 2015).
2. The posterior portion of the left inferior frontal gyrus (left IFG, mostly BA44/45) is involved in inference making (Baggio et al., 2015; see also Goel et al., 2000; Reverberi et al., 2007, 2010; Prado et al., 2011), and recent studies trace its contribution to logical forms. Specifically, activity in left IFG predicts whether or not participants extract and use the formal structure of deductive problems for generating a conclusion. Importantly left IFG activation does not predict whether the generated conclusion will be valid or not, suggesting that its role is less the active process of drawing a conclusion than that of representing the logical form (Reverberi et al., 2012a). Converging evidence suggests that left IFG is devoted to computing hierarchies and relations among trees (Pallier et al., 2011). Again, these results account for individual differences, and suggest the presence of a cascade of mental representations well predicted by the Not-So-New paradigm.
3. Functional dissociations have been reported between deductive tasks of different types, such as relative and propositional reasoning (Prado et al., 2010), or conditional and categorical reasoning (Reverberi et al., 2010). Furthermore, some part of the reasoning network

(e.g., VLPFC) have been shown to dissociate "logic" from "linguistic arguments" (Monti et al., 2009).

These results prompted revisions of too-coarse-grained versions of theories of deductive reasoning (Monti et al., 2009; Reverberi et al., 2009b; Prado et al., 2010), but they also confirmed a neuroimaging approach inspired by main tenets of the Not-So-New paradigms: content can be separated from form, logical form from inference; strict predictive relations exist between patterns of brain activities and individual differences in participants' solution strategies. By contrast, we find the New Paradigm in this context predictively sterile: we fail to see what novel or different predictions it would bring about.

Perhaps future progress can be made by changing paradigm. Certainly, we agree with Oaksford and others (e.g., Heit, 2015) that the field would benefit from computational modeling, and further theoretical development. But we believe there is still much juice to be gained by squeezing the Not-So-New paradigm. The perspective of progress it offers should not be overlooked.

ACKNOWLEDGMENTS

PC and CR were supported by the PRIN grant 2010RP5RNM_001 from the Italian Ministry of University. LB was supported by MINECO grant PSI2012-31961.

REFERENCES

Baggio, G., Cherubini, P., Pischedda, D., Gorgen, K., Blumenthal, A., Haynes, J.-D., et al. (2015). *Concept Combination with Logical Connectives*. San Francisco, CA: Presented at the Cognitive Neuroscience Society.

Braine, M. D. S., and O'Brien, D. P. (1998). *Mental Logic*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Csibra, G. (2008). Goal attribution to inanimate agents by 6.5-month-old infants. *Cognition* 107, 705–717. doi: 10.1016/j.cognition.2007.08.001

Csibra, G., and Gergely, G. (2009). Natural pedagogy. *Trends Cogn. Sci.* 13, 148–153. doi: 10.1016/j.tics.2009.01.005

De Neys, W. (2012). Bias and conflict: a case for logical intuitions. *Perspect. Psychol. Sci.* 7, 28–38. doi: 10.1177/1745691611429354

De Neys, W., and Schaeken, W. (2007). When people are more logical under cognitive load: dual task impact on scalar implicature. *Exp. Psychol.* 54, 128–133. doi: 10.1027/1618-3169.54.2.128

Endress, A. D. (2013). Bayesian learning and the psychology of rule induction. *Cognition* 127, 159–176. doi: 10.1016/j.cognition.2012.11.014

Gergely, G., Bekkering, H., and Király, I. (2002). Developmental psychology: rational imitation in preverbal infants. *Nature* 415, 755–755. doi: 10.1038/415755a

Gergely, G., Nádasdy, Z., Csibra, G., and Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition* 56, 165–193. doi: 10.1016/0010-0277(95)00661-H

Goel, V., Buchel, C., Frith, C., and Dolan, R. J. (2000). Dissociation of mechanisms underlying syllogistic reasoning. *Neuroimage* 12, 504–514. doi: 10.1006/nimg.2000.0636

Goldvarg, E., and Johnson-Laird, P. N. (2001). Naive causality: a mental model theory of causal meaning and reasoning1. *Cogn. Sci.* 25, 565–610. doi: 10.1207/s15516709cog2504_3

Gweon, H., and Schulz, L. (2011). 16-month-olds rationally infer causes of failed actions. *Science* 332, 1524–1524. doi: 10.1126/science.1204493

Heit, E. (2015). Brain imaging, forward inference, and theories of reasoning. *Front. Hum. Neurosci.* 8:1056. doi: 10.3389/fnhum.2014.01056

Henson, R. (2005). What can functional neuroimaging tell the experimental psychologist? *Q. J. Exp. Psychol. A Hum. Exp. Psychol.* 58, 193–233. doi: 10.1080/02724980443000502

Johnson-Laird, P. N. (2010). Mental models and human reasoning. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18243–18250. doi: 10.1073/pnas.1012933107

Monti, M. M., Osherson, D. N., Martinez, M. J., and Parsons, L. M. (2007). Functional neuroanatomy of deductive inference: a language-independent distributed network. *Neuroimage* 37, 1005–1016. doi: 10.1016/j.neuroimage.2007.04.069

Monti, M. M., Parsons, L. M., and Osherson, D. N. (2009). The boundaries of language and thought in deductive inference. *Proc. Natl. Acad. Sci. U.S.A.* 106, 12554–12559. doi: 10.1073/pnas.0902422106

O'Brien, D. P. (1995). "Finding logic in human reasoning requires looking in the right places," in *Perspectives on Thinking and Reasoning: Essays in Honour of Peter Wason,* eds S. E. Newstead and J. S. B. T. Evans (Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.), 189–216.

Oaksford, M. (2015). Imaging deductive reasoning and the new paradigm. *Front. Hum. Neurosci.* 9:101. doi: 10.3389/fnhum.2015.00101

Pallier, C., Devauchelle, A.-D., and Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proc. Natl. Acad. Sci. U.S.A.* 108, 2522–2527. doi: 10.1073/pnas.1018711108

Prado, J., Chadha, A., and Booth, J. R. (2011). The brain network for deductive reasoning: a quantitative meta-analysis of 28 neuroimaging studies. *J. Cogn. Neurosci.* 23, 3483–3497. doi: 10.1162/jocn_a_00063

Prado, J., Spotorno, N., Koun, E., Hewitt, E., Van der Henst, J.-B., Sperber, D., et al. (2014). Neural interaction between logical reasoning and pragmatic processing in narrative discourse. *J. Cogn. Neurosci.* 27, 692–704. doi: 10.1162/jocn_a_00744

Prado, J., Van Der Henst, J. B., and Noveck, I. A. (2010). Recomposing a fragmented literature: how conditional and relational arguments engage different neural systems for deductive reasoning. *Neuroimage* 51, 1213–1221. doi: 10.1016/j.neuroimage.2010.03.026

Reverberi, C., Bonatti, L. L., Frackowiak, R. S. J., Paulesu, E., Cherubini, P., and Macaluso, E. (2012a). Large scale brain activations predict reasoning profiles. *Neuroimage* 59, 1752–1764. doi: 10.1016/j.neuroimage.2011.08.027

Reverberi, C., Cherubini, P., Frackowiak, R. S. J., Caltagirone, C., Paulesu, E., and Macaluso, E. (2010). Conditional and syllogistic deductive tasks dissociate functionally during premise integration. *Hum. Brain Mapp.* 31, 1430–1445. doi: 10.1002/hbm.20947

Reverberi, C., Cherubini, P., Rapisarda, A., Rigamonti, E., Caltagirone, C., Frackowiak, R. S. J., et al. (2007). Neural basis of generation of conclusions in elementary deduction. *Neuroimage* 38, 752–762. doi: 10.1016/j.neuroimage.2007.07.060

Reverberi, C., Pischedda, D., Burigo, M., and Cherubini, P. (2012b). Deduction without awareness. *Acta Psychol.* 139, 244–253. doi: 10.1016/j.actpsy.2011.09.011

Reverberi, C., Rusconi, P., Paulesu, E., and Cherubini, P. (2009a). Response demands and the recruitment of heuristic strategies in syllogistic reasoning. *Q. J. Exp. Psychol. (2006)* 62, 513–530. doi: 10.1080/17470210801995010

Reverberi, C., Shallice, T., D'Agostini, S., Skrap, M., and Bonatti, L. L. (2009b). Cortical bases of elementary deductive reasoning: Inference, memory, and metadeduction. *Neuropsychologia* 47, 1107–1116. doi: 10.1016/j.neuropsychologia.2009.01.004

Rotello, C. M., and Heit, E. (2014). The neural correlates of belief bias: activation in inferior frontal cortex reflects response rate differences. *Front. Hum. Neurosci.* 8:862. doi: 10.3389/fnhum.2014.00862

Southgate, V., and Csibra, G. (2009). Inferring the outcome of an ongoing novel action at 13 months. *Dev. Psychol.* 45, 1794–1798. doi: 10.1037/a0017197

Stahl, A. E., and Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science* 348, 91–94. doi: 10.1126/science.aaa3799

Stuhlmüller, A., and Goodman, N. D. (2014). Reasoning about reasoning by nested conditioning: modeling theory of mind with probabilistic programs. *Cogn. Syst. Res.* 28, 80–99. doi: 10.1016/j.cogsys.2013.07.003

Téglás, E., Girotto, V., Gonzalez, M., and Bonatti, L. L. (2007). Intuitions of probabilities shape expectations about the future at 12 months and beyond. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19156–19159. doi: 10.1073/pnas.0700271104

Téglás, E., Ibanez-Lillo, A., Costa, A., and Bonatti, L. L. (2015). Numerical representations and intuitions of probabilities at 12 months. *Dev. Sci.* 18, 183–193. doi: 10.1111/desc.12196

Téglás, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., and Bonatti, L. L. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. *Science* 332, 1054–1059. doi: 10.1126/science.1196404

Tenenbaum, J. B., Griffiths, T. L., and Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends Cogn. Sci.* 10, 309–318. doi: 10.1016/j.tics.2006.05.009

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.

frontiers in Human Neuroscience

HYPOTHESIS AND THEORY
published: 27 October 2015
doi: 10.3389/fnhum.2015.00590

Episodes, events, and models

Sangeet S. Khemlani, Anthony M. Harrison and J. Gregory Trafton*

Naval Research Laboratory, Navy Center for Applied Research in Artificial Intelligence, Washington, DC, USA

We describe a novel computational theory of how individuals segment perceptual information into representations of events. The theory is inspired by recent findings in the cognitive science and cognitive neuroscience of event segmentation. In line with recent theories, it holds that online event segmentation is automatic, and that event segmentation yields mental simulations of events. But it posits two novel principles as well: first, discrete episodic markers track perceptual and conceptual changes, and can be retrieved to construct event models. Second, the process of retrieving and reconstructing those episodic markers is constrained and prioritized. We describe a computational implementation of the theory, as well as a robotic extension of the theory that demonstrates the processes of online event segmentation and event model construction. The theory is the first unified computational account of event segmentation and temporal inference. We conclude by demonstrating now neuroimaging data can constrain and inspire the construction of process-level theories of human reasoning.

Keywords: event segmentation, temporal reasoning, mental models, episodic memory, MDS robot, ACT-R/E

OPEN ACCESS

Edited by:
Gorka Navarrete,
Universidad Diego Portales, Chile

Reviewed by:
Carlo Reverberi,
Università Milano Bicocca, Italy
Walter Schaeken,
KU Leuven, Belgium

****Correspondence:***
Sangeet S. Khemlani
sangeet.khemlani@nrl.navy.mil

Received: *30 June 2015*
Accepted: *12 October 2015*
Published: *27 October 2015*

Citation:
Khemlani SS, Harrison AM and Trafton JG (2015) Episodes, events, and models.
Front. Hum. Neurosci. 9:590.
doi: 10.3389/fnhum.2015.00590

INTRODUCTION

How do people represent and reason about time? Calendars, clocks, and timepieces come coupled with the convenient illusion of time as a collection of discrete temporal markers, such as months and minutes, which are experienced in serial order. Events, such as *breakfast* or *the birthday party*, are perceived as hierarchical organized structures relative to those markers. In extraordinary conditions of sensory deprivation—a prisoner in solitary confinement, for example –the façade of a regimented temporal hierarchy melts away to reveal the truth: time at the scale of human experience is a continuous flow of sensory information without subdivision.

Humans organize this unabating stream of sensory input into meaningful representations of episodes and events. Brain regions are sensitive to perceptually salient event boundaries (Zacks et al., 2001a), and people learn to segment continuous actions into discrete events in their infancy (Wynn, 1996). The concept of time, temporal order, and event structure develops throughout childhood (Piaget, 1927/1969; Harner, 1975; Hudson and Shapiro, 1991). By age 3, children understand the temporal order of actions and their relations to one another in a sequence of conceptually related events (Nelson and Gruendel, 1986). Adults in turn rely on complex event structures in comprehending discourse and temporal expressions (Miller and Johnson-Laird, 1976; Moens and Steedman, 1988), in remembering autobiographical episodes (Anderson and Conway, 1993), and in planning for the future (Bower, 1982). The end result of parsing the continuous stream of sensory information appears to yield event structures that take the form of a mental model, i.e., an iconic configuration of events organized around a spatial axis (Johnson-Laird, 1983; Casasanto et al., 2010; Radvansky and Zacks, 2011; Bonato et al., 2012), from which temporal relations between can be inferred (Vandierendonck and De Vooght, 1994; Schaeken et al., 1996; Gentner, 2001).

There is an intimate link between the processes of temporal inference and the way in which the brain segments events: event segmentation yields the mental representations that permit temporal reasoning. Recent research focuses on how the brain carves continuous experiences up to build discrete temporal representations. Behavioral and imaging data suggest that to construct representations of events online, individuals rapidly integrate multiple conceptual and perceptual cues—such as a movement to a new spatial location or the introduction of a new character or object into the perceiver's environment (Zacks et al., 2007). But no theory describes how cues are accessed and encoded, how they are integrated, and how they are used to build representations of events; no extant computer program can solve the task either.

To address the discrepancy, we describe a novel approach that synthesizes these various operations to yield a unified theory of event segmentation and temporal inference. We implemented the system computationally in an embodied platform that is able to process input from its sensors to build discrete model-based representations of events. The paper begins with a review of the functional neuroanatomy of the brain mechanisms underlying the integration of conceptual and perceptual cues to mark event boundaries. It then describes a theory of how processing continuous sensory information yields episodic memory representations, as well as how those memory representations are used to build event models. It presents a computational and robotic implementation of the theory, and shows how the theory provides a foundation for an account of temporal inference. Finally, it reviews the present approach as one that marshals the insights of cognitive neuroscience to advance theories of high-level inference.

EVENT SEGMENTATION IN THE BRAIN

You walk through a hallway to enter a room, where your colleague sits behind her desk. You take a seat in front of the desk and begin to converse with her. You leave the office sometime later to head to the bar across to street to meet a friend for drinks. At some point during this sequence of continuous environmental changes, a new event began: the *meeting*. At another point, it ended and a new event began. There exists no direct, observable, physical cue that marks the beginning, duration, or end of the meeting: the meeting and its extension across time has to be perceived indirectly from an integration of multiple internal and external cues (Zacks and Tversky, 2001), and the process of perception has to yield a discrete representation of a sequence of events (Radvansky and Zacks, 2011).

People can systematically parse out meaningful events by observing sequences of everyday actions (Newtson, 1973; Newtson et al., 1977). Newtson and his colleagues pioneered the study of event segmentation behavior, and posited three hypotheses on the perception of events: first, event boundaries are distinguished by a large number of distinctive changes in perceptual stimuli. Second, event boundaries are graded—some boundaries are sharp and mark distinct separations between two separate events, whereas other boundaries are fuzzier and mark less distinguished separations. Finally, events are part of a "partonomy," i.e., a part-whole hierarchy (see Cooper and Shallice, 2006; Hard et al., 2006). For example, suppose you *wash a set of dirty dishes*. That event consists of subordinate events (e.g., *wash plate 1*, *wash plate 2*, and so on) and is itself part of a larger event (e.g., *cleaning the kitchen*).

Recent neuroimaging studies concur with Newtson's proposals. Zacks and his colleagues present decisive evidence that processes governing event segmentation are unconscious, automatic, and ongoing (Zacks et al., 2001a, 2010; Speer et al., 2003). In one study, participants passively viewed sequences of everyday activities in the scanner, and then viewed the sequences again while they explicitly segmented the event boundaries (Zacks et al., 2001a). The data revealed systematic increases in BOLD response prior to points at which boundaries were identified; likewise, there was a reliable difference in activation of frontal and posterior clusters of brain regions as a function of whether participants marked fine or course boundaries in events. These two points suggest an ongoing, automatic segmentation process that integrates cues from external stimuli in the absence of conscious deliberation. A similar study by Speer et al. (2003) revealed that evoked responses in the brain's motion sensitive area (extrastriate MT+ and the area connecting left inferior frontal and precentral sulcus) occurred in temporal proximity to participants' overt segmentation behavior as they analyzed videos of action sequences. Schubotz and colleagues show that MT activation may play a more general role in segmenting ongoing activity from movements, i.e., not just for goal-directed action sequences (Schubotz et al., 2012). Participants' behavioral data likewise provide evidence for partonomic organization of event segmentation: their subjective evaluations of coarse event boundaries overlap with their evaluations of fine boundaries (see Zacks et al., 2001a). Moreover, when asked to describe events from memory, participants' responses reveal a hierarchical structure such that superordinate events are remembered and described more frequently (Zacks et al., 2001b).

Online event segmentation is not driven by visual cues alone. Speer et al. (2009) found an association between activations in regions of the brain associated with processing event boundaries and participants' identification of event boundaries in linguistic narratives. Event boundaries were distinguished by explicit changes in characters, locations, goal-directed activities, causal antecedents, and interactions with objects in the narratives (Speer et al., 2009). Other evidence reveals brain regions that subserve online event segmentation in auditory narrative comprehension (Whitney et al., 2009) and in music (Sridharan et al., 2007).

These results dovetail with other work that suggests that understanding action narratives is similar to simulating motor movements (e.g., Aziz-Zadeh et al., 2006). Aziz-Zadeh et al. show that mirror neuron areas in the premotor cortex are active both when participants passively observe action sequences as well as when they read descriptions of those same sequences. As they argue, the results support the activation of shared mental representations for conceptually interpreting language input and for perceptually processing visual input.

In sum, neural evidence corroborates three hypotheses about event segmentation:

1. Event segmentation is an ongoing, automatic process.
2. Events are segmented into discrete representations relative to a temporal partonomy, where events are embedded within other events. An additional computational constraint is that because the brain cannot represent infinite regression, the temporal partonomy must be bounded.
3. Event segmentation is driven by detecting perceptual changes in audiovisual stimuli and in conceptual changes in mental representations of discourse (but cf. Schapiro et al., 2013).

Gaps in Theories of Event Perception

It may be unimpeachable that people systematically carve continuous experience into events, and that they do so by marking boundaries between events. Many views from philosophy, neuroscience, and psychology even concur that event structures are discrete in nature (e.g., Casati and Varzi, 2008; Radvansky and Zacks, 2011; Liverence and Scholl, 2012) and some theorists posit specific ways in which those structures can be organized relative to one another (Schapiro et al., 2013). Indeed, few would argue that representations of event structure aren't critical for making inferences about temporal, spatial, and causal relations. However, consensus over matters of event cognition does not imply completeness. No extant theory of event segmentation explains how the process yields discrete event representations. Instead, many gaps in knowledge exist about how event structures come about. Three salient questions remain unanswered by theoretical and empirical investigations: First, what is the neurocognitive representation of an event boundary? It may be a discrete representation that is encoded in memory, or it may be a transient set of activations that are rapidly extinguished once a representation of an event is constructed. Second, how does the online process of event segmentation resolve multiple perceptual and conceptual segmentation cues? Some cues appear more important than others, e.g., changes in the focus of an object may be less important than changes in location, and other cues may compete with one another. Third, how does the brain recognize an event as an event? In addition to encoding an event's spatiotemporal frame, its characters, their goals, their interactions, and the objects involved, the mind needs to represent a nested structure of events within other events, and no theory at present explains what the representation looks like or what sorts of mental operations are permitted by it.

To address these three questions, we developed a novel theory of event segmentation and temporal inference. The theory builds on the idea that changes to internal and external stimuli precipitate segmentation behavior, but goes beyond it to hypothesize that segmentation is driven by the construction of episodic representations of event boundaries. Some perceptual and conceptual cues take precedence to others to yield a precedence hierarchy, and the hierarchy determines the activations of episodic representations in memory. The episodic memories in turn allow for the direct construction of mental models of temporal relations. We present the theory in the next section.

A UNIFIED THEORY OF EVENT SEGMENTATION AND REPRESENTATION

We developed a novel, model-based theory of event segmentation and event representation. The theory inverts a common strategy in understanding event segmentation: instead of considering how individuals parse a continuous stream of information into discrete temporal units, we begin with the assumption that the end result of segmentation is the construction of a temporal mental model (Johnson-Laird, 1983; Schaeken et al., 1996; Radvansky and Zacks, 2011). Craik (1943) was the first psychologist to propose that people build and interrogate small-scale models of the world around them, but philosophers before him explored analogous notions. Mental models serve as a general account of how individuals perceive the external world, how they understand linguistic assertions, how they represent them, and how they reason from them (see Johnson-Laird, 1983; Johnson-Laird and Byrne, 1991; Johnson-Laird and Khemlani, 2014). As Johnson-Laird (1983, p. 406) writes, "Mental models owe their origin to the evolution of perceptual ability in organisms with nervous systems. Indeed, perception provides us with our richest model of the world." Hence, models serve as a way to unify perceptual and linguistic processes, as they are hypothesized to be the end result of both. They are pertinent to reasoning about abstract relations, as well as relations about time and space (Goodwin and Johnson-Laird, 2005; Ragni and Knauff, 2013). The model theory depends on three foundational principles:

1. Mental models represent distinct *possibilities*: when perceiving the world and processing language, models represent a set of discrete possibilities to which the current situation or description refers. When perceiving the world, models represent a homomorphism of the sensory input, i.e., many properties of the sensory input are omitted from the model. The properties that are represented are subject to the next principle of the theory.
2. The principle of *iconicity:* a model's structure corresponds to the structure of what it represents (see Peirce, 1931–1958, Vol. 4). Events are represented as either kinematic models that unfold in time, i.e., where time is represented by time itself akin to a mental "movie" (Khemlani et al., 2013) or else as a spatial arrangement of discrete events, where time is represented along a mental time line (Schaeken et al., 1996; Bonato et al., 2012). Logical consequences emerge from the iconic properties of the models (Goodwin and Johnson-Laird, 2005) and conceptual simulations on the models (Trickett and Trafton, 2007; Khemlani et al., 2013).
3. The principle of *parsimony*: In scenarios in which discourse is consistent with multiple alternative models, people tend to construct a single mental model, which yields rapid, intuitive inferences. Provided that the inferential task is not too difficult, they may be able to construct additional alternative models from a description. However, inferences that depend on alternative models are more difficult.

Mental models account for how people reason about time. Schaeken et al. (1996) showed that reasoners are faster and make fewer errors when reasoning about descriptions consistent with just one event model than descriptions consistent with multiple models. For example, the following description is consistent with one model:

> John takes a shower before he drinks coffee.
> John drinks coffee before he eats breakfast.

The event model consistent with premises can be depicted in the following diagram:

```
shower    coffee    breakfast
```

The diagram uses linguistic tokens arranged across spatial axis that represents a mental timeline. The tokens are for convenience, but the theory postulates that people simulate the events corresponding to each token. They make inferences by scanning the iconic representation for relations. When a token is to the left of a second token on the timeline, the event to which it refers happens before the event in the second token. Hence, reasoners have little difficulty deducing that John takes a shower *before* eats breakfast from the description. They do so rapidly and make few mistakes. In contrast, the following description is consistent with multiple models:

> John takes a shower before he drinks coffee.
> John drinks coffee before he eats breakfast.

The premises are consistent with the possibility in which the coffee precedes the breakfast:

```
shower    coffee    breakfast
```

and also with the possibility in which the breakfast precedes the coffee:

```
shower    breakfast    coffee
```

Reasoners have difficulty in deducing that no relation holds of necessity between the shower, the coffee, and the breakfast. They appear to build one model of the assertions and to refrain from considering alternatives (see also Vandierendonck and De Vooght, 1994, 1997). Vandierendonck and colleagues further showed that reasoners construct initial event models relative to their background beliefs (Dierckx et al., 2004).

The model theory accordingly serves as a viable account of temporal representation and reasoning, though the theory does not explain how events are perceived in the first place. In the following sections, we posit two novel assumptions that augment previous model-based accounts. The resulting theory can cope with how people represent durations, and also how they perceive durational events online. It accordingly provides a unified account of temporal perception and inference.

Representing Duration with Models

One fundamental challenge to the theory presented above is that it does not account for how people represent and reason about events with durations. People make inferences about durations on a routine basis: if you are scheduled to take part in a meeting from 10 a.m. to 1 p.m., and a colleague asks you to join him for lunch at 12 p.m., then you must first detect the conflict and then prioritize your schedule accordingly. Hence, reasoners base their actions on understanding durations of events. While previous incarnations of the model theory have focused on punctate and not durational events, we extend the theory to deal with both. The reason is because many events can be construed in a punctual aspect, i.e., as taking place in a single moment, as well as in a durational aspect, i.e., one that describes a scenario that endures across a temporal interval (Miller and Johnson-Laird, 1976; Moens and Steedman, 1988). Consider the following examples from Miller and Johnson-Laird (1976, p. 429–431):

(a) It exploded when he arrived.
(b) It exploded while he arrived.

In (a), the sentential connective *when* ensures that the noun phrase, *he arrived*, takes on a punctual aspect. Hence, people may build a model akin to the following:

```
arrived
exploded
```

where the two events happen at same time and are therefore vertically aligned (given a horizontal axis representing time). In (b), the connective *while* confers a durational aspect, and so people may directly represent the duration in their mental model, e.g.:

```
[   arrived   ]
exploded
```

where the brackets denote that the arrival is extended across several time points. As both punctate and durational events are pervasive in daily life, a rich account of temporal reasoning must explain how both types of events are represented and interrogated.

Durational events play an essential role in event perception. Events are almost always perceived across a temporal interval. If, as most theories of segmentation posit, people use environmental changes to mark the beginnings and endings of events, then events must extend across multiple moments in time for those changes to be registered. It may be that events are perceived at first as being durational in nature, and coalesce later into punctate moments only after being encoded in memory. Exceptions exist: the moment of birth, the moment of death, and winning the lottery may be perceived as a single moment in time. But many events are compiled into punctate representations only under retrospective analysis. The process of segmenting events assumes that segmentation is necessary to begin with, and hence, that most events subject to direct perception have duration.

An initial step to a unified theory of event segmentation and temporal inference is accordingly to explain how durations are represented in models. Models concern discrete possibilities; the theory eschews the representation of infinite sequences, and so metric information is difficult to represent with models of possibilities. One challenge is accordingly to describe a method by which durations are represented discretely. Recent work in cognitive neuroscience may provide insight into the nature of the representation. Research on rats reveals specific hippocampal neurons that fire reliably at particular moments in event sequences. These so-called "time cells" encode the event for later retrieval, as well as episodic information such as where the event takes place (MacDonald et al., 2011). Studies on adults corroborate the essential role of the hippocampus in encoding event sequences, encoding episodic information, and bridging temporal gaps between discontiguous events (Kumaran and Maguire, 2006; Lehn et al., 2009; Ross et al., 2009; Staresina and Davachi, 2009; Hales and Brewer, 2010). Ezzyat and Davachi (2011) show that event boundaries are used to bind episodic information to event representations; more generally, they posit a critical role of episodic memory in event perception. In a similar vein, Baguley and Payne (2000) present evidence that people encode episodic traces in memory, and use those traces to build event models from temporal descriptions.

We accordingly introduce the following principle about the representation of durations:

> *The principle of discrete episodes:* Reasoners represent durational events by constructing discrete episode markers as chunks in episodic memory. Episode markers represent perceived changes in goals, locations, individuals, and objects. Markers are retrieved to construct durational mental models in which one marker represents the start of an event and another marker represents its end.

The principle of discrete episodes has implications for both event segmentation and mental model construction. According to the principle, when an event boundary is identified during online event segmentation, an episode marker is constructed. The event boundary may be triggered by multiple perceptual or conceptual cues; those cues are encoded in the representation of the marker (cf. Ezzyat and Davachi, 2011). For example, consider the scenario introduced in Section Event Segmentation in the Brain of a meeting with your colleague. The meeting might begin when you enter your colleague's office. Many changes occur the moment you enter: a change in location, the introduction of a salient individual to the environment (your colleague), the start of a goal (holding the meeting), and the introduction of a salient object (e.g., a printout of data). A single episodic marker encodes all of the detected changes: the location, the individual, the goal, and the object. When the meeting ends and you leave the office, there is a change in location, which may precipitate the construction of another episodic marker. Other things may or may not change; for example, if your colleague walks with you back to your office with the printout in hand, no character- or object-based changes would be encoded.

The principle posits that episodic markers are encoded as chunks in episodic memory (Altmann and Trafton, 2002, p. 40). As such, they are highly active when they are first constructed, but memory for them gradually fades. Markers that encode many perceptual and conceptual changes start with higher activations than markers that track fewer changes. Episodic markers are maintained in long-term memory (cf. Baguley and Payne, 2000), and when they are retrieved, their activation spikes and spreads to activate associated markers, i.e., those within the same temporal context and those that track the same sorts of perceptual and conceptual changes.

Episodic markers, by definition, encode punctate episodes. They can also be used retrospectively to construct discrete representations of events, i.e., durational event models. A memory of "the meeting" would accordingly consist of two separate markers as follows:

$$\text{meeting}_{\text{START}} \qquad \text{meeting}_{\text{END}}$$

The markers may encode disparate sets of information. The start and end of a meeting may be cued by perceptual changes in location, for example, whereas the start and end of a bike ride concerns the conceptual introduction and completion of a goal (We address this issue in a thoroughgoing way in the next section). In either case, episodic markers can be used to build event models. Such models can be hierarchically organized:

$$\begin{array}{ccccccc}
\text{day}_{\text{START}} & & & & & & \text{day}_{\text{END}} \\
 & \text{meeting}_{\text{START}} & \text{meeting}_{\text{END}} & & & & \\
 & & & \text{evening}_{\text{START}} & & \text{evening}_{\text{END}} & \\
 & & & & \text{dinner} & & \\
1 & 2 & 3 & 4 & 5 & 6 & 7
\end{array}$$

In the model above, each line represents a distinct event. The model depicts a punctate event (dinner) represented within a durational event (the evening). The dinner may be conceived as durational as well, but at the bottom of the hierarchy, non-intersecting durational events are functionally equivalent to punctate events. The model is iconic and its components are discrete, i.e., it does not maintain any metric information by default, such as how many minutes the "day" event endured or how many hours the "morning" event endured; hence, people can reason about events whose durations outlast lifetimes (e.g., epochs and eons). Humans and other animals use other neural mechanisms to track and represent metric information about duration (see Allman et al., 2014, for a review). The numbers represent individual episode markers, e.g., 3 represents the episode marker that encodes the cues used to mark the end of the meeting. It is also a parsimonious representation from which to make temporal inferences. For example, the model above can be used to infer the following temporal relations:

- The dinner did not occur during the meeting.
- The meeting occurred before the evening.
- The dinner happened during the day.

Hence, relations concerning relative duration and other temporal relations can be drawn from models that maintain only discrete representations. The principle of discrete episodes posits that episode markers are used to construct events dynamically and to retrospectively build representations of events from memory or linguistic descriptions.

Constructing Models Dynamically from Episodic Information

According to the principle of discrete episodes, episode markers encode perceived changes in goals, locations, and other salient conceptual and perceptual information. But how can the system use the information encoded within an episode marker to rapidly construct event models dynamically, even as new markers are being encoded? The problem is acute because the cues used to mark the beginning of an event may not be relevant in marking the end of an event. The process of interrogating all of the information encoded by an episodic marker is cognitively implausible on account of the combinatorial explosion inherent in assessing and integrating multiple types of properties. The theory accordingly posits a more rapid procedure:

> *The principle of event prioritization:* Events are associated with a single perceptual or conceptual element whose change denotes the beginning and end of the event. Changes in elements are prioritized with respect to a given context: by default, goal events are the highest priority as they override events based on perceptual changes. When a goal is active, perceptual changes do not yield episode markers outside the context of the goal. Perceptual changes are likewise ranked in order of priority based on the ease of detecting a change: location events override events based on individuals, which in turn override those based on objects in the environment.

One way of construing the principle of event prioritization is that an ongoing event completes only when elements of the highest pertinent priority change. Recent work uncovers evidence for the prioritization and ordering of rule sets (Reverberi et al., 2012), and we extend the general idea to focus on event perception. In what follows, we describe how the principle operates for four primary sorts of conceptual and environmental changes: goals, locations, individuals, and objects.

Goals

The principle posits that goal-directed events are of utmost importance. Here we speak of goals in a narrow sense: goals are mental states that govern immediate, short-term, and ongoing sequences of actions that bring about a desired state of affairs in the world. Hence, goal-directed actions are those that subserve the completion of the goal. Life goals, career goals, and romantic goals are outside the scope of our present analysis because they do not govern immediate, short-term sequences. Many seminal studies on event representations address the integral involvement of goals in the way events are encoded, retrieved, and reconstructed (Lichtenstein and Brewer, 1980; Brewer and Dupree, 1983; Travis, 1997). Goals are of highest importance because they provide a top-down structure on event segmentation based on perceptual changes. An example of a sort of goal that falls within the purview of the principle of event prioritization is the goal to walk across town to meet a friend for a drink at a prearranged time. The goal-based event (walking across town) continues until the goal is completed. While episodic markers are constructed as the event proceeds, the perceived event remains organized relative to the goal and not on any other perceptual experience, such as the perception of changes in locations or individuals in the environment. Hence, external cues that would otherwise signal the beginning of a new event—such as a change in location—would instead signal the beginning of a new subevent organized within the context of the goal-based event.

Locations

Locations serve to organize multiple perceptual stimuli. As with the time cells discussed above, animals and people have dedicated hippocampal "place cells" that encode location information (see Moser et al., 2008, for a review). A behavioral demonstration of their importance is evident in studies by Radvansky and Copeland (2006) and Radvansky et al. (2010). They show that memory for objects drops when individuals move through a doorway from one location to another in a virtual reality environment, and explain the effect as a dynamic update to an event model. The principle of event prioritization posits that locations govern the perception of an event when a high-level goal stays constant and ongoing, or is absent altogether. Locations are also more stable than other sorts of perceptual stimuli because locations generally do not change relative to another individual's agency, whereas other sorts of perceptual cues (the individuals in the environment and the objects they interact with) do change relative to agency. We discuss them next.

Characters and Objects

Characters and objects in an environment serve as low-level perceptual cues for the dynamic construction of events in the absence of both goal- and location-based cues. When individuals have no goal to govern their actions and their locations do not change for a long period of time (e.g., when traveling on an airplane for several hours), the principle of event prioritization posits that dynamic events are constructed relative to detecting changes based on interaction, i.e., changes in individuals and changes in objects to which the perceived attends. One motivation for the deference of character- and object-based cues to goal- and location-based cues is that the former two can change rapidly, and it requires computational resources to track those changes and use them to update event models. Another motivation comes from evidence from Zacks et al. (2001b): they asked participants to describe units of activity as they identified them in an event segmentation task with instructions to mark events using a fine-grain or a coarse-grain. Participants described objects more often using fine-grain descriptions, and they used a broader variety of words to describe objects for fine-grained descriptions. These data suggest that people track objects more frequently when locations and goals do not change. The principle of event prioritization predicts that they may forget objects as locations change, in line with the results from Radvansky et al. (2010).

Summary

The unified theory of event segmentation and event representation that we posit is based on the assumption that segmentation yields and reasoning relies on mental models of temporal relations. Previous model-based accounts could not explain how durations were represented or how models were constructed dynamically, and so our unified account includes two novel assumptions: first, people track changes in their environment by automatically constructing discrete units of episodic memory, i.e., episode markers; and second, people dynamically construct events by prioritizing some cues over others. A summary of the theory is provided in **Figure 1**. To test

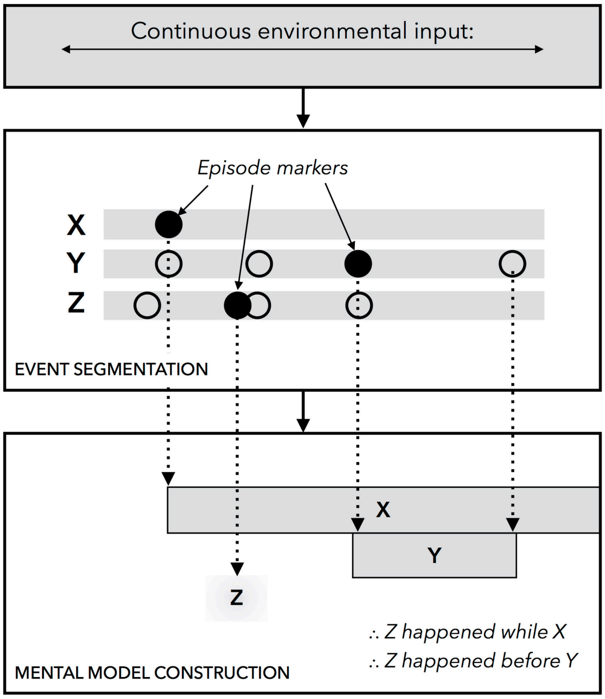


FIGURE 1 | A diagram of the unified theory of event segmentation and representation. In the event segmentation component of the system, which operates online and in parallel with other cognitive processes, changes are detected in continuous environmental input across a finite set of perceptual stimuli, marked by X, Y, and Z in the diagram. At the onset of a stimulus, which is indicated by a black circle, a new episodic marker is constructed. The offset of a stimulus likewise yields a new episodic marker. When the system is queried for information pertaining to temporal relationships, it uses the markers to build a discrete event model. The system then scans the model to make inferences.

the viability of the account, we turn next to describe its embodied computational implementation.

An Embodied Implementation of the Unified Theory

We developed an embodied, robotic implementation of the theory described in the previous section. The unorthodox approach is a result of the multifaceted nature of the tasks under investigation. The approach may be highly relevant for roboticists, because many robotic systems lack the ability to perceive and construct representations of events (Zacks, 2005; Maniadakis and Trahanias, 2011). But our goal is different. We argue that an embodied demonstration of the theory at work can help identify the types of information needed for the algorithms at each stage of the theory. A viable theory of event segmentation is one that integrates multiple perceptual and conceptual cognitive processes such as goal maintenance, location detection, person identification, and object recognition, and only a working system that integrates these perceptual processes sufficiently constrain and inform the implementational details of the theory we developed. Recent work in our laboratory has focused on each of these constituent perceptual processes: we have developed an embodied robotic platform capable of fiducial-based location tracking (see Kato and Billinghurst, 1999), person identification through face recognition (Kamgar-Parsi and Lawson, 2011) and soft biometrics (i.e., clothing, complexion, and height cues; Martinson et al., 2013) and context-sensitive object detection (Lawson et al., 2014). The platform's sensors and perceptual subsystems are interfaced with ACT-R/E, an embodied cognitive architecture for human-robot interaction (Trafton et al., 2013) based on ACT-R, a hybrid symbolic/subsymbolic production-based system for mental processing (Anderson, 2007). The system comes with multiple interoperating modules that are designed to deal with different sorts of inputs and memory representations called "chunks." Modules make chunks available through a capacity-limited buffer. Modules and buffers are mapped to the functional operation of distinct cortical regions. ACT-R/E builds on the ACT-R theory in that it can parse environmental input from perceptual systems, which is translated into chunks in a long-term memory store (the "E" stands for "embodied"). ACT-R/E is also interfaced with robotic sensors and effectors, and so it can act on the physical world. A summary of the system's sensors and its cognitive architecture is provided in **Figure 2**. We briefly review how the system implements event segmentation and the construction of event models.

Online Episodic Segmentation

The principle of discrete episodes posits that at the lowest level, an agent's experience is carved up into discrete windows of time by the encoding of episodic markers. As an agent's goals, locations, and observations of objects and people change, new episodic markers are encoded and annotated with the type of change (e.g., a change in *location*) and the contents of the change (e.g., *entered location-b*). The markers do not represent temporal durations, but rather single points in time. Encoding happens automatically as a natural consequence of attending to the environment. In the ACT-R/E cognitive architecture (Trafton et al., 2013) when the computational implementation attends to a new goal, a representation of that goal is placed within the system's goal buffer. The system monitors the buffers of relevance (i.e., the *goal buffer* for goal changes, the *configural buffer* for location changes, and the *visual buffer* for people and objects; see **Figure 2**). It creates a new episodic marker when a change in content is detected (Altmann and Trafton, 2002; Trafton et al., 2011). Each episode is symbolically annotated with information regarding environmental changes. It is also associatively linked to the prior and new contents, as well as the prior episode marker. Linking the markers in this way permits subsequent retrievals to iterate through episodes and their associated contents.

Figure 3 provides a detailed trace of the creation of discrete episodic markers. At the top of the figure is an activity trace for an individual patrolling an area. When the goal of patrolling is assigned (by, e.g., verbally issuing the directive to patrol the area), a change of goal is detected and an episodic marker (Ep-1) is encoded, and linked with the encoded goal. As the agent proceeds through the task, it encounters new locations. For each change of location, a new episodic marker is encoded (Ep-2, Ep-3, Ep-4), and populated with details regarding the changes

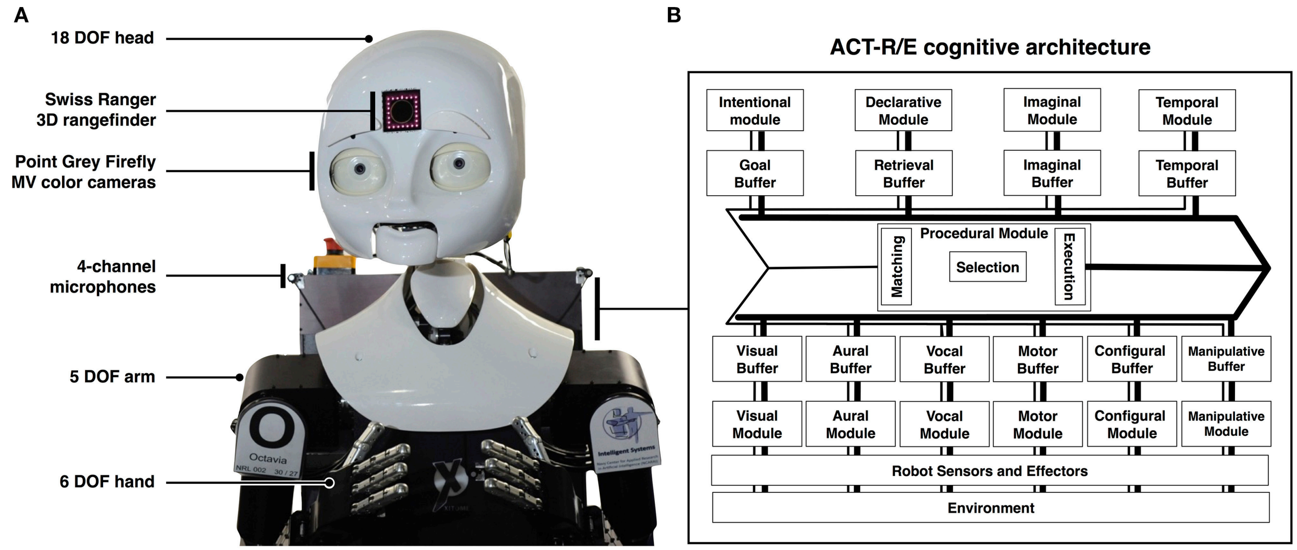


FIGURE 2 | The robotic implementation of the ACT-R/E cognitive. (A) depicts the MDS (mobile, dexterous, social) robot in use in our lab, and shows its various sensors and effectors. **(B)** provides the details of the ACT-R/E cognitive architecture (Trafton et al., 2013). The architecture is an *embodied* extension of ACT-R (Anderson, 2007), and it interfaces the robot's sensory apparatus. ACT-R/E is composed of multiple modules that mimic components of human cognition. For example, it includes modules for maintaining goals, storing declarative memories, processing visual, and auditory input, and issuing motor commands. Each module is paired with a buffer that limits the capacity that the system can process at once, and accordingly implements a processing bottleneck characteristic of human cognition. Computational implementations of cognitive processes, such as the event segmentation system we present, are developed in ACT-R/E by constructing procedural memory representations that are executed under pre-specified conditions, and which retrieve information from or else modify the contents of the system's various buffers. In the diagram, the thin lines depict the pipeline for retrieval from the contents of the buffers and the thick lines depict the pipeline for modifying the contents of the buffers.

in location, as well as the prior episodes. At one point, the agent encounters a new individual (e.g., Bob). It encodes one episodic marker to capture Bob's arrival, and another to capture Bob's departure. Once the patrolling goal is accomplished, a new marker is encoded. In line with extant theories of event segmentation, the process of encoding events is continuous. As the agent moves on to other tasks, more episodic markers are created and stored in memory.

To perceive an event as an event, the system must retrieve the markers in memory and use them to retrospectively construct an event model. We turn to this procedure.

Event Model Construction

Event segmentation occurs on an ongoing basis by default, i.e., episodic markers are encoded online. In contrast, event models are only constructed retrospectively, as a result of an external query. It is from these models that people make inferences about temporal matters. For example, the user can query the system to remember a particular location, or to infer a particular relation that holds between events, or to describe the events that occurred in a given time window. Retrospective construction is highly relevant when the system needs to make inferences about its recent experiences. For example, if the system is directed to perform a particular goal—as in the patrol example above—then it will have two separate episodic markers that highlight the start of a new goal and its completion, along with any associated environmental information that the system can detect. Now suppose that during the course of the goal, the system traveled to two separate locations. That means that the system will construct at least four separate episodic markers:

1. A marker representing the start of a new goal.
2. A marker representing the detection of a new location (*location 1*) as well as the current goal.
3. A marker representing the detection of a new location (*location 2*) as well as the current goal.
4. A marker representing the satisfaction of the goal.

These four markers will be represented in long-term memory. When the system is prompted to recall information about the particular goal, it can retrieve all four markers. It parses markers (1) and (2) to build a model of a goal's duration:

$\text{goal}_{\text{START}}$ goal_{END}

Information provided from markers (2) and (3) allow for the construction of the durational event marking location 1:

$\text{goal}_{\text{START}}$ goal_{END}
$\text{location1}_{\text{START}}$ $\text{location1}_{\text{END}}$

and information provided from markers (3) and (4) allow for the construction of the durational event marking location 2:

$\text{goal}_{\text{START}}$ goal_{END}
$\text{location1}_{\text{START}}$ $\text{location1}_{\text{END}}$
$\text{location2}_{\text{START}}$ $\text{location2}_{\text{END}}$

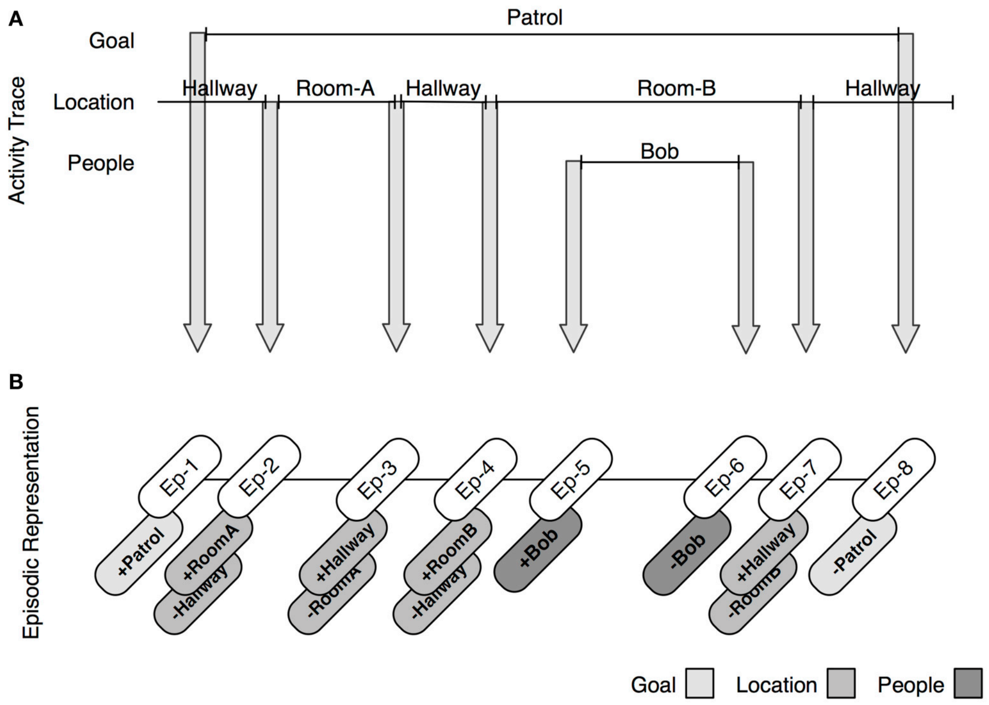


FIGURE 3 | The process by which episodes are encoded at event boundaries. (A) Shows a diagram of a trace of activity as a function of changes in goals, locations, and people. At each change, a new episodic marker is constructed (depicted as arrows). **(B)** Shows the representation of each episodic marker. Episodes are linked with symbolic information that describes the perceived changes at the time of encoding. Hence, episodes are used to uniquely describe a change in goal, location, person, and object (not depicted).

Hence, a complete event model of the relevant experiences is represented in the following mental model:

$$\begin{array}{llll} \text{goal}_{\text{START}} & & & \text{goal}_{\text{END}} \\ & \text{location1}_{\text{START}} & \text{location1}_{\text{END}} & \\ & & \text{location2}_{\text{START}} & \text{location2}_{\text{END}} \end{array}$$

From the model above, individuals can draw deductions concerning event relations, such as that visiting location 1 occurred during the goal, and the visit to location 1 occurred before the visit to location 2. The model can be revised and modified, in which case inferences would be counterfactual (Byrne, 2005). For example, reasoners can modify the event model to move the duration of the visit to location 1 *after* the visit to location 2. If no other changes are made to the model, then the reasoner might make the following counterfactual conclusion: *if the visit to location 1 had happened after the visit to location 2, then it would not have happened while the system was completing the goal*. In sum, episodic chunks can be used to build complex event models from memories. Scanning and revising the models accordingly serves as the basis of temporal reasoning.

The basic process for constructing an event model is illustrated in **Figure 4**. At the top of the figure is the episodic representation that was built in the patrolling example above (**Figure 3B**). The system constructs an event model by retrieving the earliest relevant episodic marker (e.g., Ep-1) and checking how it was triggered (e.g., goal change). From this information, a provisional event encoding is created and associated with content regarding the type and trigger for the event (e.g., a goal change initiated by following a command to patrol a given area). This information is retained until a compatible episodic marker (e.g., Ep-8) is retrieved, marking the end of the event and committing it to the event model. Each episode is retrieved and processed until there are no more markers, or some temporal limit is reached.

The process is able to produce veridical event models, such as that seen in **Figure 4B**: a veridical event model is a one-to-one mapping of marker pairs and events. Humans are unlikely to generate such complex and complete event models, particularly over long periods of time. Instead event models are influenced by the goals that triggered the retrospective construction in the first place. The principle of event prioritization constrains the construction of episodic marker types. By default, this prioritization is (from highest to lowest priority): goal, location, person, and object. During reconstruction, lower prioritized events are only encoded when they fall within the bounds of higher prioritized events. In this way, an implicit sub-event model structure can be reconstructed. **Figure 4C** shows the prioritized event model, which only represents the superordinate event, i.e., the event that characterizes the goal of patrolling an area. The principle of event prioritization, while specifying a default prioritization, does not exclude the possibility that other retrospective tasks could require other prioritizations. User queries may demand some information over others and prioritize, e.g., locations to be retrieved. The system supports the construction of partial, incremental event models.

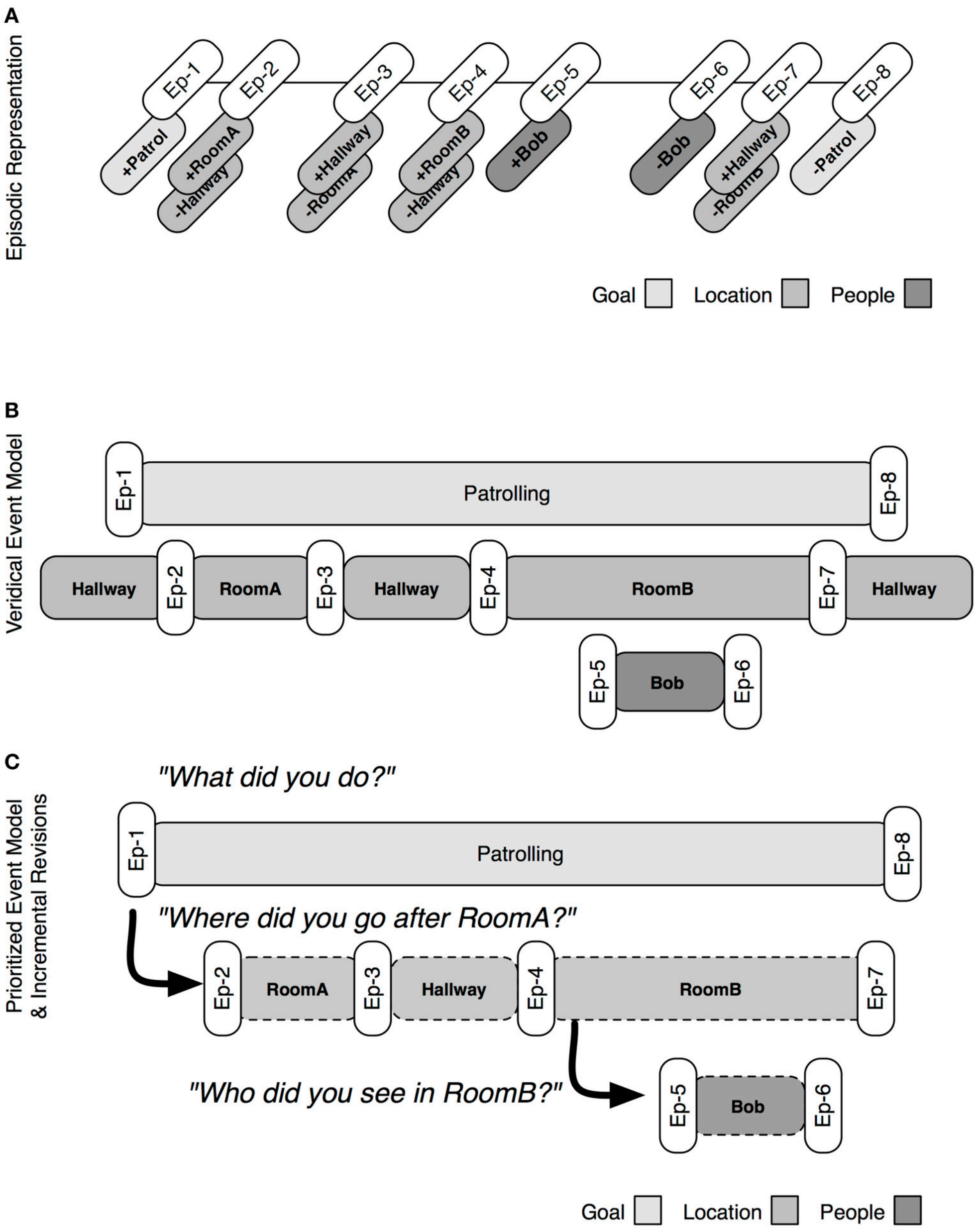


FIGURE 4 | The process by episodic markers are retrieved to build event models. (A) Shows the episodic representation (see also **Figure 3B**). **(B)** Shows a veridical event model that can be constructed by an unprioritized mapping from episodic markers to model structures. **(C)** Shows a prioritized mapping, in which the construction of a goal event takes precedence to that of other sorts of events. Additional queries can be used to revise and flesh out the prioritized event model.

A demonstration of the system for event segmentation and model construction as it occurs online is available in the **Video 1**.

GENERAL DISCUSSION

We describe a unified synthesis of event segmentation and temporal reasoning. Researchers typically focus on one process or the other. In our treatment, both are organized around the construction of discrete temporal mental models (i.e., event models). Models serve as the output of the event segmentation and the basis of temporal inference. Event segmentation is relevant in the online perception of events. Humans are capable of applying a regimented hierarchy to the continuous stream of sensory input they receive, and do so automatically and without difficulty. Yet no current theory of event segmentation or computer algorithm explains how different pieces of environmental input are used to regiment the stream of input. We accordingly developed an algorithm based on two overarching principles: (i) individuals represent events by constructing markers that track perceived changes in goals, locations, individuals, and objects; and (ii) episodic markers are constructed based on a prioritization hierarchy, in which changes in goals take precedence to changes in location, and changes in location take precedence to changes in characters and objects. The theory provides a plausible mechanism for temporal reasoning. The account thus unifies temporal cognition from how time is perceived to how temporal relations are inferred. The two principles upon which the account is based are simulated in a computational implementation of the theory, and on a robotic platform that demonstrates the viability of the hypotheses are guiding online perceptual input.

In addition to advancing temporal cognition, our theory is grounded in systematic evidence from cognitive neuroscience. The approach demonstrates a central role for neuroscientific research in the development of cognitive theory. We conclude by discussing a recent controversy on the role of cognitive neuroscience in developing and testing psychological theories of reasoning.

A central and irreproachable result from recent studies of the neuroscience of deductive inference may be that it is not modular: it implicates large swathes of the brain. A given experiment can show activation in various configurations of the basal ganglia, cerebellum, and occipital, parietal, temporal, and frontal lobes (Goel, 2007; Prado et al., 2011). Different sorts of inference recruit different brain regions (e.g., Waechter and Goel, 2005; Kroger et al., 2008; Monti et al., 2009), and a recent meta-analysis of 28 neuroimaging studies revealed systematic consistency in those regional activations for relational, quantificational, and sentential inferences (Prado et al., 2011).

Despite evidence of systematicity, many skeptics question if neuroimaging data can ever help adjudicate between theories of cognitive operations (Harley, 2004; Coltheart, 2006; Uttal, 2011). The problem is acute for students of reasoning: in order to make use of the available data, predictions of functional neuroanatomy are coaxed from psychological proposals. Most cognitive accounts of inference make no strong claims about functional neuroanatomy (Heit, 2015), i.e., they make no claims at the "implementation level" of inference (see Marr, 1982). Hence, coaxing predictions about implementation from accounts that specify only the mathematical functions to be computed for reasoning, or else the representations and algorithms that underlie reasoning, has the insidious effect of washing away theoretical nuances (Goel, 2007). Many imaging studies test the extreme view that the biological implementation of inferential procedures should rely on only one sort of mental representation, which has a distinct neural signature. The preponderance of evidence conflicts with such a view (Prado et al., 2011), which is fortunate, because the present authors know of no author or theory that defends it. And as Oaksford (2015) observes, constraints on the methodology itself may prevent diagnostic analyses. Researchers accordingly face a methodological quandary: Is it possible to marshal insights from cognitive neuroscience to inform theories of reasoning when those theories fail to make predictions of neural mechanism?

Our present approach demonstrates that it is indeed possible for theories of inferences to be informed by insights from cognitive neuroscience. As in previous work on developing an embodied theory of spatial cognition (Trafton and Harrison, 2011), we describe an embodied theory of temporal cognition whose fundamental assumptions are informed and constrained by recent work on the neuroscience of temporal processing. Cognitive neuroscience may be in its infancy, and likewise, theories of inference do not make predictions that can be tested by the imaging methodologies. Nevertheless, results from imaging studies rule out certain sorts of representations and provide mechanistic constraints on how humans may engage in particular cognitive tasks. The preceding discussion serves as a case study in how neuroimaging results can serve to guide and constrain the development of theories at Marr's "algorithmic level," which focuses on cognitive representations and processes upon those representations.

In particular, the representations we proposed in the present theory—episodic markers and event models—are supported by work on how event segmentation is carried out by the brain. Likewise, the procedures we posit, including the hypothesis that people prioritize certain changes in the environment over others, are guided by both behavioral and imaging work on mental processes that track ongoing changes in the environment. Hence, cognitive neuroscience can play a pivotal role in the development and enrichment of cognitive theories of reasoning: imaging research can serve to rule out representations that cannot be feasibly processed by complementary neural processes, and it can suggest the need for alternative representations.

The skeptics may ultimately have purchase: no psychological theory of reasoning can be said to be testable by means of neuroscientific data unless that theory makes specific predictions of neural processes. A first step toward such a theory for any domain of cognition is to provide a unified account of that domain that explains how low-level perception leads to high-level inference. In the case of temporal cognition, we provide such an account, and explain how events are perceived to build mental simulations of their temporal experience, and how reasoners make temporal inferences from those simulations.

ACKNOWLEDGMENTS

We are grateful to Bill Adams, Paul Bello, Magda Bugajska, Dan Gartenberg, Laura Hiatt, Joe Kreke, Ed Lawson, Priya Narayanan, Frank Tamborello, and Alan Schultz for their helpful comments. This work was supported by a Jerome and Isabella Karle Fellowship from the Naval Research Laboratory (to SK) and by a grant from the Office of Naval Research (to JT).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fnhum.2015.00590

Video 1 | A demonstration of the online process of perceptual segmentation, episodic marker encoding, and event model construction on an embodied robotic platform.

REFERENCES

Allman, M. J., Teki, S., Griffiths, T. D., and Meck, W. H. (2014). Properties of the internal clock: first-and second-order principles of subjective time. *Annu. Rev. Psychol.* 65, 743–771. doi: 10.1146/annurev-psych-010213-115117

Altmann, E., and Trafton, J. G. (2002). Memory for goals: an activation-based model. *Cogn. Sci.* 26, 39–83. doi: 10.1207/s15516709cog2601_2

Anderson, J. R. (2007). *How Can the Human Mind Occur in the Physical Universe?* New York, NY: Oxford University Press.

Anderson, S. J., and Conway, M. A. (1993). Investigating the structure of autobiographical memories. *J. Exp. Psychol. Learn. Mem. Cogn.* 19, 1178–1196. doi: 10.1037/0278-7393.19.5.1178

Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., and Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Curr. Biol.* 16, 1–6. doi: 10.1016/j.cub.2006.07.060

Baguley, T., and Payne, S. (2000). Long-term memory for spatial and temporal mental models includes construction processes and model structure. *Q. J. Exp. Psychol. A* 53, 479–512. doi: 10.1080/713755888

Bonato, M., Zorzi, M., and Umiltà, C. (2012). When time is space: evidence for a mental time line. *Neurosci. Biobehav. Rev.* 36, 2257–2273. doi: 10.1016/j.neubiorev.2012.08.007

Bower, G. (1982). "Plans and goals in understanding episodes," in *Discourse processing*, eds A. Flammer and W. Kintsch (Amsterdam: North-Holland Publishing Company), 2–15.

Brewer, W. F., and Dupree, D. A. (1983). Use of plan schemata in recall and recognition of goal-directed actions. *J. Exp. Psychol. Learn. Mem. Cogn.* 9, 117–129. doi: 10.1037/0278-7393.9.1.117

Byrne, R. M. J. (2005). *The Rational Imagination: How People Create Alternatives to Reality*. Cambridge, MA: MIT Press.

Casasanto, D., Fotakopoulou, O., and Boroditsky, L. (2010). Space and time in the child's mind: evidence for a cross-dimensional asymmetry. *Cogn. Sci.* 34, 387–405. doi: 10.1111/j.1551-6709.2010.01094.x

Casati, R., and Varzi, A. C. (2008). "Event concepts" in *Understanding Events: From Perception to Action*, eds T. F. Shipley and J. Zacks (New York, NY: Oxford University Press), 31–54.

Coltheart, M. (2006). What has functional neuroimaging told us about the mind (so far)? *Cortex* 42, 323–331. doi: 10.1016/S0010-9452(08)70358-7

Cooper, R. P., and Shallice, T. (2006). Hierarchical schemas and goals in the control of sequential behavior. *Psychol. Rev.* 113, 831–887. doi: 10.1037/0033-295X.113.4.887

Craik, K. (1943). *The Nature of Explanation.* Cambridge, UK: Cambridge University Press.

Dierckx, V., Vandierendonck, A., Liefhooge, B., and Christiaens, E. (2004). Plugging a tooth before anaesthetising the patient? The influence of people's beliefs on reasoning about the temporal order of actions. *Think. Reason.* 10, 371–404. doi: 10.1080/13546780442000132

Ezzyat, Y., and Davachi, L. (2011). What constitutes an episode in episodic memory? *Psychol. Sci.* 22, 243–252. doi: 10.1177/0956797610393742

Gentner, D. (2001). "Spatial metaphors in temporal reasoning," in *Spatial Schemas in Abstract Thought*, ed M. Gattis (Cambridge, MA: MIT Press), 203–222.

Goel, V. (2007). Anatomy of deductive reasoning. *Trends Cogn. Sci.* 11, 435–441. doi: 10.1016/j.tics.2007.09.003

Goodwin, G. P., and Johnson-Laird, P. N. (2005). Reasoning about relations. *Psychol. Rev.* 112, 468–493. doi: 10.1037/0033-295X.112.2.468

Hales, J. B., and Brewer, J. B. (2010). Activity in the hippocampus and neocortical working memory regions predicts successful associative memory for temporally discontiguous events. *Neuropsychologia* 48, 3351–3359. doi: 10.1016/j.neuropsychologia.2010.07.025

Hard, B. M., Tversky, B., and Lang, D. S. (2006). Making sense of abstract events: building event schemas. *Mem. Cogn.* 34, 1221–1235. doi: 10.3758/BF03193267

Harley, T. A. (2004). Does cognitive neuropsychology have a future? *Cogn. Neuropsychol.* 21, 3–16. doi: 10.1080/02643290342000131

Harner, L. (1975). Yesterday and tomorrow: development of early understanding of the terms. *Dev. Psychol.* 11, 864–865. doi: 10.1037/0012-1649.11.6.864

Heit, E. (2015). Brain imaging, forward inference, and theories of reasoning. *Front. Hum. Neurosci.* 8:01056. doi: 10.3389/fnhum.2014.01056

Hudson, J. A., and Shapiro, L. R. (1991). "From knowing to telling: the development of children's scripts, stories, and personal narratives," in *Developing Narrative Structure,* eds A. McCabe and C. Peterson (Washington, DC: American Psychological Association), 89–136.

Johnson-Laird, P. N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness.* Cambridge, MA: Harvard University Press.

Johnson-Laird, P. N., and Byrne, R. M. J. (1991). *Deduction.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Johnson-Laird, P. N., and Khemlani, S. (2014). "Toward a unified theory of reasoning," in *The Psychology of Learning and Motivation,* ed B. H. Ross (Academic Press), 1–42.

Kamgar-Parsi, B., and Lawson, W. (2011). Toward development of a face recognition system for watchlist surveillance. *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 1925–1937. doi: 10.1109/TPAMI.2011.68

Kato, H., and Billinghurst, M. (1999). "Marker tracking and HMD calibration for a video-based augmented reality conferencing system," in *Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality*. IEEE, 85–94.

Khemlani, S. S., Mackiewicz, R., Bucciarelli, M., and Johnson-Laird, P. N. (2013). Kinematic mental simulations in abduction and deduction. *Proc. Natl. Acad. Sci. U.S.A.* 110, 16766–16771. doi: 10.1073/pnas.1316275110

Kroger, J. K., Nystrom, L. E., Cohen, J. D., and Johnson-Laird, P. N. (2008). Distinct neural substrates for deductive and mathematical processing. *Brain Res.* 1243, 86–103. doi: 10.1016/j.brainres.2008.07.128

Kumaran, D., and Maguire, E. A. (2006). The dynamics of hippocampal activation during encoding of overlapping sequences. *Neuron* 49, 617–629. doi: 10.1016/j.neuron.2005.12.024

Lawson, W., Hiatt, L., and Trafton, J. G. (2014). "Leveraging cognitive context for object recognition," in *Computer Vision and Pattern Recognition Workshops (CVPRW).*

Lehn, H., Steffenach, H. A., van Strien, N. M., Veltman, D. J., Witter, M. P., and Håberg, A. K. (2009). A specific role of the human hippocampus in recall of temporal sequences. *J. Neurosci.* 29, 3475–3484. doi: 10.1523/JNEUROSCI.5370-08.2009

Lichtenstein, E. H., and Brewer, W. F. (1980). Memory for goal-directed events. *Cogn. Psychol.* 12, 412–445. doi: 10.1016/0010-0285(80)90015-8

Liverence, B. M., and Scholl, B. J. (2012). Discrete events as units of perceived time. *J. Exp. Psychol. Hum. Percept. Perform.* 38, 549–554. doi: 10.1037/a0027228

MacDonald, C. J., Lepage, K. Q., Eden, U. T., and Eichenbaum, H. (2011). Hippocampal "time cells" bridge the gap in memory for discontiguous events. *Neuron* 71, 737–749. doi: 10.1016/j.neuron.2011.07.012

Maniadakis, M., and Trahanias, P. (2011). Temporal cognition: a key ingredient of intelligent systems. *Front. Neurorobot.* 5:2. doi: 10.3389/fnbot.2011.00002

Marr, D. (1982). *Vision: A Computational Approach.* New York, NY: Freeman.

Martinson, E., Lawson, W., and Trafton, J. G. (2013). "Identifying people with soft-biometrics at Fleet Week," in *Proceedings of the 8th ACM/IEEE International Conference on Human Robot Interaction* (Tokyo: IEEE), 49–56.

Miller, G., and Johnson-Laird, P. N. (1976). *Language and Perception.* Cambridge, MA: Belknap Press.

Moens, M., and Steedman, M. (1988). Temporal ontology and temporal reference. *Comput. Linguist.* 14, 15–28.

Monti, M. M., Parsons, L. M., and Osherson, D. N. (2009). The boundaries of language and thought in deductive inference. *Proc. Natl. Acad. Sci. U.S.A.* 106, 12554–12559. doi: 10.1073/pnas.0902422106

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

Nelson, K., and Gruendel, J. (1986). "Children's scripts," in *Event knowledge: Structure and Function in Development*, ed K. Nelson (Hillsdale, NJ: Erlbaum), 21–46.

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *J. Pers. Soc. Psychol.* 28, 28–38. doi: 10.1037/h0035584

Newtson, D., Engquist, G., and Bois, J. (1977). The objective basis of behavior units. *J. Pers. Soc. Psychol.* 35, 847–862. doi: 10.1037/0022-3514.35.12.847

Oaksford, M. (2015). Imaging deductive reasoning. *Front. Hum. Neurosci.* 9:101. doi: 10.3389/fnhum.2015.00101

Peirce, C. S. (1931–1958). *Collected Papers of Charles Sanders Peirce.* Edited by C. Hartshorne, P. Weiss, and A. Burks. Cambridge, MA: Harvard University Press.

Piaget, J. (1927/1969). *The Child's Conception of Time*. London: Routledge and Kegan Paul.

Prado, J., Chadha, A., and Booth, J. R. (2011). The brain network for deductive reasoning: a quantitative meta-analysis of 28 neuroimaging studies. *J. Cogn. Neurosci.* 23, 3483–3497. doi: 10.1162/jocn_a_00063

Radvansky, G. A., and Copeland, D. E. (2006). Walking through doorways causes forgetting: situation models and experienced space. *Mem. Cogn.* 34, 1150–1156. doi: 10.3758/BF03193261

Radvansky, G. A., Tamplin, A. K., and Krawietz, S. A. (2010). Walking through doorways causes forgetting: environmental integration. *Psychon. Bull. Rev.* 17, 900–904. doi: 10.3758/PBR.17.6.900

Radvansky, G. A., and Zacks, J. M. (2011). Event perception. *WIREs Cogn. Sci.* 2, 608–620. doi: 10.1002/wcs.133

Reverberi, C., Görgen, K., and Haynes, J.-D. (2012). Distributed representations of rule identity and rule order in human frontal cortex and striatum. *J. Neurosci.* 32, 17420–17430. doi: 10.1523/JNEUROSCI.2344-12.2012

Ragni, M., and Knauff, M. (2013). A theory and a computational model of spatial reasoning with preferred mental models. *Psychol. Rev.* 120, 561–588. doi: 10.1037/a0032460

Ross, R. S., Brown, T. I., and Stern, C. E. (2009). The retrieval of learned sequences engages the hippocampus: evidence from fMRI. *Hippocampus* 19, 790–799. doi: 10.1002/hipo.20558

Schaeken, W., Johnson-Laird, P. N., and d'Ydewalle, G. (1996). Mental models and temporal reasoning. *Cognition* 60, 205–234.

Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B., and Botvinick, M. M. (2013). Neural representations of events arise from temporal community structure. *Nat. Neurosci.* 16, 486–492. doi: 10.1038/nn.3331

Schubotz, R. I., Korb, F. M., Schiffer, A.-M., Stadler, W., and von Cramon, D. Y. (2012). The fraction of an action is more than a movement: neural signatures of event segmentation in fMRI. *Neuroimage* 61, 1195–1205. doi: 10.1016/j.neuroimage.2012.04.008

Speer, N. K., Reynolds, J. R., Swallow, K. M., and Zacks, J. M. (2009). Reading stories activates neural representations of Visual and motor experiences. *Psychol. Sci.* 20, 989–999. doi: 10.1111/j.1467-9280.2009.02397.x

Speer, N. K., Swallow, K. M., and Zacks, J. M. (2003). Activation of human motion processing areas during event perception. *Cogn. Affect. Behav. Neurosci.* 3, 335–345. doi: 10.3758/CABN.3.4.335

Sridharan, D., Levitin, D. J., Chafe, C. H., Berger, J., and Menon, V. (2007). Neural dynamics of event segmentation in music: converging evidence for dissociable ventral and dorsal networks. *Neuron* 55, 521–532. doi: 10.1016/j.neuron.2007.07.003

Staresina, B. P., and Davachi, L. (2009). Mind the gap: binding experiences across space and time in the human hippocampus. *Neuron* 63, 267–276. doi: 10.1016/j.neuron.2009.06.024

Trafton, J. G., Altmann, E., and Ratwani, R. M. (2011). A memory for goals model of sequence errors. *Cogn. Syst. Res.* 12, 134–143. doi: 10.1016/j.cogsys.2010.07.010

Trafton, J. G., Harrison, A. M. (2011). Embodied spatial cognition. *Top. Cogn. Sci.* 3, 686–706. doi: 10.1111/j.1756-8765.2011.01158.x

Trafton, J. G., Hiatt, L. M., Harrison, A. M., Tamborello, F. P., Khemlani, S. S., and Schultz, A. C. (2013). ACT-R/E: an embodied cognitive architecture for human-robot interaction. *J. Hum. Robot Interact.* 2, 30–55. doi: 10.5898/JHRI.2.1.Trafton

Travis, L. L. (1997). "Goal-based organization of event memory in toddles," in *Developmental Spans in Event Comprehension and Representation: Bridging Fictional and Actual Events*, eds P. W. v. d. Broek, P. J. Bauer, and T. Bovig (Mahwah, NJ: Erlbaum), 111–138.

Trickett, S. B., and Trafton, J. G. (2007). "What if. . .": the use of conceptual simulations in scientific reasoning. *Cogn. Sci.* 31, 843–875. doi: 10.1080/03640210701530771

Uttal, W. R. (2011). *Mind and Brain: A Critical Appraisal of Cognitive Neuroscience.* Cambridge, MA: MIT Press.

Vandierendonck, A., and De Vooght, G. (1994). "The time–spatialization hypothesis and reasoning about time and space," in *Temporal Reasoning and Behavioral Variability*, eds M. Richelle, V. de Keyser, G. d'Ydewalle, and A. Vandierendonck (Liège: Interuniversity Pole of Attraction), 99–125.

Vandierendonck, A., and De Vooght, G. (1997). Working memory constraints on linear reasoning with spatial and temporal contents. *Q. J. Exp. Psychol. A* 50, 803–820. doi: 10.1080/713755735

Waechter, R. L., and Goel, V. (2005). "Resolving valid multiple model inferences activates a left hemisphere network," in *Mental Models and Cognitive Psychology, Neuroscience, and Philosophy of Mind*, eds C. Held, M. Knauff, and G. Vosgerau (New York, NY: Elsevier).

Whitney, C., Huber, W., Klann, J., Weis, S., Krach, S., and Kircher, T. (2009). Neural correlates of narrative shifts during auditory story comprehension. *Neuroimage* 47, 360–366. doi: 10.1016/j.neuroimage.2009.04.037

Wynn, K. (1996). Infants' individuation and enumeration of actions. *Psychol. Sci.* 7, 164–169. doi: 10.1111/j.1467-9280.1996.tb00350.x

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., et al. (2001a). Human brain activity time-locked to perceptual event boundaries. *Nat. Neurosci.,* 4, 651–655. doi: 10.1038/88486

Zacks, J. M. (2005). "Parsing activity into meaningful events," in *IEEE International Workshop on Robot and Human Interactive Communication* (IEEE). 190–195.

Zacks, J. M., Speer, N. K., Swallow, K. M., and Maley, C. J. (2010). The brain's cutting-room floor: segmentation of narrative cinema. *Front. Hum. Neurosci.* 4:168. doi: 10.3389/fnhum.2010.00168

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., and Reynolds, J. R. (2007). Event perception: a mind-brain perspective. *Psychol. Bull.* 133, 273–293. doi: 10.1037/0033-2909.133.2.273

Zacks, J., and Tversky, B. (2001). Event structure in perception and cognition. *Psychol. Bull.* 127, 3–21. doi: 10.1037/0033-2909.127.1.3

Zacks, J. M., Tversky, B., and Iyer, G. (2001b). Perceiving, remembering, and communicating structure in events. *J. Exp. Psychol.* 130, 29–58. doi: 10.1037/0096-3445.130.1.29

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.

frontiers in Human Neuroscience

PERSPECTIVE
published: 23 March 2015
doi: 10.3389/fnhum.2015.00148

Evidence for an inhibitory-control theory of the reasoning brain

Olivier Houdé [1,2] *and Grégoire Borst* [1]*

[1] CNRS Unit 8240, Laboratory for the Psychology of Child Development and Education, Alliance for Higher Education and Research Sorbonne-Paris-Cité, Paris Descartes University, Paris, France, [2] Institut Universitaire de France, Paris, France

In this article, we first describe our general inhibitory-control theory and, then, we describe how we have tested its specific hypotheses on reasoning with brain imaging techniques in adults and children. The innovative part of this perspective lies in its attempt to come up with a brain-based synthesis of Jean Piaget's theory on logical algorithms and Daniel Kahneman's theory on intuitive heuristics.

Keywords: inhibitory control, reasoning, heuristics/algorithm, developmental cognitive neuroscience, Piaget

Based on the numerous scientific data garnered in children of all ages, Jean Piaget (Piaget, 1983) proposed a seminal model of cognitive development according to which children's cognitive abilities developed through four different stages from the sensorimotor stage (from birth to 2 years of age) to the formal operational stage (starting at 12 years of age). Between two and 7 years of age (the so-call preoperational stage), Piaget assumed that children were mainly illogical in comparison to adults. Importantly, during the concrete operational stage, between 7 and 12 years of age, children start to reason logically in several logico-mathematical domains (e.g., number, categorization. . .). Finally, after 12 years of age, children's reasoning is not limited to concrete objects but can be applied to abstract propositions.

OPEN ACCESS

Edited by:
Ira Andrew Noveck,
Centre Nationale de la Recherche Scientifique, France

Reviewed by:
Melanie Stollstorff,
University of Colorado Boulder, USA
Einat Shetreet,
Tufts University, USA

Correspondence:
Grégoire Borst, CNRS Unit 8240,
Laboratory for the Psychology of
Child Development and Education,
Alliance for Higher Education and
Research Sorbonne-Paris-Cité, Paris
Descartes University, LaPsyDÉ,
Sorbonne, 46 rue Saint-Jacques,
75005 Paris, France
gregoire.borst@parisdescartes.fr

Received: *09 October 2014*
Accepted: *03 March 2015*
Published: *23 March 2015*

Citation:
Houdé O and Borst G (2015)
Evidence for an inhibitory-control theory of the reasoning brain.
Front. Hum. Neurosci. 9:148.
doi: 10.3389/fnhum.2015.00148

Inhibitory-Control Theory as an Alternative to Piaget's Theory

In fact, Piaget underestimated the rich precocious logical knowledge already present in infants and young children, and he overestimated the logical abilities of older children, adolescents and adults, who commit systematic errors even in very simple logical tasks (Houdé, 2000; Kahneman, 2011). These logical errors usually occur when older children, adolescents and adults rely on prepotent responses, illogical intuitions, or misleading strategies (such as heuristics) rather than on logical algorithms. Importantly, the ability to overcome those errors is directly related to the ability to inhibit these intuitive forms of thinking (Houdé, 2000; Kahneman, 2011; Houdé and Borst, 2014). Consequently, today the discrete Piagetian stages theory is replaced by an approach of cognitive development which is analogous to overlapping waves within a non-linear dynamic system (Siegler, 1999). In such a system, at any point in time and at any age, different strategies with different degrees of complexity and sophistication might be in conflict in the brain. According to this theoretical framework, the progressive ability of the prefrontal cortex to inhibit irrelevant or misleading strategies to activate the most logical one sustains the conceptual development of children and the shift from one Piagetian stage to the next (Houdé and Borst, 2014). This constitutes the central assumption of our new neo-Piagetian theory of reasoning development.

During cognitive development, children and adults have to choose, depending on the context, between two types of strategies or multiple levels of "thinking fast and slow" (Kahneman, 2011). Typically, individuals can either solve problems using heuristics (i.e., intuitions) or logico-mathematical algorithms. On the one hand, heuristics are typically defined as strategies that are effortless, rapid, often global or holistic which constitute the most adaptive response in most situations but sometimes they are misleading especially in situations in which they compete with logical algorithms. Algorithms, on the other hand, are slow, analytical and cognitively costly strategies but they always provide the correct solution independently of the context. In most contexts, children and adults spontaneously rely on heuristics. However, choosing heuristics over algorithms does not mean that children and adults are irrational *per se* (Houdé, 2000) or "happy fools" (De Neys et al., 2013). A "presumption of rationality" is sometimes the best assessment.

Brain Imaging of Reasoning-Bias Inhibition in Adults: The Example of Deductive Logic

As opposed to Piaget's theory, which assumed that children reached a logical stage of reasoning at 12 years of age (i.e., formal operational stage), a number of studies have now provided converging evidence that adolescents and adults continue to make errors in simple deductive reasoning tasks (see e.g., Evans, 1998, 2003; Houdé, 2000). For instance, in the perceptual matching bias task designed by Evans (Evans, 1998), the vast majority of participants choose a red square on the left of a yellow circle to falsify the following rule: "*if there is not a red square on the left, then there is a yellow circle on the right*". Evans attributed this error of logic to a perceptual matching bias (or heuristic) according to which participants choose the two geometrical shapes mentioned in the rules because a negation is present in the antecedent rather than using the logical truth table (in this case the algorithm). By using the logical truth table, participants would chose two geometrical shapes (e.g., a blue diamond to the left of a green square) validating a true antecedent (i.e., not a red square) and a false consequent (i.e., not a yellow circle). Critically, in order to avoid systematic logical errors in this context, participants must resist (or inhibit) the perceptual matching bias (i.e., red square on the left of a yellow circle) to activate the logical algorithm.

According to our "presumption of rationality" analysis, participants' difficulty in solving this if-then logical problem is not related to the difficulty of the deductive reasoning *per se* but to the difficulty to exert inhibitory control over the misleading heuristic (i.e., the perceptual matching bias). To provide evidence for the role of inhibitory control in overcoming deductive reasoning errors, we contrasted the effect of two types of training on the ability to perform deductive reasoning tasks. In one condition, participants were trained to inhibit the perceptual matching bias. In the other condition, participants received training focusing on explaining the underlying logic of the task. Importantly, participants were trained on a different deductive task (i.e., the Wason task, Wason, 1968) than the one performed pre- and post-training (i.e., the perceptual matching bias task, Evans, 1998). The effects of the two types of training were compared to a test-retest control condition in which participants simply performed the perceptual matching task two times. Participants who were trained to inhibit the perceptual matching heuristic were the only ones who succeeded to overcome their deductive reasoning errors. This finding suggests that logical reasoning errors are not due to a lack of logic (or experience) but to a default to inhibit a misleading heuristic. In a follow-up PET (positron emission tomography) imaging study in which we compared the cerebral activation before and after the participants were trained in inhibiting the perceptual matching bias, we observed that the brain activation shifted from the posterior perceptual regions pre-training to prefrontal executive regions post-training. This is the first micro-longitudinal neuroimaging study of deductive reasoning and it provides the first evidence that inhibitory control was critical to reason logically.

Note that this brain imaging study on reasoning errors correction was conducted on a sample of only eight participants but the strength of these results stem from the fact that the participants were their own controls in the pre-post training comparison. Such intra-individual design is scarce in brain imaging of reasoning. Indeed Fuster (Fuster, 2003), noted about our results that "the exercise of logical reasoning seems to overcome (or to inhibit) the biasing influences from the posterior cortex and to lend to prefrontal cortex the effective control of the reasoning task" (p. 231). More specifically, with respect to our results in the prefrontal cortex, we observed a left-middle-frontal gyrus activation which was likely to reflect the logical manipulation of the algorithm in working memory, and a left-inferior-frontal gyrus activation, which was likely to reflect inhibition of the reasoning bias (or heuristic) and self-regulatory inner speech (Broca's area).

In this brain imaging study, the training condition that focused on the inhibition of the misleading heuristic comprised not only cognitive but also emotional executive warnings that were not incorporated in the training condition focusing on explaining the underlying logic of the deductive problem. By directly contrasting the cerebral activity elicited by the two types of training, we found greater activity (i.e., the rCBF: regional cerebral blood flow) following inhibitory control training in the right ventromedial prefrontal cortex (Houdé et al., 2001), which is a paralimbic emotional area (Mesulam, 2000) known to be involved in getting the mind on the "logical track" and avoiding decision-making errors (Damasio et al., 1994; Damasio and Carvalho, 2013). We speculate that the right ventromedial prefrontal cortex could serve as an internal warning/self-feeling device to correct errors during deductive reasoning. Converging data on the link between emotion, conflict detection and inhibition were reported by Spiess et al. (2007) and De Neys et al. (2010).

After these two pioneer brain imaging studies on *if-then* rules (Houdé et al., 2000, 2001), a set of new studies were published during the past decade on deductive reasoning (e.g., Noveck et al., 2004; Prado and Noveck, 2007; for reviews see Goel, 2007; Prado et al., 2011). Noveck et al. (2004) studied the underlying brain network engaged in deductive reasoning on abstract contents and found that a left lateralized parietal-frontal network supported the if-then (or conditional) reasoning. Importantly, the activation within this network increased as the reasoning became more complex. As noted by Noveck et al. (2004), a critical difference between their study and the two neuroimaging studies we conducted was that solving Evans's problem required a counterintuitive solution—i.e., a solution that involved inhibiting the misleading heuristic. Prado and Noveck (2007) using a similar deductive reasoning task as the one we used provided convergent evidence that the resolution of such problems involved inhibitory control. In their study, participants were asked to determine whether a conditional rule such as "if there is not a B there is a triangle" was falsified (or verified) by an item (e.g., A and diamond). They reported increased activation in the right mid-dorsolateral prefrontal cortex (mid-DLPFC), the medial frontal areas (including the anterior cingulate area), the pre-supplementary motor area and the parietal cortices with increasing perceptual mismatch between the conditional rule and the item (i.e., when the perceptual matching bias was stronger). Critically, a psychophysiological interaction analysis revealed that the integration between the visual areas of the brain (supporting the perceptual matching heuristic) and mid-DLPFC decreased when the perceptual mismatch increased. Taken together the results suggest that overcoming the perceptual matching bias is rooted in part by the inhibitory control exerted by prefrontal regions (i.e., mid-DLPFC and the medial frontal cortex) on lower level visual regions.

Note that whereas the left lateral prefrontal structures (including the left IFG) supported the inhibition of the misleading heuristic during conditional reasoning in our studies (Houdé et al., 2000, 2001) subsequent studies reported activation in the right IFG (e.g., Noveck et al., 2004; Prado and Noveck, 2007; for reviews see Goel, 2007; Prado et al., 2011). We suspect that the activation in the left prefrontal areas of the brain reported in our seminal studies could be a consequence of the verbal nature of the executive training (given between the pre- and post-test) which would have favored using inhibitory control in verbal working memory after the training (i.e., during the post-test). This interpretation is coherent with previous studies showing that inhibition in verbal working memory is supported by the left prefrontal areas of the brain (Jonides et al., 1998).

The role of inhibitory control and the prefrontal cortex (including the inferior frontal gyrus, IFG) in deductive reasoning has been demonstrated not only using conditional reasoning but also syllogistic reasoning (De Neys and Van Gelder, 2009; Tsujii et al., 2010, 2011). For instance, Tsujii et al. (2010) investigated the network of brain areas involved in syllogistic reasoning. Critically, prefrontal regions including the right IFG—i.e., a region consistently activated when a prepotent response (or a heuristic) is inhibited (see Aron et al., 2004, 2014)—are specifically recruited when participants judge the validity of syllogisms in which the logical validity of the conclusion is in conflict with the belief of the participants (e.g., Valid incongruent syllogism: *No mammals are dogs/All German Shepherd are mammals/No German Shepherd are dogs*). Importantly, a follow-up study revealed that the ability to reason on belief laden syllogisms is impaired when the activity of the right IFG is disrupted using rTMS (i.e., repetitive Transcranial Magnetic Stimulation). This study provided additional evidence for a causal relation between the right IFG and the ability to overcome logical errors through the inhibition of heuristic thinking.

Brain Imaging of Reasoning-Bias Inhibition in Children: The Example of Number Conservation

One of the most famous Piagetian problems used for testing reasoning in children is the number-conservation task (Piaget, 1983). In this problem, the child is first presented with two rows of tokens with the same number of tokens and the same length. After the child acknowledges that the two rows contain the same number of objects, the tokens in one of the rows are spread apart and the child is asked whether the two rows contain yet the same number of tokens. Children younger than 6 or 7 years of age tend to report that the longer row contains more tokens. According to Piaget (1952), young children make systematic errors in the number-conservation problem because they rely on an intuitive "illogical" mode of thinking which is a hallmark of the preoperational stage of cognitive development. When children reach 6 or 7 years of age, they successfully solve the number conservation task by understanding the reversibility of operations (any transformation can be cancelled out by the reverse transformation) which is evidence that children are in the concrete operational stage of development.

Following Piaget's pioneer work, a growing number of studies were proposed to investigate the cognitive development of numeracy and raised numerous criticisms of Piaget's theory. For instance, studies have demonstrated that newborns and infants understand that there is an invariance between number and physical transformations, even in contexts extremely similar as the one created in the number-conservation problem (Antell and Keating, 1983; see also Dehaene, 2011). A critical question for developmental psychologist is thus to understand why newborns and infants who have some knowledge of the relation between number and space will later on make systematic errors in the number-conservation problem until age 6 or 7. This non-linear pattern of development could be explained by the fact that children learn a number of heuristics during their childhood that are most of the times appropriate to find the solution except in context in which they are misleading and need to be inhibited (Houdé, 2000; Houdé and Borst, 2014). For instance, in Piaget's number-conservation problem, children tend to rely on the misleading length-equals-number heuristic rather than on a counting or operational reversibility algorithm.

One of the challenges of today's research in developmental psychology is thus to shift from the Piagetian (Piaget, 1983) and neo-Piagetian (see Demetriou, 1988 for a review) views that the conceptual change exclusively relies on the growing ability to coordinate multiple systems of operations to a view according to which conceptual change is in part rooted in a domain-general ability of selection-inhibition of competing strategies, i.e., heuristics (or intuitions) and logico-mathematical algorithms. Critically, at each age and in each situation the strengths of the heuristics and the algorithms fluctuate within a nonlinear dynamical system (Siegler, 1999; Houdé, 2000; Houdé and Borst, 2014). According to this new model, cognitive development occurs in bursts with sometimes errors occurring after success in both children and adults. This model is coherent with what we know of the structural changes of the brain from childhood to adulthood (Casey et al., 2005). Indeed, the inhibition of heuristics could remain challenging because the maturation of the prefrontal cortex sustaining inhibitory-control ability continues throughout childhood and adolescence.

To determine whether the growing ability to perform Piaget's number-conservation problem is rooted in the growing ability to inhibit the *length-equals-number* heuristic due to the progressive maturation of the prefrontal cortex, we asked 60 children aged 5–10 to solve Piagetian problems in a functional magnetic resonance imaging (fMRI) study. We found that children who succeed in solving Piaget's number-conservation problems (i.e., children aged 7 and older) recruited a parieto-frontal network including the right IFG and the bilateral intra parietal sulcus (IPS; Houdé et al., 2011)—two regions respectively involved in inhibition (e.g., Aron et al., 2004, 2014) and numeracy (e.g., Dehaene, 2011). In a subsequent fMRI study (Poirel et al., 2012), we provided evidence that the recruitment of the right IFG was directly related to the need to inhibit a heuristic by reporting a significant positive correlation between the BOLD (i.e., the blood-oxygen-level-dependent) signal in the rIFG and the inhibitory control efficiency as measured by an Animal Stroop task (Wright et al., 2003)—a Stroop task adapted for non-reading children. The results we garnered in schoolchildren are coherent with the ones we reported above in adolescents and adults for which failure to inhibit a heuristic led to systematic logical errors although they reached the formal operational stage according to Piaget's theory. Note, however, that our developmental study on number conservation shows a right-inferior-frontal gyrus activation for inhibition (in line with Aron et al. (2004, 2014) meta-analysis reviews), while our adults study on deductive reasoning (Houdé et al., 2000) showed a left-inferior-frontal gyrus activation for inhibition. In this last study, there was no Stroop-correlation control, but the leftward lateralization was probably due to the strong verbal component (rules) of the logical task, involving self-regulatory inner speech. The number conservation problem is, inversely, a visuospatial task which fits well with a rightward lateralization of the activation.

Conclusion

In this review we want to argue that learning to inhibit misleading heuristics from System 1 (i.e., intuitive system) when they interfere with the activation of the logical algorithms from System 2 (i.e., analytical system, see e.g., Evans, 2003; Kahneman, 2011) is the critical process that allows one to reason logically (Houdé, 2000; Goel, 2007; Prado and Noveck, 2007; De Neys and Van Gelder, 2009; Tsujii et al., 2010, 2011; Prado et al., 2011; Houdé and Borst, 2014). The new post-Piagetian theoretical framework we propose allows us to better understand why newborns and infants who possess an early ability to reason logically in different domains will later in life have the tendency to reason illogically. Typically, at all ages, overcoming systematic logical errors relies on blocking (i.e., inhibiting) our intuitions, a process that is highly dependent on the maturation of the prefrontal cortex (Borst et al., 2013). Finally, the ability to inhibit misleading heuristics remains challenging throughout our lifetime. Thus children, adolescents and adults may sometimes need "prefrontal pedagogy" to help them overcome their tendency to rely on intuitive heuristics and biases in reasoning tasks (Houdé, 2007).

References

Antell, S. E., and Keating, D. P. (1983). Perception of numerical invariance in neonates. *Child Dev.* 54, 695–701. doi: 10.2307/1130057

Aron, A., Robbins, T., and Poldrack, R. (2004). Inhibition and the right inferior frontal cortex. *Trends Cogn. Sci.* 8, 170–177. doi: 10.1016/j.tics.2004.02.010

Aron, A., Robbins, T., and Poldrack, R. (2014). Inhibition and the right inferior frontal cortex: one decade on. *Trends Cogn. Sci.* 18, 177–185. doi: 10.1016/j.tics.2013.12.003

Borst, G., Moutier, S., and Houdé, O. (2013). "Negative priming in logicomathematical reasoning," in *New Approaches in Reasoning Research*, eds W. De Neys and M. Osman (New York: Psychology Press), 34–50.

Casey, B., Tottenham, N., Liston, C., and Durston, S. (2005). Imaging the developing brain. *Trends Cogn. Sci.* 9, 104–110. doi: 10.1016/j.tics.2005.01.011

Damasio, A., and Carvalho, G. B. (2013). The nature of feelings. *Nat. Rev. Neurosci.* 14, 143–152. doi: 10.1038/nrn3403

Damasio, H., Grabowski, T., Frank, R., Galaburda, A., and Damasio, A. (1994). The return of Phineas Gage: clues about the brain from the skull of a famous patient. *Science* 264, 1102–1105. doi: 10.1126/science.8178168

Dehaene, S. (2011). *The Number Sense.* New York: Oxford University Press.

Demetriou, A. (ed) (1988). *The Neo-Piagetian Theories of Cognitive Development.* Amsterdam: North-Holland.

De Neys, W., Moyens, E., and Vansteenwegen, D. (2010). Feeling we're biased: autonomic arousal and reasoning conflict. *Cogn. Affect. Behav. Neurosci.* 10, 208–216. doi: 10.3758/CABN.10.2.208

De Neys, W., Rossi, S., and Houdé, O. (2013). Bats, balls and substitution sensitivity. *Psychon. Bull. Rev.* 20, 269–273. doi: 10.3758/s13423-013-0384-5

De Neys, W., and Van Gelder, E. (2009). Logic and belief across the life span: the rise and fall of belief inhibition during syllogistic reasoning. *Dev. Sci.* 12, 123–130. doi: 10.1111/j.1467-7687.2008.00746.x

Evans, J. (1998). Matching bias in conditional reasoning. *Think. Reason.* 4, 45–82. doi: 10.1080/135467898394247

Evans, J. (2003). In two minds: dual-process accounts of reasoning. *Trends Cogn. Sci.* 7, 454–459. doi: 10.1016/j.tics.2003.08.012

Fuster, J. (2003). *Cortex and Mind.* New York: Oxford University Press.

Goel, V. (2007). Anatomy of deductive reasoning. *Trends Cogn. Sci.* 11, 435–441. doi: 10.1016/j.tics.2007.09.003

Houdé, O. (2000). Inhibition and cognitive development. *Cogn. Dev.* 15, 63–73. doi: 10.1016/s0885-2014(00)00015-0

Houdé, O. (2007). First insights on neuropedagogy of reasoning. *Think. Reason.* 13, 81–89. doi: 10.1080/13546780500450599

Houdé, O., and Borst, G. (2014). Measuring inhibitory control in children and adults. *Front. Psychol.* 5:616. doi: 10.3389/fpsyg.2014.00616

Houdé, O., Pineau, A., Leroux, G., Poirel, N., Perchey, G., Lanoë, C., et al. (2011). Functional MRI study of Piaget's conservation-of-number task in preschool and school-age children. *J. Exp. Child Psychol.* 110, 332–346. doi: 10.1016/j.jecp.2011.04.008

Houdé, O., Zago, L., Crivello, F., Moutier, S., Pineau, A., Mazoyer, B., et al. (2001). Access to deductive logic depends on a right ventromedial prefrontal area devoted to emotion and feeling. *Neuroimage* 14, 1486–1492. doi: 10.1006/nimg.2001.0930

Houdé, O., Zago, L., Mellet, E., Moutier, S., Pineau, A., Mazoyer, B., et al. (2000). Shifting from the perceptual brain to the logical brain. *J. Cogn. Neurosci.* 12, 721–728. doi: 10.1162/089892900562525

Jonides, J., Smith, E. E., Marshuetz, C., Koeppe, R. A., and Reuter-Lorenz, P. A. (1998). Inhibition in verbal working memory revealed by brain activation. *Proc. Natl. Acad. Sci. U S A* 95, 8410–8413. doi: 10.1073/pnas.95.14.8410

Kahneman, H. (2011). *Thinking Fast and Slow.* London: Allen Lane.

Mesulam, M. (2000). *Principles of Behavioral and Cognitive Neurology.* New York: Oxford University Press.

Noveck, I., Goel, V., and Smith, K. (2004). The neural basis of conditional reasoning with arbitrary content. *Cortex* 40, 613–622. doi: 10.1016/s0010-9452(08)70157-6

Piaget, J. (1952). *The Child's Conception of Number.* New York: Routledge and Kegan Paul.

Piaget, J. (1983). "Piaget's theory," in *Handbook of Child Psychology* (Vol. 1), ed P. H. Mussen (New York: Wiley), 103–128.

Poirel, N., Borst, G., Simon, G., Rossi, S., Cassotti, M., Pineau, A., et al. (2012). Number conservation is related to children's prefrontal inhibitory control. *PLoS One* 7:e40802. doi: 10.1371/journal.pone.0040802

Prado, J., Chadha, A., and Booth, J. R. (2011). The brain network for deductive reasoning: a quantitative meta-analysis of 28 neuroimaging studies. *J. Cogn. Neurosci.* 23, 3483–3497. doi: 10.1162/jocn_a_00063

Prado, J., and Noveck, I. (2007). Overcoming perceptual features in logical reasoning: a parametric fMRI study. *J. Cogn. Neurosci.* 19, 642–657. doi: 10.1162/jocn.2007.19.4.642

Siegler, R. (1999). Strategic development. *Trends Cogn. Sci.* 3, 430–435. doi: 10.1016/S1364-6613(99)01372-8

Spiess, J., Etard, O., Mazoyer, B., Tzourio-Mazoyer, N., and Houdé, O. (2007). The skin-conductance component of error correction in a logical reasoning task. *Current Psychology Letters* 23. Available online at: http://cpl.revues.org/2872

Tsujii, T., Masuda, S., Akiyama, T., and Watanabe, S. (2010). The role of inferior frontal cortex in belief-bias reasoning: an rTMS study. *Neuropsychologia* 48, 2005–2008. doi: 10.1016/j.neuropsychologia.2010.03.021

Tsujii, T., Sakatani, K., Masuda, S., Akiyama, T., and Watanabe, S. (2011). Evaluating the roles of the inferior frontal gyrus and superior parietal lobule in deductive reasoning: an rTMS study. *Neuroimage* 58, 640–646. doi: 10.1016/j.neuroimage.2011.06.076

Wason, P. C. (1968). Reasoning about a rule. *Q. J. Exp. Psychol.* 20, 273–281. doi: 10.1080/14640746808400161

Wright, I., Waterman, M., Prescott, H., and Murdoch-Eaton, D. (2003). A new stroop-like measure of inhibitory function development. *J. Child Psychol. Psychiatry* 44, 561–575. doi: 10.1111/1469-7610.00145

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.

frontiers
in Human Neuroscience

PERSPECTIVE
published: 24 April 2015
doi: 10.3389/fnhum.2015.00222

How can we study reasoning in the brain?

David Papo *

GISC and Laboratory of Biological Networks, Center for Biomedical Technology, Universidad Politécnica de Madrid, Madrid, Spain

The brain did not develop a dedicated device for reasoning. This fact bears dramatic consequences. While for perceptuo-motor functions neural activity is shaped by the input's statistical properties, and processing is carried out at high speed in hardwired spatially segregated modules, in reasoning, neural activity is driven by internal dynamics and processing times, stages, and functional brain geometry are largely unconstrained *a priori*. Here, it is shown that the complex properties of spontaneous activity, which can be ignored in a short-lived event-related world, become prominent at the long time scales of certain forms of reasoning. It is argued that the neural correlates of reasoning should in fact be defined in terms of non-trivial generic properties of spontaneous brain activity, and that this implies resorting to concepts, analytical tools, and ways of designing experiments that are as yet non-standard in cognitive neuroscience. The implications in terms of models of brain activity, shape of the neural correlates, methods of data analysis, observability of the phenomenon, and experimental designs are discussed.

Keywords: cognitive neuroscience, reasoning, scaling, non-stationarity, non-ergodicity, characteristic scales, observation time, resting brain activity

OPEN ACCESS

Edited by:
Ira Andrew Noveck,
Centre Nationale de la Recherche Scientifique, France

Reviewed by:
Jascha Ruesseler,
University of Bamberg, Germany
Jérôme Prado,
Centre National de la Recherche Scientifique, France

***Correspondence:**
David Papo,
GISC and Laboratory of Biological Networks, Center for Biomedical Technology, Universidad Politécnica de Madrid, Calle Ramiro de Maeztu, 7, 28040 Madrid, Spain
papodav@gmail.com

Received: *18 November 2014*
Accepted: *08 April 2015*
Published: *24 April 2015*

Citation:
Papo D (2015) How can we study reasoning in the brain? Front. Hum. Neurosci. 9:222. doi: 10.3389/fnhum.2015.00222

Introduction

Consider an individual trying to solve a problem and reasoning for 10 min before attaining a solution. Take the middle 5 min. Clearly, though containing no behaviorally salient event, these 5 min represent a genuine, indeed rather general, instance of reasoning. What do we know about the brain regime far from its conclusion? Can we use this regime to predict a solution, and a solution to retrodict this regime?

Here, I concentrate on a form of reasoning, of which the above scenario constitutes an example, which can broadly be defined as "thinking in which there is a conscious intent to reach a conclusion and in which methods are used that are logically justified" (Moshman, 1995), with no a priori assumption on the type of reasoning process that may take place during it. It is argued that finding the *generic* properties of this form of reasoning entails addressing the following fundamental issues: What are reasoning's temporal and spatial scales? When is a given observation time sufficient? How should we integrate the information contained in various reasoning episodes?

A Mini Literature Review

The neural correlates of reasoning have traditionally been expressed in terms of brain spatial coordinates. Early neuropsychological work viewed reasoning as emerging from global brain processing (Gloning and Hoff, 1969), consistent with evidence indicating that it is negatively affected by diffuse brain damage (Lezak, 1995). Neuroimaging studies have framed the neural correlates of reasoning in terms of local functionally specialized brain activity, either by taking a normative approach to reasoning (Goel et al., 1997, 1998; Osherson et al., 1998; Parsons and Osherson, 2001; Noveck et al., 2004; Prado et al., 2011), or by fractionating it into subcomponent processes (Houdé et al., 2001; Acuna et al., 2002; Kroger et al., 2002; Reverberi et al., 2012). The results often lack specificity to reasoning (Papo et al., 2007). Most importantly, these investigations provide a static characterization of reasoning.

The neuroimaging literature mostly focused on short-term and normative forms of reasoning (Prado et al., 2011; Bonnefond et al., 2013, 2014). This minimizes variability in reasoning episode length and allows segmenting reasoning episodes into separable chunks, but does that at the price of limitations in the phenomenology and ecologic value of its stimuli. Some neuroimaging (Luo et al., 2004; Subramaniam et al., 2008) and electrophysiological (Jung-Beeman et al., 2004; Mai et al., 2004; Kounios et al., 2006, 2008; Lang et al., 2006; Bowden and Jung-Beeman, 2007; Qiu et al., 2008; Sandkühler and Bhattacharya, 2008; Sheth et al., 2008) studies examined more ecological forms of reasoning, viz. insight problems (Knoblich et al., 1999). However, even electrophysiological studies, despite optimal temporal resolution, adopted an event-related perspective, concentrating on activity occurring a few seconds before insight emergence, which only documents the *outcome* of the reasoning process, not the process itself.

Event-related neural activity associated with the solution of riddles with insight was found to be related to properties of preceding resting activity (Kounios et al., 2006, 2008). These studies had the remarkable merit of using spontaneous brain activity to characterize reasoning, but in essence provided a comparative statics description. Although some behavioral studies treated reasoning as a dynamical process (Stephen et al., 2009), a comparable neurophysiological characterization is still incomplete.

The Problem(s) with Reasoning

The generalized form of reasoning considered in this study comes in episodes offering scant behaviorally salient events with no characteristic temporal length. Each episode is a non-reproducible instance, as a reasoning task can be carried out in multiple ways. Brain activity associated with reasoning is not event-related, and many neurophysiological processes interact in a wide range of spatial and temporal scales.

These phenomena can all be traced back to a basic fact: the brain did not develop a dedicated device for reasoning. Hard-wired partially segregated modules ensure that perceptuo-motor functions are carried out at great speed, with stereotyped duration and time-varying profile, and identifiable stages, largely determined by input statistical properties. Reasoning, on the contrary, is associated with an internally-driven dynamics: processing times and stages, and functional brain geometry are largely unconstrained.

Considering these extraordinary challenges, can we still find general reasoning properties, over and above specific task demands and individual differences? What sort of process is reasoning in its general form? Is it a series of simpler reasoning cycles? Can we segment it into stages? What are the best neural variables and tools to make these properties observable?

Characterizing the Reasoning Process

Robust characterizations of reasoning should incorporate properties consistently appearing across different subjects and in different periods of time, and select analytical tools accordingly. For instance, perceptual response sensitivity to incoming signals, stability against noise, and minimal dependence on initial conditions favor tools capturing transient dynamics, which naturally reproduce these properties under appropriate conditions, over tools handling asymptotic activity, which fail to do so (Rabinovich et al., 2008).

Reasoning's relative instability and inefficiency suggest that optimal circuitry may need constant reconstruction and protection from interference, summoning protracted support of energetically costly long-range communications. Reasoning may be a sort of resonant regime, where functional efficiency would be achieved with specific, though unstable, spatio-temporal patterns. This suggests that reasoning should be studied with tools which can describe spatially-extended dynamic transients and can quantify information transfer and the corresponding energetic cost.

Reasoning Dynamics

Each cognitive process can be translated in dynamical terms and corresponding aspects of neural activity.

Perceptual processes are relaxational, quasi-stereotyped short duration processes. The brain can *prima facie* be modeled as an excitable medium: perturbations above a threshold induce a dynamical cycle before the system reverts to its initial silent state.

Learning too is a relaxational process. Following a gradient dynamics, the brain incorporates the environment's statistical relationships by representing them in terms of its functional connectivity (Sporns et al., 2000). Cycles can be of much longer duration and non-trivial shape than perceptual ones. No single instant summarizes the entire process, and the dynamics consists of fluctuations much shorter than the whole process.

Reasoning may not be purely relaxational. As in the case of learning, no instant summarizes the whole dynamics but, contrary to learning, there is no clear gradient. Neural activity is an out-of-equilibrium endogenously modulated spontaneous brain activity. Its phenomenology is considerably more complex than the equilibrium event-related short time-scale one of perception or the gradient-driven regression to equilibrium dynamics of learning.

To study reasoning, one should therefore first consider properties of spontaneous activity that are *generic* (i.e., that hold for almost all conditions) at long time scales and then see how these properties are modulated during reasoning (Papo, 2014a).

The Starting Point: Spontaneous Brain Activity

When observed long enough, brain fluctuations appear to be characterized by structured patterns (Kenet et al., 2003). The temporal sequence with which these patterns are re-edited across the cortical space also appears to have non-random structure (Beggs and Plenz, 2003, 2004; Cossart et al., 2003; Ikegaya et al., 2004; Dragoi and Tonegawa, 2011; Betzel et al., 2012). The structure with which these fluctuations appear can be described in the same way one would describe an object, characterizing its component parts, the relationships between them, and the way one can inspect it. For instance, if we think of brain fluctuations as the steps of a random walker, one can describe the *phase space*, i.e., the space of all states attainable by the system's dynamics, but also of traveled distances, times to reach a given target and memory of previous steps.

In the equilibrium world of perceptual scientists, brain steps are Gaussian distributed, and memory of past steps is lost so rapidly that no structure is apparent when considering the time course of activity. Spontaneous activity has no evident temporal structure and can be treated as a null state to which the brain reverts in the absence of stimulation.

At the long time scales of reasoning, the random walker takes steps from a non-Gaussian distribution. Like a fractal object, it displays similar properties at all scales (Novikov et al., 1997; Linkenkaer-Hansen et al., 2001; Gong et al., 2002; Freeman et al., 2003; Stam and de Bruin, 2004; Expert et al., 2010; van de Ville et al., 2010; Fraiman and Chialvo, 2012). While self-similarity may not be exact (Suckling et al., 2009; Zilber et al., 2012), these scaling patterns indicate that activity at different temporal scales is characterized by non-trivial relationships between them (Bacry et al., 2001; Friedrich et al., 2011; Papo, 2013b). Not all regions of the phase space are equally visited, with some taking an extremely long time to be reached (Bianco et al., 2007). Transitions from one region to the other depend on past history of the dynamics (Gilboa et al., 2005). Memory of past steps decays so slowly that the time it takes two time-points to totally decorrelate may diverge, so that a characteristic time ceases to exist (Grigolini et al., 1999; Fairhall et al., 2001; Gilboa et al., 2005; Lundstrom et al., 2008). Temporal correlations are not stationary, but time-dependent (Bianco et al., 2007). If, rather than an ordinary watch, one measured time with a watch ticking at every step taken by the walker, the passage of time would appear to be highly irregular and clustered, alternating between relatively quiet phases and more turbulent ones (Gong et al., 2007; Allegrini et al., 2010).

The temporal structure can be used to define landmarks within time-windows where no behaviorally salient event occurs. This can be done by identifying segments that can be considered stationary (Kaplan et al., 2005). The distribution of these segments' durations and their correlations and specific sequences may help clarify whether reasoning far away from both problem presentation and solution is merely a repetition of simple cycles seen in more controlled forms of reasoning, or is of a qualitatively different nature, and if so, may help determine the time scales at which simpler cycles are reedited.

To fully describe the phase space, one needs to consider that the brain as a whole consists of a great number of local random walkers. Local walkers interact to form transient patterns of connectivity. These patterns can be endowed with topological properties at all spatial scales by resorting to complex networks theory (Bullmore and Sporns, 2009). Eventually, one deals with an abstract structure consisting of spatial patterns endowed with topological properties, the temporal evolution of which displays the complex properties described above.

Overall, the space in which the random walker turns out to live, and which reflects the brain's dynamical repertoire, can be represented as a complex spatio-temporal structure (Zaslavsky, 2002). This structure can be described in terms of symmetries and universal properties, which are robust with respect to the nature of microscopic details, by resorting to a variety of methods, e.g., algebraic and differential topology, renormalization group methods etc. (Lesne, 2008; Petri et al., 2014). Using these methods it is possible (1) to partition the phase space, (2) to identify dynamical pathways leading to specific regions of this space, and (3) to relate descriptions of the same brain at different scales and of different brains exhibiting the same large-scale behavior (Lesne, 2008).

From Spontaneous Activity to Reasoning

Cognitive processes can be thought of as selections and orchestrations of cortical states already present in spontaneous activity (Kenet et al., 2003; Fiser et al., 2004; Luczak et al., 2009). Each process reveals a specific part of the phase space, and can be associated with its own topological properties and symmetries, and characteristic kinematics, memory, aging properties, degree of ergodicity, and internal clock (Papo, 2014a). For example, different conditions under which subjects carried out a reasoning task were shown to modulate the scaling regime of fluctuations of the corresponding brain activity (Buiatti et al., 2007), suggesting that reasoning may modulate not brain activity's amplitude but its functional form (Papo, 2014a), e.g., by forcing the system's stationary distribution to equal a target one. These modulations may correspond to cross-overs between universality classes, resulting from transitions between different dynamical regimes (Burov and Barkai, 2008).

The statistics of fluctuations can be used to study insight and to evaluate whether insight occurrence can be predicted. The sudden onset of insight may be thought of as an extreme event comparable to earthquakes, financial crashes, or epileptic seizures (Contoyiannis and Eftaxias, 2008; Osorio et al., 2010), e.g., as a rupture phenomenon, and the route to it as a long charging process, with nested hierarchical "earthquakes." The probability distribution of fluctuations gives an estimate of the likelihood of the occurrence of such events: for a Gaussian distribution, extreme events are exponentially rare. However, for non-Gaussian distributions, such events do occur with non-zero probability. It is tempting to conjecture that, in analogy with results of studies of these phenomena, insight onset may be predicted by monitoring changes in anomalous diffusion parameters (Contoyiannis and Eftaxias, 2008), Gaussianity (Manshour et al., 2009), or fractal spectrum complexity (de Arcangelis and Herrmann, 1989; Kapiris et al., 2004).

Assessing Reasoning: from Dynamics to Thermodynamics and Information

Considering the functions reasoning fulfills and the constraints the brain faces while performing it can shed light on ways in which brain fluctuations can help quantify how the brain carries out reasoning.

Reasoning, as other cognitive processes, e.g., memory recall (Rhodes and Turvey, 2007; Baronchelli and Radicchi, 2013), can be represented as a search process similar to that of animals foraging in an unknown environment (Viswanathan et al., 2011). This search process can be characterized in terms of random walks (Shlesinger et al., 1993; Codling et al., 2008; Lomholt et al., 2008; Bénichou et al., 2011). Importantly, the statistics of random steps and their correlations indicate the extent to which a given trajectory optimizes search, given the characteristics of the explored space and the resources available to the individual (Bénichou et al., 2011). Such a characterisation would allow assessing in a context-specific way the quality of both the reasoning and the "reasoned." That behavioral aspects of human cognition (Rhodes and Turvey, 2007; Baronchelli and Radicchi, 2013) and brain activity both show non-Gaussian, heavy-tailed distributions might indicate search optimality (Lomholt et al., 2008; Humphries et al., 2012). However, because these properties are generic in spontaneous activity, reasoning's quality can only be described in terms of its modulations, and finding the neural property and spatial scale showing such scaling modulations are the crucial steps.

Because it lacks a hardwired structure, reasoning faces both a stability and an energetic problem. Fluctuation dynamics can help address the first issue, but may not be sufficient *per se* to address the second. While a graph theoretical representation of functional brain activity may provide indications as to the ways the brain tackles both problems (Bullmore and Sporns, 2012; Papo et al., 2014), a direct characterization can be achieved by considering the brain as a very complex engine and by characterizing its thermodynamics. Crucially, thermodynamics can be deduced from dynamics (Sekimoto, 1998). Such a characterisation could be used to quantify variations in thermodynamic variables such as free energy, entropy, or temperature (Papo, 2013a) during a reasoning task, but also possible transitions in some other property of neural activity, for particular values of these variables. For instance, a suitably modified equilibrium temperature accounting for the non-equilibrium nature of brain activity (Cugliandolo, 2011) can quantify deviations of each spatio-temporal scale from equilibrium, entropy production, etc. (Papo, 2014b).

Finally, one may want to quantify reasoning in terms of the information created, erased, and transferred during its execution. Simple fluctuations can be thought of as letters of an alphabet, fluctuation complexes as words, and the reasoning process represented as a network traffic regulation problem. Characterizing traffic regulation and phenomena such as overload or jamming may involve using information-theoretical tools and complex network theory and understanding the interplay between the underlying network's topology, the dynamics of information packets and the shape of fluctuation distributions (DeDeo and Krakauer, 2012; Delvenne et al., 2013; Lambiotte et al., 2013). Although only causal information (Shalizi and Moore, 2003) may directly serve reasoning purposes, the total information encoded in the network may describe the noise-control mechanisms indirectly optimizing it. Interestingly, non-equilibrium systems such as the brain, information, and thermodynamics can be thought of as the opposite side of the same coin (Parrondo et al., 2015). Ultimately, the information content of reasoning-related neural activity could be extracted from its dynamics, via thermodynamics.

From Theory to Experiment

Observing Reasoning

Reasoning is a difficult phenomenon to observe: tasks can be executed in more than one-way, each possibly corresponding to a neural phase space with convoluted geometry and the processes involved in reasoning may evolve over time-scales exceeding those typical of laboratory testing.

Proper observation of a given process requires that the *observation time* be much larger than any scale in the system. A process is observable if it has a finite ratio between the characteristic time of the independent variable and the length of the available time series (Reiner, 1964). Factors including long-term memory, aging and weak ergodicity breaking may result in a diverging ratio (Rebenshtok and Barkai, 2007).

The observation time should also be much larger than the time needed to visit the neural phase space. The time needed to explore this space may far exceed the typical reasoning episode duration. Cognitive neuroscientists observe phenomena through experiments where subjects typically carry out given tasks a large number of times, assumed to be independent realizations of the same observable, and to adequately sample the phase space of task-related brain activity. However, in the presence of complex fluctuations, trials may not *self-average*, i.e., dispersion would not vanish even for an infinite number of trials (Aharony and Harris, 1996). Thus, trials may explore different aspects of the space of available strategies and may therefore improve phase space exploration rather than the signal-to-noise ratio (Ghosh et al., 2007).

Experimental Implications

Reasoning's characteristics, particularly its lack of characteristic temporal duration, have implications at various levels. First, episodes cannot be compared in an event-related fashion. Second, defining reliable neural correlates of reasoning requires defining its characteristic temporal scales. Third, measures of brain activity should be invariant with respect to overall duration. Scaling exponents, data collapse and universality of fluctuations statistics (Bramwell et al., 1998; Bhattacharya, 2009; Friedman et al., 2012), or explicit evolution equations for the particle's momenta and for the cross-scale fluctuation probabilities (Friedrich et al., 2011) can be retrieved from data and applied to unevenly lengthen trials. Thermodynamic quantities such as free energy or temperature can also be estimated for stochastic trajectories over finite time durations (Ruelle, 1978; Beck and Schlögl, 1997; Canessa, 2000; Olemskoi and Kokhan, 2006; Papo, 2014b). In all cases, the reconstruction of the underlying dynamics improves with the recording device's resolution.

Reasoning presents a dilemma between ensuring complete phase space exploration, which may require extremely long trials, and signal stationarity, which is guaranteed only for time scales much shorter than the reasoning episodes' duration. At fast time scales, the window in which relevant quantities are calculated should not introduce spurious time scales, filtering out genuine ones. Altogether, reasoning's inherently unstable nature suggests that describing it may boil down to characterizing non-stationarities and their aetiologies.

Reasoning tasks may be so difficult that only few participants manage to produce solutions within a reasonable time. This represents a shortcoming when trials are considered as independent and identically distributed, as the signal-to-noise ratio improves with the square root of the number of trials. Smoothing response times is a frequent strategy to obviate this problem, but limits or distorts the reasoning process. Furthermore, however many, short trials may insufficiently explore the phase space. Designs with few long trials may express richer spatio-temporal brain dynamics than many short ones of equivalent overall length.

Finally, while observed scaling properties may help us understand whether insight is *predictable*, i.e., whether it is an outlier or it is generated by the same distribution producing anonymous events, predicting insight onset in real data appears to be a challenging task, as reasoning episodes are various orders of magnitude shorter than earthquake, financial, or epilepsy time series (Sornette, 2002).

Conclusions

Reasoning elicits an exceptionally rich repertoire of otherwise unexpressed neural properties. Its neural correlates are therefore as helpful to neuroscientists, who are compelled to consider hitherto neglected brain properties, as they are to psychologists who strive to understand its underlying processes.

Defining general and robust mechanistic properties of healthy and dysfunctional reasoning will require as yet non-standard brain metrics, experimental designs, and analytical tools, and may ultimately help us understand and fine-tune the action of brain enhancers.

Acknowledgments

The author acknowledges the support of MINECO (FIS201238949-C03-01).

References

Acuna, B. D., Eliassen, J. C., Donoghue, J. P., and Sanes, J. N. (2002). Frontal and parietal lobe activation during transitive inference in humans. *Cereb. Cortex* 12, 1312–1321. doi: 10.1093/cercor/12.12.1312

Aharony, A., and Harris, A. B. (1996). Absence of self-averaging and universal fluctuations in random systems near critical points. *Phys. Rev. Lett.* 77, 3700–3703. doi: 10.1103/PhysRevLett.77.3700

Allegrini, P., Menicucci, D., Paradisi, P., and Gemignani, A. (2010). Fractal complexity in spontaneous EEG metastable-state transitions: new vistas on integrated neural dynamics. *Front. Physiol.* 1:128. doi: 10.3389/fphys.2010.00128

Bacry, E., Delour, J., and Muzy, J. F. (2001). Multifractal random walk. *Phys. Rev. E* 64:026103. doi: 10.1103/PhysRevE.64.026103

Baronchelli, A., and Radicchi, F. (2013). Lévy flights in human behaviour and cognition. *Chaos Solitons Fract.* 56, 101–105. doi: 10.1016/j.chaos.2013.07.013

Beck, C., and Schlögl, F. (1997). *Thermodynamics of Chaotic Systems: An Introduction*. Cambridge: Cambridge University press.

Beggs, J. M., and Plenz, D. (2003). Neuronal avalanches in neocortical circuits. *J. Neurosci.* 23, 11167–11177.

Beggs, J. M., and Plenz, D. (2004). Neuronal avalanches are diverse and precise activity patterns that are stable for many hours in cortical slice cultures. *J. Neurosci.* 24, 5216–5229. doi: 10.1523/JNEUROSCI.0540-04.2004

Bénichou, O., Loverdo, C., Moreau, M., and Voituriez, R. (2011). Intermittent search strategies. *Rev. Mod. Phys.* 83:81. doi: 10.1103/RevModPhys.83.81

Betzel, R. F., Erickson, M. A., Abell, M., O'Donnell, B. F., Hetrick, W. P., and Sporns, O. (2012). Synchronization dynamics and evidence for a repertoire of network states in resting EEG. *Front. Comput. Neurosci.* 6:74. doi: 10.3389/fncom.2012.00074

Bhattacharya, J. (2009). Increase of universality in human brain during mental imagery from visual perception. *PLoS ONE* 4:e4121. doi: 10.1371/journal.pone.0004121

Bianco, S., Ignaccolo, M., Rider, M. S., Ross, M. J., Winsor, P., and Grigolini, P. (2007). Brain, music, and non-poisson renewal processes. *Phys. Rev. E* 75:061911. doi: 10.1103/PhysRevE.75.061911

Bonnefond, M., Kaliuzhna, M., Van der Henst, J. B., and De Neys, W. (2014). Disabling conditional inferences: an EEG study. *Neuropsychologia* 56, 255–262. doi: 10.1016/j.neuropsychologia.2014.01.022

Bonnefond, M., Noveck, I., Baillet, S., Cheylus, A., Delpuech, C., Bertrand, O., et al. (2013). What MEG can reveal about reasoning: the case of if…then sentences. *Hum. Brain Mapp.* 34, 684–697. doi: 10.1002/hbm.21465

Bowden, E. M., and Jung-Beeman, M. (2007). Methods for investigating the neural components of insight. *Methods* 42, 87–99. doi: 10.1016/j.ymeth.2006.11.007

Bramwell, S. T., Holdsworth, P. C. W., and Pinton, J.-F. (1998). Universality of rare fluctuations in turbulence and critical phenomena. *Nature* 396, 552–554. doi: 10.1038/25083

Buiatti, M., Papo, D., Baudonnière, P. M., and van Vreeswijk, C. (2007). Feedback modulates the temporal scale-free dynamics of brain electrical activity in a hypothesis testing task. *Neuroscience* 146, 1400–1412. doi: 10.1016/j.neuroscience.2007.02.048

Bullmore, E. T., and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10, 186–198. doi: 10.1038/nrn2575

Bullmore, E. T., and Sporns, O. (2012). The economy of brain network organization. *Nat. Rev. Neurosci.* 13, 336–349. doi: 10.1038/nrn3214

Burov, S., and Barkai, E. (2008). Critical exponent of the fractional Langevin equation. *Phys. Rev. Lett.* 100:070601. doi: 10.1103/PhysRevLett.100.070601

Canessa, E. (2000). Multifractality in time series. *J. Phys. A Math. Gen.* 33, 3637–3651. doi: 10.1088/0305-4470/33/19/302

Codling, E. A., Plank, M. J., and Benhamou, S. (2008). Random walk models in biology. *J. R. Soc. Interface* 5, 813–834. doi: 10.1098/rsif.2008.0014

Contoyiannis, Y. F., and Eftaxias, K. A. (2008). Tsallis and Levy statistics in the preparation of an earthquake. *Nonlinear Process. Geophys.* 15, 379–388. doi: 10.5194/npg-15-379-2008

Cossart, R., Aronov, D., and Yuste, R. (2003). Attractor dynamics of network UP states in the neocortex. *Nature* 423, 283–288. doi: 10.1038/nature01614

Cugliandolo, L. F. (2011). The effective temperature. *J. Phys. A Math. Theor.* 44, 483001. doi: 10.1088/1751-8113/44/48/483001

de Arcangelis, L., and Herrmann, H. J. (1989). Scaling and multiscaling laws in random fuse networks. *Phys. Rev. B* 39:2678. doi: 10.1103/PhysRevB.39.2678

DeDeo, S., and Krakauer, D. C. (2012). Dynamics and processing in finite self-similar networks. *J. R. Soc. Interface* 9, 2131–2144. doi: 10.1098/rsif.2011.0840

Delvenne, J.-C., Lambiotte, R., and Rocha, L. E. C. (2013). Bottlenecks, burstiness, and fat tails regulate mixing times of non-poissonian random walks. arXiv:1309.4155.

Dragoi, G., and Tonegawa, S. (2011). Preplay of future place cell sequences by hippocampal cellular assemblies. *Nature* 469, 397–401. doi: 10.1038/nature09633

Expert, P., Lambiotte, R., Chialvo, D. R., Christensen, K., Jensen, H. J., Sharp, D. J., et al. (2010). Self-similar correlation function in brain resting-state functional magnetic resonance imaging. *J. R. Soc. Interface* 8, 472–479. doi: 10.1098/rsif.2010.0416

Fairhall, A. L., Lewen, G. D., Bialek, W., and de Ruyter van Steveninck, R. (2001). "Multiple timescales of adaptation in a neural code," in *Advances in Neural Information Processing Systems 13*, eds T. K. Leen, T. G. Dietterich, and V.Tresp (Cambridge, MA: MIT Press), 124–130.

Fiser, J., Chiu, C., and Weliky, M. (2004). Small modulation of ongoing cortical dynamics by sensory input during natural vision. *Nature* 431, 573–578. doi: 10.1038/nature02907

Fraiman, D., and Chialvo, D. R. (2012). What kind of noise is brain noise: anomalous scaling behavior of the resting brain activity fluctuations. *Front. Physiol.* 3:307. doi: 10.3389/fphys.2012.00307

Freeman, W. J., Holmes, M. D., Burke, B. C., and Vanhatalo, S. (2003). Spatial spectra of scalp EEG and EMG from awake humans. *Clin. Neurophysiol.* 114, 1053–1068. doi: 10.1016/S1388-2457(03)00045-2

Friedman, N., Ito, S., Brinkman, B. A., Shimono, M., Deville, R. E., Dahmen, K. A., et al. (2012). Universal critical dynamics in high resolution neuronal avalanche data. *Phys. Rev. Lett.* 108:208102. doi: 10.1103/PhysRevLett.108.208102

Friedrich, R., Peinke, J., Sahimi, M., and Reza Rahimi Tabar, M. (2011). Approaching complexity by stochastic methods: from biological systems to turbulence. *Phys. Rep*. 506, 87–162. doi: 10.1016/j.physrep.2011.05.003

Ghosh, A., Rho, Y., McIntosh, A. R., Kötter, R., and Jirsa, V. K. (2007). Noise during rest enables the exploration of the brain's dynamic repertoire. *PLoS Comput. Biol.* 4:e1000196. doi: 10.1371/journal.pcbi.1000196

Gilboa, G., Chen, R., and Brenner, N. (2005). History-dependent multiple-timescale dynamics in a single-neuron model. *J. Neurosci.* 25, 6479–6489. doi: 10.1523/JNEUROSCI.0763-05.2005

Gloning, K., and Hoff, H. (1969). "Cerebral localization of disorders of higher nervous activity," in *Handbook of Clinical Neurology, Vol. 3, Disorders of Higher Nervous Activity,* eds P. J. Vincken and G. N. Bruyn (New York, NY: Wiley).

Goel, V., Gold, B., Kapur, S., and Houle, S. (1997). The seats of reason? An imaging study of deductive and inductive reasoning. *Neuroreport* 8, 1305–1310.

Goel, V., Gold, B., Kapur, S., and Houle, S. (1998). Neuroanatomical correlates of human reasoning. *J. Cogn. Neurosci.* 10, 293–302. doi: 10.1162/089892998562744

Gong, P., Nikolaev, A. R., and van Leeuwen, C. (2002). Scale-invariant fluctuations of the dynamical synchronization in human brain electrical activity. *Neurosci. Lett*. 336, 33–36. doi: 10.1016/S0304-3940(02)01247-8

Gong, P., Nikolaev, A. R., and van Leeuwen, C. (2007). Intermittent dynamics underlying the intrinsic fluctuations of the collective synchronization patterns in electrocortical activity. *Phys. Rev. E* 76:011904. doi: 10.1103/PhysRevE.76.011904

Grigolini, P., Rocco, A., and West, B. J. (1999). Fractional calculus as a macroscopic manifestation of randomness. *Phys. Rev. E* 59, 2603–2613. doi: 10.1103/PhysRevE.59.2603

Houdé, O., Zago, L., Crivello, F., Moutier, S., Pineau, A., Mazoyer, B., et al. (2001). Access to deductive logic depends on a right ventromedial prefrontal area devoted to emotion and feeling: evidence from a training paradigm. *Neuroimage* 14, 1486–1492. doi: 10.1006/nimg.2001.0930

Humphries, N. E., Weimerskirch, H., Queiroz, N., Southall, E. J., and Sims, D. W. (2012). Foraging success of biological Lévy flights recorded *in situ*. *Proc. Natl. Acad. Sci. U.S.A.* 109, 7169–7174. doi: 10.1073/pnas.1121201109

Ikegaya, Y., Aaron, G., Cossart, R., Aronov, D., Lampl, I., Ferster, D., et al. (2004). Synfire chains and cortical songs: temporal modules of cortical activity. *Science* 304, 559–564. doi: 10.1126/science.1093173

Jung-Beeman, M., Bowden, E. M., Haberman, J., Frymiare, J. L., Arambel-Liu, S., Greenblatt, R., et al. (2004). Neural activity when people solve verbal problems with insight. *PLoS Biol.* 2:E97. doi: 10.1371/journal.pbio.0020097

Kapiris, P. G., Eftaxias, K. A., and Chelidze, T. L. (2004). Electromagnetic signature of prefracture criticality in heterogeneous media. *Phys. Rev. Lett*. 92:065702. doi: 10.1103/PhysRevLett.92.065702

Kaplan, A. Y., Fingelkurts, A. A., Fingelkurts, A. A., Borisov, B. S., and Darkhovsky, B. S. (2005). Nonstationary nature of the brain activity as revealed by EEG/MEG: methodological, practical and conceptual challenges. *Signal Process.* 85, 2190–2212. doi: 10.1016/j.sigpro.2005.07.010

Kenet, T., Bibitchkov, D., Tsodyks, M., Grinvald, A., and Arieli, A. (2003). Spontaneously emerging cortical representations of visual attributes. *Nature* 425, 954–956. doi: 10.1038/nature02078

Knoblich, G., Ohlsson, S., Haider, H., and Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 1534–1556. doi: 10.1037/0278-7393.25.6.1534

Kounios, J., Fleck, J. I., Green, D. L., Payne, L., Stevenson, J. L., Bowden, E. M., et al. (2008). The origins of insight in resting-state brain activity. *Neuropsychologia* 46, 281–291. doi: 10.1016/j.neuropsychologia.2007.07.013

Kounios, J., Frymiare, J. L., Bowden, E. M., Fleck, J. I., Subramaniam, K., Parrish, T. B., et al. (2006). The prepared mind: neural activity prior to problem presentation predicts subsequent solution by sudden insight. *Psychol. Sci.* 17, 882–890. doi: 10.1111/j.1467-9280.2006.01798.x

Kroger, J. K., Sabb, F. W., Fales, C. L., Bookheimer, S. Y., Cohen, M. S., and Holyoak, K. J. (2002). Recruitment of anterior dorsolateral prefrontal cortex in human reasoning: a parametric study of relational complexity. *Cereb. Cortex* 12, 477–485. doi: 10.1093/cercor/12.5.477

Lambiotte, R., Tabourier, L., and Delvenne, J.-C. (2013). Burstiness and spreading on temporal networks. *Eur. Phys. J. B* 86, 320. doi: 10.1140/epjb/e2013-40456-9

Lang, S., Kanngieser, N., Jaśkowski, P., Haider, H., Rose, M., and Verleger, R. (2006). Precursors of insight in event-related brain potentials. *J. Cogn. Neurosci.* 18, 2052–2066. doi: 10.1162/jocn.2006.18.12.2152

Lesne, A. (2008). "Regularization, renormalization, and renormalization groups: relationships and epistemological aspects," in *Vision of Oneness*, eds I. Licata and A. Sakaji (Roma, QL: Aracne), 121–154.

Lezak, M. D. (1995). *Neuropsychological Assessment, 3rd Edn*. Oxford: Oxford University Press.

Linkenkaer-Hansen, K., Nikouline, V. V., Palva, J. M., and Ilmoniemi, R. (2001). Long-range temporal correlations and scaling behavior in human oscillations. *J. Neurosci.* 15, 1370–1377.

Lomholt, M., Tal, K., Metzler, R., and Joseph, K. (2008). Lévy strategies in intermittent search processes are advantageous. *Proc. Natl. Acad. Sci. U.S.A.* 105, 11055–11059. doi: 10.1073/pnas.0803117105

Luczak, A., Barthó, P., and Harris, K. D. (2009). Spontaneous events outline the realm of possible sensory responses in neocortical populations. *Neuron* 62, 413–425. doi: 10.1016/j.neuron.2009.03.014

Lundstrom, B. N., Higgs, M. H., Spain, W. J., and Fairhall, A. L. (2008). Fractional differentiation by neocortical pyramidal neurons. *Nat. Neurosci.* 11, 1335–13342. doi: 10.1038/nn.2212

Luo, J., Niki, K., and Phillips, S. (2004). Neural correlates of the 'Aha! reaction'. *Neuroreport* 15, 2013–2017. doi: 10.1097/00001756-200409150-00004

Mai, X.-Q., Luo, J., Wu, J.-H., and Luo, Y.-J. (2004). "Aha!" Effects in a guessing riddle task: an event-related potential study. *Hum. Brain Mapp.* 22, 261–270. doi: 10.1002/hbm.20030

Manshour, P., Saberi, S., Sahimi, M., Peinke, J., Pacheco, A. F., and Rahimi Tabar, M. R. (2009). Turbulencelike behavior of seismic time series. *Phys. Rev. Lett.* 102:014101. doi: 10.1103/PhysRevLett.102.014101

Moshman, D. (1995). Reasoning as self-constrained thinking. *Hum. Dev.* 38, 53–64. doi: 10.1159/000278299

Noveck, I. A., Goel, V., and Smith, K. W. (2004). The neural basis of conditional reasoning with arbitrary content. *Cortex* 40, 613–622. doi: 10.1016/S0010-9452(08)70157-6

Novikov, E., Novikov, A., Shannahoff-Khalsa, D., Schwartz, B., and Wright, J. (1997). Scale-similar activity in the brain. *Phys. Rev. E* 56, R2387–R2389. doi: 10.1103/PhysRevE.56.R2387

Olemskoi, A., and Kokhan, S. (2006). Effective temperature of self-similar time series: analytical and numerical developments. *Phys. A* 360, 37–58. doi: 10.1016/j.physa.2005.06.048

Osherson, D., Perani, D., Cappa, S., Schnur, T., Grassi, F., and Fazio, F. (1998). Distinct brain foci in deductive versus probabilistic reasoning. *Neuropsychologia* 36, 369–376. doi: 10.1016/S0028-3932(97)00099-7

Osorio, I., Frei, M. G., Sornette, D., Milton, J., and Lai, Y. C. (2010). Epileptic seizures: quakes of the brain? *Phys. Rev. E* 82:021919. doi: 10.1103/PhysRevE.82.021919

Papo, D. (2013a). Brain temperature: what it means and what it can do for (cognitive) neuroscientists. arXiv:1310.2906v1.

Papo, D. (2013b). Time scales in cognitive neuroscience. *Front. Physiol.* 4:86. doi: 10.3389/fphys.2013.00086

Papo, D. (2014a). Functional significance of complex fluctuations in brain activity: from resting state to cognitive neuroscience. *Front. Syst. Neurosci.* 8:112. doi: 10.3389/fnsys.2014.00112

Papo, D. (2014b). Measuring brain temperature without a thermometer. *Front. Physiol.* 5, 124. doi: 10.3389/fphys.2014.00124

Papo, D., Douiri, A., Bouchet, F., Bourzeix, J.-C., Caverni, J.-P., and Baudonnière, P.-M. (2007). Time-frequency intracranial source localization of feedback-related EEG activity in hypothesis testing. *Cereb. Cortex* 17, 1314–1322. doi: 10.1093/cercor/bhl042

Papo, D., Zanin, M., Pineda, J. A., Boccaletti, S., and Buldú, J. M. (2014). Brain networks: great expectations, hard times, and the big leap forward. *Philos. Trans. R. Soc. B Biol. Sci.* 369, 20130525. doi: 10.1098/rstb.2013.0525

Parrondo, M. R., Horowitz, J. M., and Sagawa, T. (2015). Thermodynamics of information. *Nat. Phys.* 11, 131–139. doi: 10.1038/nphys3230

Parsons, L. M., and Osherson, D. (2001). New evidence for distinct right and left brain systems for deductive versus probabilistic reasoning. *Cereb. Cortex* 11, 954–965. doi: 10.1093/cercor/11.10.954

Petri, G., Expert, P., Turkheimer, F., Carhart-Harris, R., Nutt, D., Hellyer, J., et al. (2014). Homological scaffolds of brain functional networks. *J. R. Soc. Interface* 11:20140873. doi: 10.1098/rsif.2014.0873

Prado, J., Chadha, A., and Booth, J. R. (2011). The brain network for deductive reasoning: a quantitative meta-analysis of 28 neuroimaging studies. *J. Cogn. Neurosci.* 23, 3483–3497. doi: 10.1162/jocn_a_00063

Qiu, J., Li, H., Yang, D., Luo, Y., Li, Y., Wu, Z., et al. (2008). The neural basis of insight problem solving: an event-related potential study. *Brain Cogn.* 68, 100–106. doi: 10.1016/j.bandc.2008.03.004

Rabinovich, M., Huerta, R., and Laurent, G. (2008). Transient dynamics for neural processing. *Science* 321, 48–50. doi: 10.1126/science.1155564

Rebenshtok, A., and Barkai, E. (2007). Distribution of time-averaged observables for weak ergodicity breaking. *Phys. Rev. Lett.* 99:210601. doi: 10.1103/PhysRevLett.99.210601

Reiner, M. (1964). The Deborah number. *Phys. Today* 17, 62. doi: 10.1063/1.3051374

Reverberi, C., Bonatti, L. L., Frackowiak, R. S., Paulesu, E., Cherubini, P., and Macaluso, E., (2012). Large scale brain activations predict reasoning profiles. *Neuroimage* 59, 1752–1764. doi: 10.1016/j.neuroimage.2011.08.027

Rhodes, T., and Turvey, M. T. (2007). Human memory retrieval as Lévy foraging. *Phys. A* 385, 255–260. doi: 10.1016/j.physa.2007.07.001

Ruelle, D. (1978). *Thermodynamic Formalism*. Reading, MA: Addison Wesley Publ. Co.

Sandkühler, S., and Bhattacharya, J. (2008). Deconstructing insight: EEG correlates of insightful problem solving. *PLoS ONE* 3:e1459. doi: 10.1371/journal.pone.0001459

Sekimoto, K. (1998). Langevin equation and thermodynamics. *Prog. Theor. Phys. Suppl.* 130, 17–27. doi: 10.1143/PTPS.130.17

Shalizi, C. R., and Moore, C. (2003). What is a macrostate? Subjective observations and objective dynamics. arXiv:cond-mat/0303625v1.

Sheth, B. R., Sandkühler, S., and Bhattacharya, J. (2008). Posterior beta and anterior gamma oscillations predict cognitive insight. *J. Cogn. Neurosci.* 21, 1269–1279. doi: 10.1162/jocn.2009.21069

Shlesinger, M., Zaslavsky, G., and Klafter, J. (1993). Strange kinetics. *Nature* 363, 31–37. doi: 10.1038/363031a0

Sornette, D. (2002). Predictability of catastrophic events: material rupture, earthquakes, turbulence, financial crashes, and human birth. *Proc. Natl. Acad. Sci. U.S.A.* 99, 2522–2529. doi: 10.1073/pnas.022581999

Sporns, O., Tononi, G., and Edelman, G. M. (2000). Connectivity and complexity: the relationship between neuroanatomy and brain dynamics. *Neural Netw.* 13, 909–922. doi: 10.1016/S0893-6080(00)00053-8

Stam, C. J., and de Bruin, E. A. (2004). Scale-free dynamics of global functional connectivity in the human brain. *Hum. Brain Mapp.* 22, 97–109. doi: 10.1002/hbm.20016

Stephen, D. G., Boncoddo, R. A., Magnuson, J. S., and Dixon, J. A. (2009). The dynamics of insight: mathematical discovery as a phase transition. *Mem. Cogn.* 37, 1132–1149. doi: 10.3758/MC.37.8.1132

Subramaniam, K., Kounios, J., Parrish, T. B., and Jung-Beeman, M. (2008). A brain mechanism for facilitation of insight by positive affect. *J. Cogn. Neurosci.* 21, 415–432. doi: 10.1162/jocn.2009.21057

Suckling, J., Wink, A. M., Bernard, F. A., Barnes, A., and Bullmore, E. (2009). Endogenous multifractal brain dynamics are modulated by age, cholinergic blockade and cognitive performance. *J. Neurosci. Methods* 174, 292–300. doi: 10.1016/j.jneumeth.2008.06.037

van de Ville, D., Britz, J., and Michel, C. M. (2010). EEG microstate sequences in healthy humans at rest reveal scale-free dynamics. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18179–18184. doi: 10.1073/pnas.1007841107

Viswanathan, G. M., da Luz, M. G. E., Raposo, E. P., and Stanley, H. E. (2011). *The Physics of Foraging: an Introduction to Random Searches and Biological Encounters*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511902680

Zaslavsky, G. M. (2002). Chaos, fractional kinetics, and anomalous transport. *Phys. Rep.* 371, 461–580. doi: 10.1016/S0370-1573(02)00331-9

Zilber, N., Ciuciu, P., Abry, P., and van Wassenhove, V. (2012). Modulation of scale-free properties of brain activity in MEG. *IEEE I. S. Biomed. Imaging (Barcelona)* 1531–1534. doi: 10.1109/ISBI.2012.6235864

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.

frontiers in
HUMAN NEUROSCIENCE

OPINION ARTICLE
published: 03 February 2015
doi: 10.3389/fnhum.2015.00041

Investigating reasoning with multiple integrated neuroscientific methods

Matthew E. Roser*[1]*, *Jonathan St. B. T. Evans*[1], *Nicolas A. McNair*[2], *Giorgio Fuggetta*[3], *Simon J. Handley*[1], *Lauren S. Carroll*[1] and *Dries Trippas*[4]

[1] *School of Psychology, Plymouth University, Plymouth, UK*
[2] *School of Psychology, The University of Sydney, Sydney, NSW, Australia*
[3] *School of Psychology, The University of Leicester, Leicester, UK*
[4] *Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany*
*Correspondence: *matt.roser@plymouth.ac.uk*

Edited by:
Ira Andrew Noveck, Centre Nationale de la Recherche Scientifique, France

Reviewed by:
David Papo, Universidad Politecnica de Madrid, Spain
Wim De Neys, Centre National de la Recherche Scientifique, France

Keywords: reasoning, neuroscience methods, methodological integration, brain connectivity, neuronavigated TMS

Recent years have seen increased application of functional MRI (fMRI), transcranial magnetic stimulation (TMS) and event-related potentials (ERP), to questions of human rationality. This has both illuminated the brain bases of these functions and contributed to theoretical advances (Goel, 2007; Prado et al., 2008). Most studies have, however, employed only one method and the developing literatures run somewhat parallel with only informal integration of results across methods. Results from other fields (Sarfeld et al., 2012) demonstrate the potential benefits of integration of multiple neuroscientific methods within studies of human reasoning, allowing findings from one method to influence the application of other methods, or constrain the interpretation of data derived therefrom. Including data on regional brain volume, structural and functional connectivity, individual differences and development and aging is particularly appropriate to the study of neural mechanisms of human reasoning, which are likely to be formed from networks of numerous widely-distributed brain regions. Here we briefly describe how the integration of several neuroscientific methods within a single study may advance investigations of the reasoning brain.

fMRI has now been applied to a large number of reasoning paradigms (Goel, 2007). Consideration of what appears initially as a disparate set of brain activations reveals consistencies suggestive of several underlying neural systems. A formal analysis (Prado et al., 2011) of 28 studies found similar consistency of activation across studies and reasoning paradigms, but no monolithic neural system for reasoning. Instead, a collection of subsystems incorporating widely distributed areas of the brain is apparent. This widespread activation, encompassing frontal and posterior areas, in response to high-level tasks with long processing times complicates interpretation.

Approaches which move beyond mapping the spatial extent of activation to consider the quality of brain activity seen in separate regions promise to clarify the distributed-network nature of the reasoning brain. Analyses may focus on the time-courses of activation within brain regions (Rodriguez-Moreno and Hirsch, 2009), identifying subsets of regions involved at different stages of reasoning, or, as in our current research (ESRC Grant RES-062-23-3285), the correlation in the degree of activation seen in separate clusters with individual differences (Reverberi et al., 2012).

Formal analyses of functional connectivity, or correlated activity (Friston, 2011), between brain regions active during the resting state have revealed the effects of prolonged practice on a reasoning task (Mackey et al., 2013). The application of functional-connectivity analyses to brain activity elicited by reasoning, rather than rest, awaits. While many imaging studies of reasoning speak of the "networks" involved it would be more accurate to speak of distributed regions of task-related activation as no studies have formally tested functional connectivity between regions. This is in contrast to other areas, such as research in memory, attention and task control, in which functional-connectivity analyses are commonplace and have greatly advanced the characterization of implicated brain networks (Vincent et al., 2008). Functional-connectivity analyses have the potential to further clarify how subsets of the numerous regions found active in fMRI studies of reasoning group together to form dynamic networks that are reconfigured across extended periods of reasoning-task performance.

A further step is analysis of effective connectivity in which causal networks of distributed regions are modeled and tested against observed data (Friston, 2011). Models incorporate information about brain *structural* connectivity into predictions of inter-regional *functional* connectivity. These structural data have traditionally come from monkey section studies but human diffusion-tensor imaging (DTI) data are now being used, as described in a recent survey of methods and applications for fusing fMRI and DTI data (Zhu et al., 2014). DTI is a MRI technique which allows the microstructural connectivity of brain tissue to be probed (Le Bihan, 2003). The data can be acquired in a scan lasting only around 10 min, which could feasibly be included in a fMRI study. DTI data have informed researchers about the constraining effect of structural connectivity upon functional connectivity in non-reasoning tasks (Honey et al., 2009). Structural-connectivity maps of direct and indirect connections between brain regions were tested as predictors of resting-state inter-regional functional connectivity, leading to a model in which functional connectivity is determined by a combination of direct and indirect structural connections. Ultimately, the integration of fMRI and DTI datasets could allow the development of richer models of dynamic networks of distributed brain regions supporting reasoning performance. Putative networks of brain regions activated by reasoning tasks may be merely regions of correlated activity that do not exist in a causative relationship, or they may be comprised of two or more overlapping and commonly-activated sub-networks. These possibilities can be tested using models informed by integrated methods.

The further integration of a developmental or aging perspective, to which DTI is sensitive (Sullivan and Pfefferbaum, 2006), would allow the organization and degeneration of brain structural connectivity, and its role in supporting reasoning, to be traced over the lifespan. Information on brain regional and connective development and degeneration is of great relevance to a growing literature (Salthouse, 2005) of age effects on reasoning. The anterior to posterior progression of degeneration in the aged brain, apparent in DTI studies (Sullivan and Pfefferbaum, 2006), predicts that reasoning processes that draw heavily on frontal support will be more affected by age than are reasoning processes that primarily involve posterior regions. Also of relevance is information about brain regional volume, as assessed by MRI, which has been shown to be abnormal in some populations, such as people with autism (Mcalonan et al., 2005; Redcay and Courchesne, 2005), who are also of interest to investigators of reasoning (Mckenzie et al., 2010; Morsanyi and Holyoak, 2010).

The incorporation of structural and functional MRI into studies of reasoning using repetitive TMS has promise to increase the power and accuracy of a technique which can probe the causal relationship between brain activity and reasoning performance. Previous rTMS studies (Tsujii et al., 2010, 2011) guided stimulation using structural MRI but selected cortical targets somewhat arbitrarily from a set of areas implicated in fMRI studies. An improvement is to integrate results from an fMRI study using the same paradigm and stimuli to target specific locations found to be functionally active. As considerable variation in reasoning-associated activation across studies using similar, but non-identical, paradigms, and stimuli has been observed (Goel, 2007) the targeting of specific areas activated by specific experimental designs is important. We (ESRC Grant RES-062-23-3285) are doing this by warping the standard-space group-analysis results from our fMRI study of conditional reasoning into the individual TMS-subject space to identify functionally-relevant targets. Furthermore, using a within-trial, short-burst rTMS paradigm (Fuggetta et al., 2008), allows greater temporal specificity in rTMS application. By disrupting activity in ventral and dorsal prefrontal cortex at different stages of conditional-reasoning trials we predict a double dissociation of the effect of rTMS on belief bias at the two locations over the two stages of the trial. This result would advance our understanding of the processes involved in conditional reasoning, and of the roles of the two brain regions, and is an example of how method integration might inform psychological theory.

ERP studies of reasoning differ in the degree to which they preserve the traditional behavioral paradigms (Qiu et al., 2009; Luo et al., 2013), which typically involve extended reading, and the temporal specificity with which they are able to resolve reasoning processes by adapting orthodox paradigms shown to elicit well-defined ERPs (Prado et al., 2008; Banks and Hope, 2014). Despite this heterogeneity, evidence is accumulating that ERPs and oscillatory activity associated with expectation and inhibition are modulated by performance on reasoning tasks (Bonnefond and Van der Henst, 2009; Bonnefond et al., 2014). Initial steps to identify the neural sources of observed ERPs (Qiu et al., 2009; Luo et al., 2013) could be greatly improved by using results from fMRI studies to constrain the fitting of source models. The ultimate aim is to conduct simultaneous recordings of EEG and fMRI (Baumeister et al., 2014), illuminating sequential activations across distributed networks, as are revealed by the less-available technique of magnetoencephalography (Bonnefond et al., 2013).

A full characterization of the reasoning brain will require models that describe functional connectivity between widespread brain regions, constrained and shaped by structural connectivity, which varies between and within individuals across time and space. This implies a conceptualization of the reasoning brain as a spatially-extended dynamical system. Models of this type will necessarily integrate data derived from many different methods and may require mathematical tools not previously applied to investigations of reasoning (Siegelmann, 2010). At present most of these techniques are being applied to the study of the reasoning brain, but in a parallel fashion. The lesson from other areas of investigation (Calhoun and Lemieux, 2014) is that their integration can yield more than the sum of their parts.

ACKNOWLEDGMENTS

This work was supported by the Economic and Social Research Council Grant RES-062-23-3285. Dual processes in reasoning: A neuropsychological study of the role of working memory.

REFERENCES

Banks, A. P., and Hope, C. (2014). Heuristic and analytic processes in reasoning: an event-related potential study of belief bias. *Psychophysiology* 51, 290–297. doi: 10.1111/psyp.12169

Baumeister, S., Hohmann, S., Wolf, I., Plichta, M. M., Rechtsteiner, S., Zangl, M., et al. (2014). Sequential inhibitory control processes assessed through simultaneous EEG–fMRI. *Neuroimage* 94, 349–359. doi: 10.1016/j.neuroimage.2014.01.023

Bonnefond, M., and Van der Henst, J.-B. (2009). What's behind an inference? An EEG study with conditional arguments. *Neuropsychologia* 47, 3125–3133. doi: 10.1016/j.neuropsychologia.2009.07.014

Bonnefond, M., Mariia, K., Van der Henst, J.-B., and De Neys, W. (2014). Disabling conditional inferences: an EEG study. *Neuropsychologia* 56, 255–262. doi: 10.1016/j.neuropsychologia.2014.01.022

Bonnefond, M., Noveck, I., Baillet, S., Cheylus, A., Delpuech, C., Bertrand, O., et al. (2013). What MEG can reveal about inference making: the case of if... then sentences. *Hum. Brain Mapp.* 34, 684–697. doi: 10.1002/hbm.21465

Calhoun, V. D., and Lemieux, L. (2014). Neuroimage: special issue on multimodal data fusion. *Neuroimage* 102, 1–2. doi: 10.1016/j.neuroimage.2014.04.070

Friston, K. J. (2011). Functional and effective connectivity: a review. *Brain Connect.* 1, 13–36. doi: 10.1089/brain.2011.0008

Fuggetta, G., Pavone, E. F., Fiaschi, A., and Manganotti, P. (2008). Acute modulation of cortical oscillatory activities during short trains of high-frequency repetitive transcranial magnetic stimulation of the human motor cortex: a combined EEG and TMS study. *Hum. Brain Mapp.* 29, 1–13. doi: 10.1002/hbm.20371

Goel, V. (2007). Anatomy of deductive reasoning. *Trends Cogn. Sci.* 11, 435–441. doi: 10.1016/j.tics.2007.09.003

Honey, C., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J.-P., Meuli, R., et al. (2009). Predicting human resting-state functional connectivity from structural connectivity. *Proc. Natl. Acad. Sci. U.S.A.* 106, 2035–2040. doi: 10.1073/pnas.0811168106

Le Bihan, D. (2003). Looking into the functional architecture of the brain with diffusion MRI. *Nat. Rev. Neurosci.* 4, 469–480. doi: 10.1038/nrn1119

Luo, J., Liu, X., Stupple, E. J., Zhang, E., Xiao, X., Jia, L., et al. (2013). Cognitive control in belief-laden reasoning during conclusion processing: an ERP study. *Int. J. Psychol.* 48, 224–231. doi: 10.1080/00207594.2012.677539

Mackey, A. P., Singley, A. T. M., and Bunge, S. A. (2013). Intensive reasoning training alters patterns of brain connectivity at rest. *J. Neurosci.* 33, 4796–4803. doi: 10.1523/JNEUROSCI.4141-12.2013

Mcalonan, G. M., Cheung, V., Cheung, C., Suckling, J., Lam, G. Y., Tai, K., et al. (2005). Mapping the brain in autism. A voxel-based MRI study of volumetric differences and intercorrelations in autism. *Brain* 128, 268–276. doi: 10.1093/brain/awh332

Mckenzie, R., Evans, J. S. B., and Handley, S. J. (2010). Conditional reasoning in autism: activation and integration of knowledge and belief. *Dev. Psychol.* 46, 391. doi: 10.1037/a0017412

Morsanyi, K., and Holyoak, K. J. (2010). Analogical reasoning ability in autistic and typically developing children. *Dev. Sci.* 13, 578–587. doi: 10.1111/j.1467-7687.2009.00915.x

Prado, J., Chadha, A., and Booth, J. R. (2011). The brain network for deductive reasoning: a quantitative meta-analysis of 28 neuroimaging studies. *J. Cogn. Neurosci.* 23, 3483–3497. doi: 10.1162/jocn_a_00063

Prado, J., Kaliuzhna, M., Cheylus, A., and Noveck, I. A. (2008). Overcoming perceptual features in logical reasoning: an event-related potentials study. *Neuropsychologia* 46, 2629–2637. doi: 10.1016/j.neuropsychologia.2008.04.017

Qiu, J., Li, H., Luo, Y., Zhang, Q., and Tu, S. (2009). The neural basis of syllogistic reasoning: an event-related potential study. *Brain Res.* 1273, 106–113. doi: 10.1016/j.brainres.2009.03.054

Redcay, E., and Courchesne, E. (2005). When is the brain enlarged in autism? A meta-analysis of all brain size reports. *Biol. Psychiatry* 58, 1–9. doi: 10.1016/j.biopsych.2005.03.026

Reverberi, C., Bonatti, L. L., Frackowiak, R. S., Paulesu, E., Cherubini, P., and Macaluso, E. (2012). Large scale brain activations predict reasoning profiles. *Neuroimage* 59, 1752–1764. doi: 10.1016/j.neuroimage.2011.08.027

Rodriguez-Moreno, D., and Hirsch, J. (2009). The dynamics of deductive reasoning: an fMRI investigation. *Neuropsychologia* 47, 949–961. doi: 10.1016/j.neuropsychologia.2008.08.030

Salthouse, T. A. (2005). "Effects of Aging on Reasoning," in *The Cambridge Handbook of Thinking and Reasoning*, eds K. J. Holyoak and R. G. Morrison (New York, NY: Cambridge University Press), 589–605.

Sarfeld, A. S., Diekhoff, S., Wang, L. E., Liuzzi, G., Uludağ, K., Eickhoff, S. B., et al. (2012). Convergence of human brain mapping tools: neuronavigated TMS parameters and fMRI activity in the hand motor area. *Hum. Brain Mapp.* 33, 1107–1123. doi: 10.1002/hbm.21272

Siegelmann, H. T. (2010). Complex systems science and brain dynamics. *Front. Comput. Neurosci.* 4:7. doi: 10.3389/fncom.2010.00007

Sullivan, E. V., and Pfefferbaum, A. (2006). Diffusion tensor imaging and aging. *Neurosci. Biobehav. Rev.* 30, 749–761. doi: 10.1016/j.neubiorev.2006.06.002

Tsujii, T., Masuda, S., Akiyama, T., and Watanabe, S. (2010). The role of inferior frontal cortex in belief-bias reasoning: an rTMS study. *Neuropsychologia* 48, 2005–2008. doi: 10.1016/j.neuropsychologia.2010.03.021

Tsujii, T., Sakatani, K., Masuda, S., Akiyama, T., and Watanabe, S. (2011). Evaluating the roles of the inferior frontal gyrus and superior parietal lobule in deductive reasoning: an rTMS study. *Neuroimage* 58, 640–646. doi: 10.1016/j.neuroimage.2011.06.076

Vincent, J. L., Kahn, I., Snyder, A. Z., Raichle, M. E., and Buckner, R. L. (2008). Evidence for a frontoparietal control system revealed by intrinsic functional connectivity. *J. Neurophysiol.* 100, 3328–3342. doi: 10.1152/jn.90355.2008

Zhu, D., Zhang, T., Jiang, X., Hu, X., Chen, H., Yang, N., et al. (2014). Fusing DTI and fMRI data: a survey of methods and applications. *Neuroimage* 102, 184–191. doi: 10.1016/j.neuroimage.2013.09.071

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

Received: 29 October 2014; accepted: 16 January 2015; published online: 03 February 2015.

*Citation: Roser ME, Evans JSBT, McNair NA, Fuggetta G, Handley SJ, Carroll LS and Trippas D (2015) Investigating reasoning with multiple integrated neuroscientific methods. Front. Hum. Neurosci. **9**:41. doi: 10.3389/fnhum.2015.00041*

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

frontiers in
HUMAN NEUROSCIENCE

MINI REVIEW ARTICLE
published: 09 January 2015
doi: 10.3389/fnhum.2014.01056

Brain imaging, forward inference, and theories of reasoning

Evan Heit*

School of Social Sciences, Humanities and Arts, University of California Merced, Merced, CA, USA

Edited by:
Jérôme Prado, Centre National de la Recherche Scientifique, France

Reviewed by:
Martin M. Monti, University of California Los Angeles, USA
Sandrine Rossi, Université de Caen Basse-Normandie, France
Ira Andrew Noveck, Centre National de la Recherche Scientifique, France

****Correspondence:***
Evan Heit, School of Social Sciences, Humanities and Arts, University of California, 5200 North Lake Road, Merced, CA 95343, USA
e-mail: eheit@ucmerced.edu

This review focuses on the issue of how neuroimaging studies address theoretical accounts of reasoning, through the lens of the method of forward inference (Henson, 2005, 2006). After theories of deductive and inductive reasoning are briefly presented, the method of forward inference for distinguishing between psychological theories based on brain imaging evidence is critically reviewed. Brain imaging studies of reasoning, comparing deductive and inductive arguments, comparing meaningful versus non-meaningful material, investigating hemispheric localization, and comparing conditional and relational arguments, are assessed in light of the method of forward inference. Finally, conclusions are drawn with regard to future research opportunities.

Keywords: reasoning, neuroimaging, deduction, induction, forward inference

How can neuroimaging techniques help address theoretical questions in reasoning research? To be more specific, how can techniques such as functional magnetic resonance imaging (fMRI) help researchers distinguish between psychological theories of reasoning? There have been thousands of behavioral experiments on reasoning, and the field as a whole has several competing theories without a consensus of which one best account for the behavioral data. Potentially, new evidence on patterns of brain activity during reasoning tasks could help resolve these long-standing debates.

This article will first briefly outline several psychological theories of deductive and inductive reasoning. Next, a particular method (forward inference; Henson, 2005, 2006) for using neuroimaging data to test predictions from psychological theories will be critically discussed. Then, example neuroimaging studies of deductive and inductive reasoning will be reviewed, through the lens of the method of forward inference. By no means is forward inference the only possible means to advance psychological theory in the context of neuroimaging. This exercise will provide some perspective both on neuroimaging studies of reasoning and on the method of forward inference.

THEORIES OF REASONING

Researchers have studied reasoning on both problems of deduction and problems of induction. Problems of deduction require drawing a valid, logical conclusion that must follow based on a set of given premises. In contrast, problems of induction require drawing probabilistic conclusions from given information as well as other relevant knowledge (Heit, 2007; Hayes et al., 2010). One open question in reasoning research is whether deduction and induction simply refer to two different kinds of reasoning problems – in terms of the structure and/or content of the problems themselves – or if there are truly two different kinds of reasoning, deductive reasoning and inductive reasoning, with different cognitive processes (or different mixtures of cognitive processes) involved (Rotello and Heit, 2009; Heit and Rotello, 2010; Heit et al., 2012).

According to dual-process accounts [e.g., Kahneman (2011), Evans and Stanovich (2013)], there are two kinds of underlying mechanisms, heuristic processing and analytic processing. Both induction and deduction could be influenced by these two processes, but in different mixtures (Rotello and Heit, 2009; Heit and Rotello, 2010; Heit et al., 2012). Under this mixture account, induction judgments could be particularly influenced by heuristic processes that tap into associations and knowledge that do not necessarily make an argument logically valid. In contrast, deduction judgments could be more heavily influenced by slower analytic processes that encompass more deliberative, and typically more accurate, reasoning. However, for present purposes, the crucial point is that there are two processes, not the details of any possible mixture.

In comparison, single-process accounts explain reasoning in terms of a common set of mechanisms across multiple forms of reasoning, although typically these theories focus more on either deduction or induction. Mental model theory (Johnson-Laird, 1994) asserts that a reasoner assesses an argument by constructing a visuospatial model of the premises then looking for counterexamples. Although this theory is typically applied to problems of deduction, it has also been applied to problems of induction. Bayesian accounts of reasoning address performance on problems of deduction in terms of making probabilistic judgments (Oaksford and Chater, 2007); hence, they are inductive in nature. Indeed, related models of inductive reasoning are also Bayesian in nature (Heit, 1998; Tenenbaum and Griffiths, 2001). Additionally, there are some models of inductive reasoning (Osherson et al., 1990; Sloman, 1993) that focus on problems of induction but can address performance on some problems of deduction as well. Finally, mental logic theory (Rips, 1994; Braine and O'Brien, 1998) has focused on deduction, asserting that people reason on problems of deduction by carrying out syntactic operations using a system of logical rules.

DRAWING THEORETICAL INFERENCES

Although there has been skepticism about drawing inferences about psychological theories from neuroimaging data [e.g., Coltheart (2006), Harley (2004), Uttal (2011), Van Orden and Paap (1997)]. Henson (2005, 2006) has outlined a rationale for doing so, adopting standard notions from experimental psychology on employing behavioral data. Henson (2006) referred to this process as "forward inference," namely, "the use of qualitatively different patterns of activity over the brain to distinguish between competing cognitive theories." The key idea is that if theory 1 predicts that the same cognitive processes underlie two different experimental tasks, and theory 2 predicts that the tasks differ in terms of at least one cognitive process, then theory 2 will be supported when patterns of brain activity differ between the two tasks. This inference depends on the assumption that there is at least some systematic mapping between cognitive processes and brain regions, namely, the weak assumption that within the experimental comparison of interest, the same cognitive process is not supported by different brain regions.

Forward inference itself has some limitations, such as its asymmetrical nature, that is, theory 1 can be supported by null results, whereas theory 2 could potentially be supported numerous differences. Also, as Henson (2006) noted, forward inferences are theory-dependent, namely, theories 1 and 2 may both be incorrect, and some alternative account such as theory 3 may be correct. If that alternative is not considered by the researcher, then forward inferences based on theories 1 and 2 will be misleading. Another pitfall is that there can be other reasons for differences in localization, namely, if two experimental tasks differ in patterns of brain activity, the reason may not be differences in cognitive processes but differences in rate of responding "yes" [Nosofsky et al. (2012); for a related argument, involving task complexity, see Johnson (1993)]. Going beyond the issue of which regions are activated is the matter of how these activations are causally related to each other [e.g., Chiong et al. (2013)]. In general, as Monti and Osherson (2012) point out, reasoning "should be regarded as a collection of processes and representations" [cf., Anderson (1978)], hence observed differences may correspond not to processing differences but differences in the content being processed.

A more fundamental problem for forward inference is that the theories of interest simply may not make predictions about brain activity. In Marr's (1982) terms, the theories may be at the algorithmic or computational level of description, without strong connections to the implementation level. Henson (2005) was optimistic, however, that brain imaging could either directly address the algorithmic level of processing or do so indirectly, by illuminating the implementation level which itself would constrain the algorithmic level.

A companion article to Henson (2006), by Poldrack (2006), described "reverse inference," by which the presence of a particular cognitive process is inferred from a pattern of brain activity [see Del Pinal and Nathan (2013), for a critical review]. Poldrack noted that a researcher's confidence in a reverse inference can be explained in terms of Bayes's Theorem, with the conditional probability that the cognitive process is engaged when a particular brain region is activated depending, in part, on the prior likelihood that cognitive process appears in the experimental context. Put another way, if the cognitive process is implausible in absolute terms, then the researcher should not be greatly confident that it is tied to any particular brain region. This point echoes the situation in forward inference that if two theories being compared are both incorrect, then imaging results could only give misleading support for one over the other. The conditional probability also depends on the selectivity of the brain region. For example, if the brain region is so large that it is activated by many cognitive processes, then it will be difficult to infer the engagement of any one process when the region is activated.

Although reverse inference is not used to directly compare theories, it is a part of the scientific process that could be used to develop theories. Moreover, Poldrack's (2006) Bayesian formulation of reverse inference inspires a Bayesian generalization of forward inference, as shown in Eq. 1.

$$P\left(\text{theory}_1|\text{results}\right) = \frac{P\left(\text{results}|\ \text{theory}_1\right) P(\text{theory}_1)}{P\left(\text{results}|\ \text{theory}_1\right) P(\text{theory}_1) + P\left(\text{results}|\ \text{theory}_2\right) \times P(\text{theory}_2) + \ldots + P\left(\text{results}|\ \text{theory}_n\right) P(\text{theory}_n)} \tag{1}$$

Here, the conditional probability that theory 1 is correct after observing a set of neuroimaging results depends on the conditional probability of the results under that theory, as well as the prior likelihood of the theory. This probability must be normalized in terms of the likelihood of other, competing theories. Forward inference is a special case with two theories and the observed results being either the same pattern of brain activity across two experimental tasks or different patterns of brain activity.

PREDICTIONS ABOUT BRAIN ACTIVITY

Next, several examples of neuroimaging studies of reasoning, aiming to address theoretical views, will be reviewed in the light of the method of forward inference.

DEDUCTION VERSUS INDUCTION

At least one of the contrasts made in imaging research on reasoning is a good example of forward inference. Several studies (Goel et al., 1997; Osherson et al., 1998; Parsons and Osherson, 2001; Goel and Dolan, 2004) have compared deductive and inductive reasoning tasks. One class of theories (including mental model theory and Bayesian accounts) has suggested that deduction and induction are performed by a common set of processes. Another class of theories (dual-process theories) has suggested that there are two types of underlying mechanisms of reasoning, heuristic and analytic processing, which would contribute differentially to deduction and induction. To the extent that different patterns of brain activity are observed for deduction versus induction tasks, holding everything else equal between experimental conditions, by forward inference, dual-process accounts will be supported over single-process alternatives. (Note that three of these studies, all but Goel and Dolan, 2004, used exactly the same materials for the two conditions, but simply asked a deduction question or an induction question.) Indeed, these four studies all found somewhat different patterns of brain activation for deduction versus induction. Three of these studies (Goel et al., 1997; Osherson et al., 1998; Goel and Dolan, 2004) found increased activation for induction, relative to deduction, in left frontal cortex, although in somewhat different regions at a finer level. Although it would be valuable to have an understanding of why the regions differ between studies, which is not crucial for the method of forward inference.

Overall, these results do make a good case for dual-process theories over single-process theories, notwithstanding the limitations of forward inference described above. To accommodate, these results would require single-process theories to assume somewhat different processes for deduction versus induction, e.g., to become more like dual-process theories. In a related line of work Houdé et al. (2000, 2001) compared brain activity before and after a training session aimed at improving logical reasoning, rather than comparing reasoning under two sets of instructions. In terms of the method of forward inference, the qualitatively different patterns of activity pre- versus post-training would be a challenge for single-process accounts, without assuming that deduction before and after training engages different processes.

MEANINGFUL VERSUS NON-MEANINGFUL MATERIAL

Another contrast is a slightly less clear example of forward inference. Several studies have varied the content of arguments while otherwise keeping the task the same, e.g., abstract versus concrete materials (Goel et al., 2000; Goel and Dolan, 2001), materials that agree, disagree, or are neutral with respect to prior knowledge (Goel and Dolan, 2003), and visual versus spatial relations such as "fatter than" versus "is a descendant of" (Knauff et al., 2003). To apply forward inference, what is needed is one theory that predicts the same cognitive processes between conditions, and another theory that predicts different cognitive processes between conditions. With regard to the abstract/concrete and prior knowledge studies, the results were greater bilateral parietal activation for abstract or neutral content, and in two of the studies, greater left temporal activation for concrete or knowledge-related materials. With regard to the study on visual versus spatial relations, the finding was that visual problems led to enhanced activity in visual association cortex. Although these differences in brain activity would be consistent with dual-process accounts assuming that somewhat different mechanisms are employed depending on content, the problem is that even single-process accounts would need to make some assumptions to explain how content affects reasoning. So it is unclear that single-process accounts are ruled out [cf., Keren (2013)]. From the perspective of forward inference, the problem is the lack of well-defined theories making sharply different predictions.

LEFT VERSUS RIGHT HEMISPHERE

A frequent prediction addressed in brain imaging research on reasoning is whether the left or right hemisphere is activated. It is tempting to link mental logic theory, having a propositional nature, with left hemisphere activation and mental model theory, having a visuospatial nature, with right hemisphere activation. Therefore, by looking at which hemisphere is predominantly activated during a reasoning task, one might see which theory has greater support. With regard to mental model theory, the origin of this prediction appears to be Johnson-Laird (1994), and it has been tested in many studies (Goel et al., 1997, 1998, 2000; Parsons and Osherson, 2001; Knauff et al., 2002, 2003; Noveck et al., 2004; Monti et al., 2007, 2009). Although reasoning tasks are typically associated with left hemisphere activation, the results have actually been mixed (Goel, 2007), with many studies showing activation in both hemispheres.

Of greater concern is not the result but the soundness of the hemispheric prediction. An inference of the form "if theory X is correct then brain region Y will be activated" is neither forward inference nor reverse inference. Indeed, no proponent of either theory of reasoning would likely abandon their beliefs based on tests of these predictions. Noveck et al. (2004) suggested that no proponent of mental logic theory has even made predictions about brain regions. Moreover, the predictions about brain regions are not unique, e.g., alternative predictions can also be made for mental model theory, such as parietal activation (Knauff et al., 2003) or activation in the anterior prefrontal cortex (Fangmeier et al., 2006). Knauff et al. even suggested that left hemisphere activation may be consistent with mental model theory, because comprehension of arguments will recruit linguistic areas of the brain.

A final problem with the hemispheric prediction is that it sets up a comparison between two theories that are not the only possibilities. In terms of Eq. 1, other theories need to be considered. For example, the studies reviewed here did not consider Bayesian accounts of deduction (Oaksford and Chater, 2007), yet these accounts have amassed a growing set of successes in the domain of reasoning.

CONDITIONAL VERSUS RELATIONAL ARGUMENTS

Other neuroimaging studies (Knauff et al., 2002; Prado et al., 2010) have compared reasoning about two types of deduction problems, conditional (if-then) arguments and relational arguments (e.g., regarding relative spatial position). The Knauff et al. study was largely concerned with hemispheric predictions comparing mental model and mental logic theory. There were some differences in activation when comparing the two argument types; however, these differences were bilateral and not interpreted strongly. Prado et al. were more directly interested in comparing the two argument types, and indeed observed that the left inferior frontal gyrus is activated more for conditional arguments and the right temporo-parieto-occipital region is activated more for spatial arguments. These results were interpreted as evidence against "unitary" accounts of deduction and evidence for "fractionated" accounts of deduction. To the extent that unitary views predict that the same cognitive processes are used for the two tasks, and fractionated views predict that different processes are used, this is a good example of forward inference. Prado et al. took a particularly nuanced approach, pointing out that although mental model and mental logic theory can be treated as unitary accounts, it is possible to imagine "hybrid" versions predicting somewhat different cognitive processes depending on argument type. Hence, the results are useful in ruling out basic versions of single-process accounts of reasoning. However, the problem, in terms of forward inference and Eq. 1, is that multiple theories of the fractionated type, which is multiple theories that predict that different processes will underlie different problems, are still possible. So there is negative evidence against some theories but the distinctive, positive evidence for other theories is less clear.

For further discussion, including a meta-analysis of brain imaging studies across argument types and presentation modalities, see Prado et al. (2011) for an extended argument that deductive reasoning is better described in terms of multiple systems than a single mechanism.

CONCLUSION

Just as researchers spell out all of the methodological details of brain imaging studies, it is valuable when researchers spell out the details of their own reasoning, e.g., list alternative theories, give sources for predictions, examine alternative predictions, and explain the rationale of testing predictions. The method of forward inference is one such rationale, although as discussed, it is not without its own limitations. This review of brain imaging studies of reasoning has shown that some comparisons, namely, deduction versus induction and conditional arguments versus relational arguments, have made profitable use of forward inference. The possible theoretical contributions of other studies reviewed here appears to lie outside of forward inference, likely reflecting limitations of forward inference as well as cases where the studies need a more fully spelled-out rationale for making theoretical comparisons.

Looking to the future, another approach with great promise is to combine neuroimaging with mathematical modeling, to test well-specified psychological theories. Indeed, some methods of combining neuroimaging and modeling can be seen as extensions or generalizations of the method of forward inference, providing alternative methods for distinguishing between psychological processing accounts using neuroimaging data. For example, rather than comparing a single-process account to a dual-process account, McClure et al. (2007) implemented a mixture model comprising two processes, with the aim of linking model parameters to localized brain activity. Staresina et al. (2013) used the method of state-trace analysis to look for non-monotonic patterns of brain activity across experimental conditions that would rule out single-process accounts. Mack et al. (2013) compared patterns of brain activation to latent model representations for competing psychological models, assessing the match between brain activity and model predictions across multiple experimental manipulations. Finally, Rotello and Heit (2014) reinterpreted brain imaging studies of conflicts between prior beliefs and deductive reasoning, seeming to show multiple reasoning processes, using an algebraic analysis based on signal detection theory.

ACKNOWLEDGMENTS

I am grateful to Caren Rotello, Haruka Swendsen, and Changquan Long for discussions of this line of research. This material is based upon work while the author was serving at the National Science Foundation (USA). Any opinion, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Anderson, J. R. (1978). Arguments concerning representations for mental imagery. *Psychol. Rev.* 85, 249–277. doi:10.1037/0033-295X.85.4.249

Braine, M. D., and O'Brien, D. P. (eds) (1998). *Mental Logic*. Mahwah, NJ: Erlbaum.

Chiong, W., Wilson, S. M., D'Esposito, M., Kayser, A. S., Grossman, S. N., Poorzand, P., et al. (2013). The salience network causally influences default mode network activity during moral reasoning. *Brain* 136, 1929–1941. doi:10.1093/brain/awt066

Coltheart, M. (2006). What has functional neuroimaging told us about the mind (so far. *Cortex* 42, 323–331. doi:10.1016/S0010-9452(08)70358-7

Del Pinal, G., and Nathan, M. J. (2013). There and up again: on the uses and misuses of neuroimaging in psychology. *Cogn. Neuropsychol.* 30, 233–252. doi:10.1080/02643294.2013.846254

Evans, J. S. B., and Stanovich, K. E. (2013). Dual-process theories of higher cognition advancing the debate. *Perspect. Psychol. Sci.* 8, 223–241. doi:10.1177/1745691612460685

Fangmeier, T., Knauff, M., Ruff, C. C., and Sloutsky, V. (2006). fMRI evidence for a three-stage model of deductive reasoning. *J. Cogn. Neurosci.* 18, 320–334. doi:10.1162/jocn.2006.18.3.320

Goel, V. (2007). Anatomy of deductive reasoning. *Trends Cogn. Sci.* 11, 435–441. doi:10.1016/j.tics.2007.09.003

Goel, V., Buchel, C., Frith, C., and Dolan, R. J. (2000). Dissociation of mechanisms underlying syllogistic reasoning. *Neuroimage* 12, 504–514. doi:10.1006/nimg.2000.0636

Goel, V., and Dolan, R. J. (2001). Functional neuroanatomy of three-term relational reasoning. *Neuropsychologia* 39, 901–909. doi:10.1016/S0028-3932(01)00024-0

Goel, V., and Dolan, R. J. (2003). Explaining modulation of reasoning by belief. *Cognition* 87, B11–B22. doi:10.1016/S0010-0277(02)00185-3

Goel, V., and Dolan, R. J. (2004). Differential involvement of left prefrontal cortex in inductive and deductive reasoning. *Cognition* 93, B109–B121. doi:10.1016/j.cognition.2004.03.001

Goel, V., Gold, B., Kapur, S., and Houle, S. (1997). The seats of reason? An imaging study of deductive and inductive reasoning. *Neuroreport* 8, 1305–1310. doi:10.1097/00001756-199703240-00049

Goel, V., Gold, B., Kapur, S., and Houle, S. (1998). Neuroanatomical correlates of human reasoning. *J. Cogn. Neurosci.* 10, 293–302. doi:10.1162/089892998562744

Harley, T. A. (2004). Does cognitive neuropsychology have a future? *Cogn. Neuropsychol.* 21, 3–16. doi:10.1080/02643290342000131

Hayes, B. K., Heit, E., and Swendsen, H. (2010). Inductive reasoning. *Wiley Interdiscip. Rev. Cogn. Sci.* 1, 278–292. doi:10.1002/wcs.44

Heit, E. (1998). "A Bayesian analysis of some forms of inductive reasoning," in *Rational Models of Cognition*, eds M. Oaksford and N. Chater (Oxford: Oxford University Press), 248–274.

Heit, E. (2007). "What is induction and why study it?," in *Inductive Reasoning*, eds A. Feeney and E. Heit (Cambridge: Cambridge University Press), 1–24.

Heit, E., and Rotello, C. M. (2010). Relations between inductive reasoning and deductive reasoning. *J. Exp. Psychol. Learn. Mem. Cogn.* 36, 805–812. doi:10.1037/a0018784

Heit, E., Rotello, C. M., and Hayes, B. K. (2012). "Relations between memory and reasoning," in *Psychology of Learning and Motivation*, Vol. 57, ed. B. H. Ross (San Diego, CA: Academic Press), 57–101.

Henson, R. (2005). What can functional neuroimaging tell the experimental psychologist? *Q. J. Exp. Psychol. A* 58, 193–233. doi:10.1080/02724980443000502

Henson, R. (2006). Forward inference using functional neuroimaging: dissociations versus associations. *Trends Cogn. Sci.* 10, 64–69. doi:10.1016/j.tics.2005.12.005

Houdé, O., Zago, L., Crivello, F., Moutier, S., Pineau, A., Mazoyer, B., et al. (2001). Access to deductive logic depends on a right ventromedial prefrontal area devoted to emotion and feeling: evidence from a training paradigm. *Neuroimage* 14, 1486–1492. doi:10.1006/nimg.2001.0930

Houdé, O., Zago, L., Mellet, E., Moutier, S., Pineau, A., Mazoyer, B., et al. (2000). Shifting from the perceptual brain to the logical brain: the neural impact of cognitive inhibition training. *J. Cogn. Neurosci.* 12, 721–728. doi:10.1162/089892900562525

Johnson, R. (1993). On the neural generators of the P300 component of the event-related potential. *Psychophysiology* 30, 90–97. doi:10.1111/j.1469-8986.1993.tb03208.x

Johnson-Laird, P. N. (1994). Mental models and probabilistic thinking. *Cognition* 50, 189–209. doi:10.1016/0010-0277(94)90028-0

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Keren, G. (2013). A tale of two systems: a scientific advance or a theoretical stone soup? Commentary on Evans & Stanovich (2013). *Perspect. Psychol. Sci.* 8, 257–262. doi:10.1177/1745691613483474

Knauff, M., Fangmeier, T., Ruff, C. C., and Johnson-Laird, P. N. (2003). Reasoning, models, and images: behavioral measures and cortical activity. *J. Cogn. Neurosci.* 15, 559–573. doi:10.1162/089892903321662949

Knauff, M., Mulack, T., Kassubek, J., Salih, H. R., and Greenlee, M. W. (2002). Spatial imagery in deductive reasoning: a functional MRI study. *Cogn. Brain Res.* 13, 203–212. doi:10.1016/S0926-6410(01)00116-1

Mack, M. L., Preston, A. R., and Love, B. C. (2013). Decoding the brain's algorithm for categorization from its neural implementation. *Curr. Biol.* 23, 2023–2027. doi:10.1016/j.cub.2013.08.035

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: WH Freeman and Company.

McClure, S. M., Ericson, K. M., Laibson, D. I., Loewenstein, G., and Cohen, J. D. (2007). Time discounting for primary rewards. *J. Neurosci.* 27, 5796–5804. doi:10.1523/JNEUROSCI.4246-06.2007

Monti, M. M., and Osherson, D. N. (2012). Logic, language and the brain. *Brain Res.* 1428, 33–42. doi:10.1016/j.brainres.2011.05.061

Monti, M. M., Osherson, D. N., Martinez, M. J., and Parsons, L. M. (2007). Functional neuroanatomy of deductive inference: a language-independent distributed network. *Neuroimage* 37, 1005–1016. doi:10.1016/j.neuroimage.2007.04.069

Monti, M. M., Parsons, L. M., and Osherson, D. N. (2009). The boundaries of language and thought: neural basis of inference making. *Proc. Natl. Acad. Sci. U.S.A.* 106, 12554–12559. doi:10.1073/pnas.0902422106

Nosofsky, R. M., Little, D. R., and James, T. W. (2012). Activation in the neural network responsible for categorization and recognition reflects parameter changes. *Proc. Natl. Acad. Sci. U.S.A.* 109, 333–338. doi:10.1073/pnas.1111304109

Noveck, I. A., Goel, V., and Smith, K. W. (2004). The neural basis of conditional reasoning with arbitrary content. *Cortex* 40, 613–622. doi:10.1016/S0010-9452(08)70157-6

Oaksford, M., and Chater, N. (2007). *Bayesian Rationality: The Probabilistic Approach to Human Reasoning*. Oxford: Oxford University Press.

Osherson, D., Perani, D., Cappa, S., Schnur, T., Grassi, F., and Fazio, F. (1998). Distinct brain loci in deductive versus probabilistic reasoning. *Neuropsychologia* 36, 369–376. doi:10.1016/S0028-3932(97)00099-7

Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., and Shafir, E. (1990). Category-based induction. *Psychol. Rev.* 97, 185–200. doi:10.1037/0033-295X.97.2.185

Parsons, L. M., and Osherson, D. (2001). New evidence for distinct right and left brain systems for deductive versus probabilistic reasoning. *Cereb. Cortex* 11, 954–965. doi:10.1093/cercor/11.10.954

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends Cogn. Sci.* 10, 59–63. doi:10.1016/j.tics.2005.12.004

Prado, J., Chadha, A., and Booth, J. R. (2011). The brain network for deductive reasoning: a quantitative meta-analysis of 28 neuroimaging studies. *J. Cogn. Neurosci.* 23, 3483–3497. doi:10.1162/jocn_a_00063

Prado, J., Van Der Henst, D., Van, J. B., and Noveck, I. A. (2010). Recomposing a fragmented literature: how conditional and relational arguments engage different neural systems for deductive reasoning. *Neuroimage* 51, 1213–1221. doi:10.1016/j.neuroimage.2010.03.026

Rips, L. J. (1994). *The Psychology of Proof: Deductive Reasoning in Human Thinking*. Cambridge, MA: MIT Press.

Rotello, C., and Heit, E. (2014). The neural correlates of belief bias: activation in inferior frontal cortex reflects response rate differences. *Front. Hum. Neurosci.* 8:1–4. doi:10.3389/fnhum.2014.00862

Rotello, C. M., and Heit, E. (2009). Modeling the effects of argument length and validity on inductive and deductive reasoning. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 1317–1330. doi:10.1037/a0016648

Sloman, S. A. (1993). Feature-based induction. *Cogn. Psychol.* 25, 231–280. doi:10.1006/cogp.1993.1006

Staresina, B. P., Fell, J., Dunn, J. C., Axmacher, N., and Henson, R. N. (2013). Using state-trace analysis to dissociate the functions of the human hippocampus and perirhinal cortex in recognition memory. *Proc. Natl. Acad. Sci. U.S.A.* 110, 3119–3124. doi:10.1073/pnas.1215710110

Tenenbaum, J. B., and Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behav. Brain Sci.* 24, 629–641. doi:10.1017/S0140525X01000061

Uttal, W. R. (2011). *Mind and Brain: A Critical Appraisal of Cognitive Neuroscience*. Cambridge, MA: MIT Press.

Van Orden, G. C., and Paap, K. R. (1997). Functional neuroimages fail to discover pieces of mind in the parts of the brain. *Philos. Sci.* 64, S85–S94. doi:10.1086/392589

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: 26 September 2014; accepted: 18 December 2014; published online: 09 January 2015.

Citation: Heit E (2015) Brain imaging, forward inference, and theories of reasoning. Front. Hum. Neurosci. ***8****:1056. doi: 10.3389/fnhum.2014.01056*

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

frontiers in
HUMAN NEUROSCIENCE

OPINION ARTICLE
published: 21 October 2014
doi: 10.3389/fnhum.2014.00862

The neural correlates of belief bias: activation in inferior frontal cortex reflects response rate differences

***Caren M. Rotello*[1]* and Evan Heit*[2]**

[1] Department of Psychological and Brain Sciences, University of Massachusetts, Amherst, MA, USA
[2] School of Social Sciences, Humanities and Arts, University of California, Merced, CA, USA
*Correspondence: caren@psych.umass.edu

Edited by:
Vinod Goel, York University, Canada

Reviewed by:
Matt Roser, Plymouth University, UK
Oshin Vartanian, Toronto Research Centre, Canada
Henrik Singmann, Albert-Ludwigs-Universität Freiburg, Germany

Keywords: belief bias, reasoning, neuroscience, accuracy, congruency

The belief bias effect in reasoning (Evans et al., 1983) is the tendency for logical problems with believable conclusions (e.g., some addictive things are not cigarettes) to elicit more positive responses than those with unbelievable conclusions (some cigarettes are not addictive things). The effect of believability interacts with conclusion validity (see the lower rows of **Table 1** for example data), leading many researchers to conclude that reasoning accuracy is greater for problems with unbelievable conclusions (e.g., Oakhill and Johnson-Laird, 1985; Newstead et al., 1992; Quayle and Ball, 2000). Dube et al. (2010, 2011) [see also Heit and Rotello (2014)] demonstrated that the typical ANOVA analysis of these behavioral data was inappropriate, and showed that a signal detection based interpretation of the data reached a different conclusion, namely that the effect of conclusion believability was to shift subjects' response bias to be more liberal. Trippas et al. (2013) also concluded that conclusion believability consistently affected response bias, but that reasoning accuracy was additionally affected by believability under certain conditions (i.e., higher cognitive ability, complex syllogisms, unlimited decision time).

The belief bias effect has also been studied in the neuroscience literature, although the focus has been slightly different. Whereas in the behavioral literature, researchers have focused on the accuracy with which subjects can discriminate valid from invalid conclusions, in the neuroscience literature, questions have centered on the brain regions responsible for resolving the conflict between the logically correct response to a problem and the believability of its conclusion. That is, neuroscience analyses have divided test trials into those for which validity and believability lead to the same conclusion (congruent trials) and those for which they lead to different conclusions (incongruent trials). A consistent finding is that the percentage of correct responses is higher for congruent than incongruent trials, an effect attributed to the competition between System 1, which drives belief-based responding, and System 2, which drives logic-based decisions (e.g., Goel et al., 2000; Tsujii and Watanabe, 2010; cf. Evans and Curtis-Holmes, 2005). A similarly consistent finding is the selective activation of right prefrontal cortex (rPFC) for incongruent, and not congruent, test trials, suggesting a role for rPFC in conflict detection and/or resolution (fMRI: Goel et al., 2000; Goel and Dolan, 2003; Stollstorff et al., 2012; fNIRS: Tsujii and Watanabe, 2009, 2010; Tsujii et al., 2010b; TMS: Tsujii et al., 2010a). For example, Stollstorff et al. (2012) noted that right lateral PFC "is consistently engaged to resolve conflict in deductive reasoning" (p. 28). In ERP, a late positivity for incongruent trials has been interpreted similarly (Luo et al., 2008, 2013). These data suggest that rPFC activation inhibits System 1 responding, a conclusion that is broadly consistent with the assumed inhibitory function of right inferior frontal cortex (Aron et al., 2014).

We will begin by showing that the partitioning of trials and subsequent analysis are based on faulty logic, such that the intended comparison of accuracy for congruent versus incongruent trials actually reflects differences in the "valid" response rates to believable and unbelievable problems. Using simple algebra, we show that accuracy for congruent and incongruent trials can only be equal when the 'valid' response rate does not vary with believability. Second, we will turn to the interpretation of the corresponding brain data, arguing that it is also flawed because of its dependence on those very same accuracy differences. Finally, we will suggest an alternative interpretation of rPFC activation in the belief bias task.

In belief bias studies, accuracy for the congruent trials, A_c, is measured using percent correct. It is simply the average of the "valid" (hit) response rate in the believable condition (H_B) and the "invalid" (correct rejection) response rate in the unbelievable condition (CR_U):

$$A_C = \frac{1}{2}(H_B + CR_U) \qquad (1)$$

Likewise, accuracy for the incongruent trials, A_I, is simply the average of the hit rate in the unbelievable condition and the correct rejection rate in the believable condition:

$$A_I = \frac{1}{2}(H_U + CR_B) \qquad (2)$$

For example, for the representative data in the lower rows of **Table 1**, $A_C = 0.5(0.86 + 0.68) = 0.77$, and $A_I = 0.5$ $(0.68 + 0.39) = 0.54$, implying that accuracy is higher for the congruent than the incongruent trials. Interestingly, the accuracy advantage seen for congruent trials is observed even though believability did

Table 1 | Data from Dube et al. (2010).

Experiment	Condition	Response rates				
		H = *P*("valid"| Valid)	Miss = *P*("invalid"| Valid)	*F* = *P*("valid"| invalid)	CR = *P*("invalid"| invalid)	Overall "valid" response rate
1	Liberal	0.79	0.21	0.67	0.33	0.730
	Conservative	0.55	0.45	0.31	0.69	0.430
2	Believable	0.86	0.14	0.61	0.39	0.735
	Unbelievable	0.68	0.32	0.32	0.68	0.500

not affect validity discrimination in this experiment (Dube et al., 2010, Exp. 2).

The interpretation of the neuroscience data on belief bias depends crucially on the difference in accuracy for congruent and incongruent trials. To understand these data, we first show that interpretation of the percent correct accuracy measure actually depends on response rate differences. Let us spend a moment examining how the accuracy difference could come about, by starting with the question of when accuracy for the two trial types would be equal. In other words, under what conditions does $A_C = A_I$, or, equivalently, when is Eq. 3 true?

$$\frac{1}{2}(H_B + CR_U) = \frac{1}{2}(H_U + CR_B) \quad (3)$$

Because the correct rejection rate, CR, equals 1 minus the false alarm rate, F, we can rewrite Eq. 3:

$$\frac{1}{2}(H_B + 1 - F_U) = \frac{1}{2}(H_U + 1 - F_B) \quad (4)$$

Some reorganization and simplification yields

$$\frac{1}{2}(H_B + F_B) = \frac{1}{2}(H_U + F_U) \quad (5)$$

Equation 5 is revealing, because the average of the hit and false alarm rates equals the "yes" rate (assuming equal number of target and lure trials). As Macmillan and Creelman (2005) showed, the yes rate is a measure of response bias, not accuracy. Thus, Eq. 5 shows that the congruent and incongruent trials can only yield equal accuracy (measured with percent correct; a related argument applies to d') if the response rates to believable and unbelievable problems are the same. This bias restriction is unlikely to be met, because the belief bias effect itself is a difference in positive response rates with conclusion believability (e.g., Evans et al., 1983; Dube et al., 2010, 2011; Trippas et al., 2013). Believable problems tend to elicit more positive responses both for valid and invalid conclusions; thus, it is easy to see that the congruency analysis will produce $A_C > A_I$. Starting with a version of Eq. 4 that assumes $A_C > A_I$

$$\frac{1}{2}(H_B + 1 - F_U) > \frac{1}{2}(H_U + 1 - F_B) \quad (6)$$

we can simplify and reorganize to see that $A_C > A_I$ whenever

$$H_B - H_U > F_U - F_B \quad (7)$$

Because both the hit and false alarm rate are higher to problems with believable conclusions, the left side of the inequality in Eq. 7 will be positive, and the right side will be negative: A_C will always be greater than A_I if believable conclusions elicit more positive responses than unbelievable conclusions. This observation generalizes to any empirical manipulation that elicits a response rate difference, as long as the more liberal condition is treated as analogous to the believable problems. For example, the upper rows of **Table 1** show data from Dube et al. (2010) (Exp. 1), which was a syllogistic reasoning task on abstract problems that were structurally identical to those in their belief bias experiments. One group of subjects was told that 85% of the problems had a valid conclusion, and another group was told that 15% of the conclusions were valid, though in fact both groups were given identical problem sets in which 50% of conclusions were logically valid. Treating the liberal condition as analogous to the believable problems, and letting the conservative condition play the role of the unbelievable problems, we can compute $A_C = 0.74$ and $A_I = 0.44$, implying that accuracy is higher for the congruent than the incongruent trials despite the absence of any believable (or unbelievable) content.

We turn now to the neuroscience literature, for which we argue that differences in response rates have been misinterpreted as accuracy differences. Neuroscience studies of belief bias have consistently found selective activation of rPFC to incongruent trials (Goel et al., 2000; Goel and Dolan, 2003; Tsujii and Watanabe, 2009, 2010; Tsujii et al., 2010a,b; Stollstorff et al., 2012). Indeed, Tsujii and Watanabe (2009, 2010) and Tsujii et al. (2010b) took this general finding a step further. In each of these three studies, they reported a positive correlation between the magnitude of activation in rIFC and the difference in accuracy levels for incongruent and congruent trials. Tsujii and Watanabe (2009) wrote "subjects with enhanced activation in the right IFC could also perform better in conflicting [incongruent] reasoning trials" (p. 121). As we have seen, however, accuracy differences as a function of congruency simply reflect a different "valid" response rate to problems with believable and unbelievable conclusions. So, a better interpretation of these data is that right IFC activation correlates with the magnitude of that response rate difference. The scatter plots in each of these studies show that the highest degree of selective activation (largest difference for incongruent compared to congruent trials) corresponds to accuracy differences (incongruent minus congruent) that are zero or positive, meaning that those subjects showed an atypical response to the belief bias task: either they showed no response rate difference with believability (and thus had no accuracy difference, see Eq. 5) or they made more positive responses to unbelievable than believable conclusions (and thus had higher accuracy for incongruent trials than congruent, see Eq. 7).

Tsujii et al. (2010a) used TMS to show that disruption to right IFC increased the magnitude of the accuracy difference with congruency: subjects showed large accuracy advantages for congruent trials, which can only occur because of large response rate effects of believability (Eq. 7). Interestingly, disruption to left IFC eliminated the accuracy advantage for congruent trials, meaning that the "valid" response rate to believable and unbelievable conclusions was at least roughly equated (Eq. 5).

Our analysis of the accuracy effect of congruency shows that the analyses in the neuroscience literature on belief bias have not directly addressed why congruency differences occur, the brain regions responsible for conflict detection/resolution, or the relative involvement of reasoning Systems 1 (belief) and 2 (logic). None of those processes have been shown to be involved in the appearance of an accuracy difference with congruency (see Eqs 5 and 7). Instead, the selective activation of prefrontal cortex in response to incongruent problems must be a consequence of the response rate difference for believable and unbelievable problems.

The failure to consider response rate differences across conditions has also lead to the misinterpretation of behavioral data in a variety of domains (e.g., Verde and Rotello, 2003; Rotello et al., 2005; Dougal and Rotello, 2007; Evans et al., 2009; Mickes et al., 2012) and of other neuroscience data. For example, fMRI evidence from perceptual categorization and recognition tasks had been interpreted as showing distinct cortical systems for these tasks (e.g., Reber et al., 1998). However, Nosofsky et al. (2012) noted that the "yes" response rate also differs by task: categorization naturally suggests a more liberal response criterion than recognition. When activation patterns were compared for categorization tasks and a recognition task in which subjects were instructed to use a liberal recognition criterion, no differences in brain activation were found; the distinct patterns were attributable to the response bias difference.

Some recent neuroscience studies have explicitly manipulated the decision criterion across trials. In simple perceptual tasks such as line length discrimination, this can be accomplished by showing participants the length of the line to use as the boundary between "short" and "long" responses. Using this strategy, White et al. (2012) found left inferior temporal cortex, which is responsible for representing objects, was activated in response to the decision criterion itself. They suggested that the criterion value (here, an explicitly provided line length) was stored much like any other stimulus, and so its particular brain location would vary with the task. In the case of syllogistic reasoning, the decision criterion represents a level of evidence for the validity of the conclusion. Where this information would be stored is an interesting question to consider, but it seems that one possible place to starting looking would be in the right inferior frontal cortex. More generally, we see much promise in future neuroscience studies of belief bias that take account of what can be inferred from analysis of behavioral measures.

AUTHOR CONTRIBUTIONS

Caren M. Rotello identified the problem and wrote the first draft. Evan Heit provided critical revisions.

ACKNOWLEDGMENTS

This study is based upon work while Evan Heit was serving at the National Science Foundation. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Aron, A. R., Robbins, T. W., and Poldrack, R. A. (2014). Inhibition and the right inferior frontal cortex: one decade on. *Trends Cogn. Sci.* 18, 177–185. doi:10.1016/j.tics.2013.12.003

Dougal, S., and Rotello, C. M. (2007). "Remembering" emotional words is based on response bias, not recollection. *Psychon. Bull. Rev.* 14, 423–429. doi:10.3758/BF03194083

Dube, C., Rotello, C. M., and Heit, E. (2010). Assessing the belief bias effect with ROCs: it's a response bias effect. *Psychol. Rev.* 117, 831–863. doi:10.1037/a0019634

Dube, C., Rotello, C. M., and Heit, E. (2011). The belief bias effect is aptly named: a reply to Klauer and Kellen (2011). *Psychol. Rev.* 118, 155–163. doi:10.1037/a0021774

Evans, J. S. B. T., Barston, J. L., and Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Mem. Cognit.* 11, 295–306. doi:10.3758/BF03196976

Evans, J. S. B. T., and Curtis-Holmes, J. (2005). Rapid responding increases belief bias: evidence for the dual-process theory of reasoning. *Think. Reason.* 11, 382–389. doi:10.1080/13546780542000005

Evans, K., Rotello, C. M., Li, X., and Rayner, K. (2009). Scene perception and memory revealed by eye movements and ROC analyses: does a cultural difference truly exist? *Q. J. Exp. Psychol.* 62, 276–285. doi:10.1080/17470210802373720

Goel, V., Buchel, C., Frith, C., and Dolan, R. (2000). Dissociation of mechanisms underlying syllogistic reasoning. *Neuroimage* 12, 504–514. doi:10.1006/nimg.2000.0636

Goel, V., and Dolan, R. (2003). Explaining modulation of reasoning by belief. *Cognition* 87, 11–22. doi:10.1016/S0010-0277(02)00185-3

Heit, E., and Rotello, C. M. (2014). Traditional difference-score analyses of reasoning are flawed. *Cognition* 131, 75–91. doi:10.1016/j.cognition.2013.12.003

Luo, J., Liu, X., Stupple, E. J. N., Zhang, E., Xiao, X., Jia, L., et al. (2013). Cognitive control in belief-laden reasoning during conclusion processing: an ERP study. *Int. J. Psychol.* 48, 224–231. doi:10.1080/00207594.2012.677539

Luo, J., Yuan, J., Qiu, J., Zhang, Q., Zhong, J., and Huai, Z. (2008). Neural correlates of the belief-bias effect in syllogistic reasoning: an event-related potential study. *Neuroreport* 19, 1073–1078. doi:10.1097/WNR.0b013e3283052fe1

Macmillan, N. A., and Creelman, C. D. (2005). *Detection Theory: A User's Guide*, 2nd Edn. Mahwah, NJ: Lawrence Erlbaum Associates.

Mickes, L., Flowe, H. D., and Wixted, J. T. (2012). Receiver operating characteristic analysis of eyewitness memory: comparing the diagnostic accuracy of simultaneous versus sequential lineups. *J. Exp. Psychol.* 18, 361–376. doi:10.1037/a0030609

Newstead, S. E., Pollard, P., Evans, J. S., and Allen, J. (1992). The source of belief bias effects in syllogistic reasoning. *Cognition* 45, 257–284. doi:10.1016/0010-0277(92)90019-E

Nosofsky, R. M., Little, D. R., and James, T. W. (2012). Activation in the neural network responsible for categorization and recognition reflects parameter changes. *Proc. Natl. Acad. Sci. U.S.A.* 109, 333–338. doi:10.1073/pnas.1111304109

Oakhill, J. V., and Johnson-Laird, P. (1985). The effects of belief on the spontaneous production of syllogistic conclusions. *Q. J. Exp. Psychol. A* 37, 553–569. doi:10.1080/14640748508400919

Quayle, J., and Ball, L. (2000). Working memory, metacognitive uncertainty, and belief bias in syllogistic reasoning. *Q. J. Exp. Psychol. A* 53, 1202–1223. doi:10.1080/02724980050156362

Reber, P. J., Stark, C. E. L., and Squire, L. R. (1998). Contrasting cortical activity associated with category memory and recognition memory. *Learn. Mem.* 5, 420–428.

Rotello, C. M., Macmillan, N. A., Reeder, J. A., and Wong, M. (2005). The remember response: subject to bias, graded, and not a process-pure indicator of recollection. *Psychon. Bull. Rev.* 12, 865–873. doi:10.3758/BF03196778

Stollstorff, M., Vartanian, O., and Goel, V. (2012). Levels of conflict in reasoning modulate right lateral prefrontal cortex. *Brain Res.* 1428, 24–32. doi:10.1016/j.brainres.2011.05.045

Trippas, D., Handley, S. J., and Verde, M. F. (2013). The SDT model of belief bias: complexity, time, and cognitive ability mediate the effects of believability.

J. Exp. Psychol. Learn. Mem. Cogn. 39, 1393–1402. doi:10.1037/a0032398

Tsujii, T., Masuda, S., Akiyama, T., and Watanabe, S. (2010a). The role of inferior frontal cortex in belief-bias reasoning: an rTMS study. *Neuropsychologia* 48, 2005–2008. doi:10.1016/j.neuropsychologia.2010.03.021

Tsujii, T., Okada, M., and Watanabe, S. (2010b). Effects of aging on hemispheric asymmetry in inferior frontal cortex activity during belief-bias syllogistic reasoning: a near-infrared spectroscopy study. *Behav. Brain Res.* 210, 178–183. doi:10.1016/j.bbr.2010.02.027

Tsujii, T., and Watanabe, S. (2009). Neural correlates of dual-task effect on belief-bias syllogistic reasoning: a near-infrared spectroscopy study. *Brain Res.* 1287, 118–125. doi:10.1016/j.brainres.2009.06.080

Tsujii, T., and Watanabe, S. (2010). Neural correlates of belief-bias reasoning under time pressure: a near-infrared spectroscopy study. *Neuroimage* 50, 1320–1326. doi:10.1016/j.neuroimage.2010.01.026

Verde, M. F., and Rotello, C. M. (2003). Does familiarity change in the revelation effect? *J. Exp. Psychol. Learn. Mem. Cogn.* 29, 739–746. doi:10.1037/0278-7393.29.5.739

White, C. N., Mumford, J. A., and Poldrack, R. A. (2012). Perceptual criteria in the human brain. *J. Neurosci.* 32, 16716–16724. doi:10.1523/JNEUROSCI.1744-12.2012

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: 23 July 2014; accepted: 07 October 2014; published online: 21 October 2014.

*Citation: Rotello CM and Heit E (2014) The neural correlates of belief bias: activation in inferior frontal cortex reflects response rate differences. Front. Hum. Neurosci. **8**:862. doi: 10.3389/fnhum.2014.00862*

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

frontiers in
HUMAN NEUROSCIENCE

MINI REVIEW ARTICLE
published: 22 December 2014
doi: 10.3389/fnhum.2014.01014

Neural correlates of causal power judgments

Denise Dellarosa Cummins *

Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, IL, USA

Edited by:
Vinod Goel, York University, Canada

Reviewed by:
Mike Oaksford, Birkbeck College, University of London, UK
Jonathan Fugelsang, University of Waterloo, Canada

Correspondence:
Denise Dellarosa Cummins, Department of Psychology, University of Illinois at Urbana-Champaign, IL, USA
e-mail: denise.cummins87@gmail.com

Causal inference is a fundamental component of cognition and perception. Probabilistic theories of causal judgment (most notably causal Bayes networks) derive causal judgments using metrics that integrate contingency information. But human estimates typically diverge from these normative predictions. This is because human causal power judgments are typically strongly influenced by beliefs concerning underlying causal mechanisms, and because of the way knowledge is retrieved from human memory during the judgment process. Neuroimaging studies indicate that the brain distinguishes causal events from mere covariation, and also distinguishes between perceived and inferred causality. Areas involved in error prediction are also activated, implying automatic activation of possible exception cases during causal decision-making.

Keywords: causal power, causal reasoning, causal judgment, causality, neural correlates of causality

Causal inference is a fundamental component of cognition and perception, binding together conceptual categories, imposing structures on perceived events, and guiding decision-making. A type of causal inference that is of particular interest to decision scientists is *causal power judgment*. Causal power refers to the ability of a particular cause alone (when it is present) to elicit an effect, relative to other causes (Cheng, 1997). For example, selective serotonin-reuptake inhibitors (SSRI) may be considered more effective in alleviating depression than a placebo if greater depression alleviation is observed when an SSRI is ingested than when a placebo is ingested.

In probabilistic theories of causal judgment, causal power is assessed through metrics that integrate contingency information. One such normative metric is defined as

$$\Delta \mathrm{P} = (\mathrm{E}|\mathrm{C}) - \mathrm{P}(\mathrm{E}| \sim \mathrm{C})$$

that is, the probability of the effect occurring in the presence of the cause minus the probability of the effect occurring in the absence of the cause. (This metric is referred to as ΔP by Cheng (1997) and as *PNS* by Pearl (2000)). An extension of ΔP that normalizes the metric by means of the base rate of the effect measures the power of the candidate cause to generate or prevent the effect *relative to other possible causes*. Cheng (1997) defined this metric for causes that *generate* an effect as

$$\mathrm{P}_c = \Delta \mathrm{P}/1 - \mathrm{P}(\mathrm{E}| \sim C).$$

This is equivalent to the metric defined by Pearl (2000) as *PS*. For causes that *prevent* the effect, Cheng (1997) defined causal power as

$$\mathrm{P}_c = -\Delta \mathrm{P}/\mathrm{P}(\mathrm{E}| \sim C).$$

The difficulty with the probabilistic approach is that human causal power judgments frequently depart from the normative values predicted by these metrics. This is because human causal power judgments are typically strongly influenced by beliefs concerning underlying causal mechanisms, and because of the way knowledge is retrieved from memory during the judgment process.

CAUSAL MECHANISMS

Causality is distinct from mere contingency or covariation. In causality, one event has the power to bring about another event. In covariation and contingency, two events are simply statistically dependent on one another. People cognize causal events differently than they do simple contingency or covariation, and this is apparent in neuro-imaging results: When viewing launching displays, significantly higher levels of relative activation is observed in the right middle frontal gyrus and the right inferior parietal lobule for causal relative to non-causal events (Fugelsang et al., 2005). Another study contrasted displays of normal causality with magic tricks that appear to violate causality and those that are surprising but do not violate causality (Parris et al., 2009). The results indicated that brain areas responsible for detecting expectancy violations in general (i.e., anterior cingulate cortex and left ventral prefrontal cortex) are not responsible for detecting causality violations. This function appears to be specific to the dorsolateral prefrontal cortex. In another study, identical pairs of words were judged for causal or associative relations in different blocks of trials. Causal judgments, beyond associative judgments, generated distinct activation in left dorsolateral prefrontal cortex and right precuneus, again substantiating the particular involvement of these areas in assessments of causality (Satpute et al., 2005).

Other research indicates that perceptual causality can be neurally distinguished from inferential causality. Inferential causality activates the medial frontal cortex (Fonlupt, 2003). Research involving callosotomy (split-brain) patients also indicates particular left hemispheric involvement (Roser et al., 2005). In contrast, perception of causality can be influenced by the application of transcranial direct stimulation to the right parietal lobe, suggesting that the right parietal lobe is involved in the processing of spatial attributes of causality (Straube and Chatterjee, 2010; Straube et al., 2011).

In short, neuroimaging studies show that the brain distinguishes causal events from non-causal events, and this distinction cannot simply be attributed to the surprising nature of non-causal event displays. It also distinguishes between perceived and inferred causality.

The importance of causal mechanism assessment looms particularly large in causal decision-making. People typically discount even strong covariation/contingency information if no plausible causal mechanism appears responsible for the covariation or contingency (Ahn et al., 1995). In a classic study by Fugelsang and Dunbar (2005), people read either plausible or implausible causal hypotheses and were shown covariation data that were either consistent or inconsistent with these hypotheses. A consistent case was one in which a plausible hypothesis was accompanied by strong covariation (high ΔP) or an implausible hypothesis was accompanied by weak covariation data (low ΔP). An inconsistent scenario was on in which a plausible hypothesis was accompanied by weak covariation data (low ΔP) or an implausible hypothesis was accompanied by strong covariation (high ΔP). The task was to estimate the effectiveness of the purported cause in bringing about the effect. The results showed quite clearly the impact of causal plausibility on behavioral judgments and neural processing. Areas associated with thinking (executive processing and working memory) were more active when people encountered data while evaluating plausible causal scenarios. Areas associated with learning and memory (caudate, parahippocampal gyrus) were activated when data and theory were consistent (plausible + strong data OR implausible + weak data). But when data and theory were *in*consistent (implausible + strong data OR plausible + weak data), attentional and executive processing areas were active (anterior cingulate cortex, prefrontal cortex, precuneus) Attentional and executive processing areas (anterior cingulate gyrus, prefrontal cortex, precuneus) were particularly active when plausible theories encountered disconfirming (weak) covariation. These results were interpreted to mean that people focus on theories that are consistent with their beliefs (plausible causal scenarios). They also attend to disconfirming data, but they do not necessarily revise beliefs in light of disconfirming data. This phenomenon is sometimes referred to as truth maintenance (Doyle, 1979) or belief revision conservatism (Kelly et al., 1997; Corner et al., 2010). Both strategies seek to maintain coherence in one's knowledge base by minimizing changes to current belief in light of new information.

KNOWLEDGE RETRIEVAL

Different types of knowledge are activated when reasoning from cause to effect than when reasoning from effect to cause. When reasoning from cause to effect, disablers are spontaneously activated; when reasoning from effect to cause, alternative causes are spontaneously activated. (Preventive causes in this literature are referred to as disablers.) Consider, for example, arguments of the form *"If Marilyn takes SSRI medication, then her depression will lift/Marilyn is taking SSRI medication/Therefore, Marilyn's depression will lift"*. People's willingness to accept such arguments is inversely proportional to the number of disablers activated in memory (factors that could prevent Marilyn's depression from lifting even though she's taking SSRI medication.) This effect has been observed in adults (e.g., Cummins et al., 1991; Cummins, 1995, 1997; De Neys et al., 2002, 2003; Vershueren et al., 2004) as well as children (Markovits et al., 1998; Janveau-Brennan and Markovits, 1999).

Recently, two models have been proposed to capture the impact of disablers on causal power judgments. In the first model, proposed by Cummins (2010), causal power judgments are captured by the following equation:

$$W_c = B(\alpha/(\alpha + \text{disablers}))$$

W_c represents the decision-maker's estimated probability that the cause will in fact bring about the effect. B is a parameter that reflects the believability of the causal mechanism underlying the purported causal relationship. The inclusion of this parameter is motivated by ample research showing that people ignore or discount covariation information if no they can think of no plausible causal mechanism whereby the purported cause can bring about the effect (e.g., Ahn et al., 1995). In the model, if a decision-maker does not believe the two events are causally related, B = 0 and disablers are irrelevant and hence not activated in memory. Only when they believe a causal mechanism exists that empowers one event to evoke another (B = 1) do disablers become relevant.

The term $\alpha/(\alpha+\text{disablers})$ is a memory activation function—a positively accelerated curve—in which the first few disablers retrieved from memory have greater impact on judgment than those retrieved later. Activation spreads throughout the network of associated disablers, and likelihood estimates drop off significantly the farther it spreads. This is because stronger disablers are presumed to be activated earlier than weaker ones, and therefore have greater impact on judgment outcomes. In other words, the psychological difference between 0 and (e.g.,) 3 items is greater than the psychological difference between (e.g.,) 4 and 7. α is a free parameter; it simply expresses the steepness of the curve, and its value is determined empirically. **Figure 1** depicts causal power likelihood estimates for different disabler and α values when B = 1.

The model captures the likelihood of an effect occurring when a cause is present and disablers are absent, and its crucial prediction is that the number of disablers and the order of disabler retrieval both matter.

The inclusion of α as a parameter is motivated by research on reasoning with causal conditional arguments. De Neys et al. (2003) reported that while "thinking aloud", reasoners did not halt the retrieval process upon retrieving a single counterexample. Instead, they continued to retrieve disablers until a final judgment was made, and willingness to accept causal conclusions declined as more disablers were activated in memory. Their results suggested a non-linear retrieval function, however, in which a

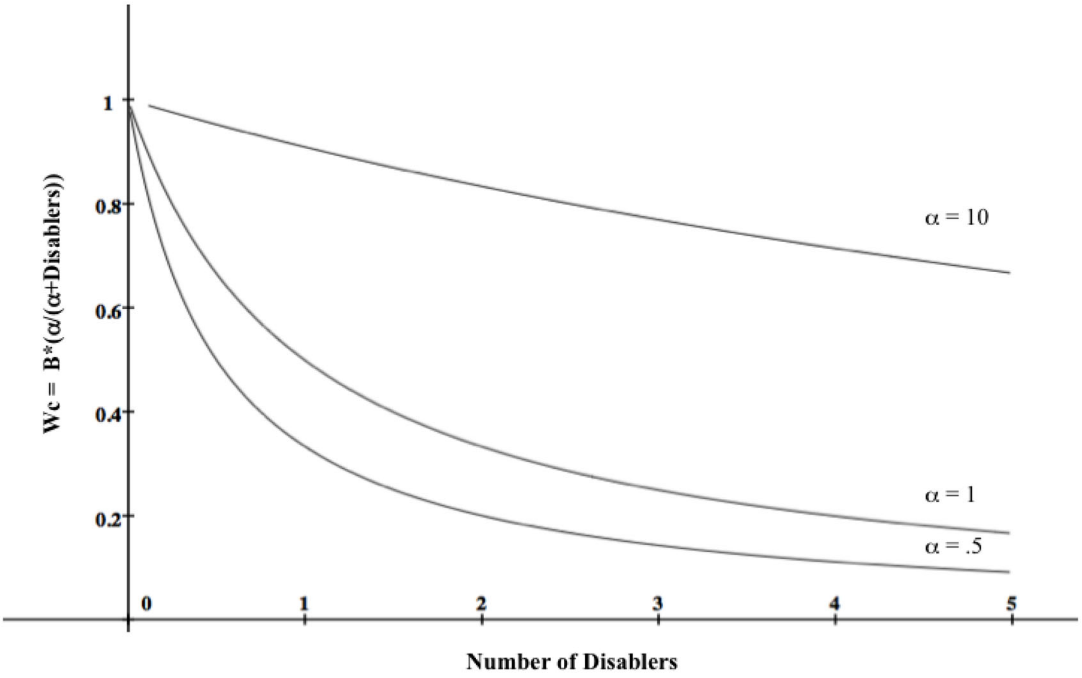


FIGURE 1 | A model of causal power values (W_c) as a function of belief that a causal mechanism underlies the contingency (B) and number of disablers for different values of α, a free parameter whose value is determined empirically. In the graph, B = 1, meaning that the decision-maker believes the contingency reflects a causal relationship. The function shows that the first few disablers retrieved have greater impact on causal power estimates than ones retrieved later.

threshold occurred at about 3 retrieved items, after which argument acceptance ratings changed very little.

In the second model, proposed by Fernbach and Erb (2013), causal power judgments are based on an aggregate disabling probability. Each disabler has some prior likelihood of being present (P_d) and, when present, a likelihood of preventing the effect from occurring, which constitutes its strength (W_d). The disabling probability of any given disabler (A_i) is equal to the product of its prior probability and its strength

$$A_i = P_{di}{}^{*}W_{di}$$

The likelihood that the cause will successfully bring about an effect is the aggregate of these individual disabling probabilities:

$$\begin{aligned} A' \;&=\; \sum\nolimits_{i=1}^{n} A_i - \sum\nolimits_{i,j:i<j} A_iA_j + \sum\nolimits_{i,j,k:i<j<k} A_iA_jA_k - \cdots \\ &\quad + (-1)^{n-1} \prod\nolimits_{i=1}^{n} A_i \end{aligned}$$

As an example, if there are two disablers, then the resulting equation is

$$A' = A_1 + A_2 - A_1{}^{*}A_2$$

If there are three, then it becomes

$$A' = A_1 + A_2 + A_3 - A_1{}^{*}A_2 - A_1{}^{*}A_3 + A_1{}^{*}A_2{}^{*}A_3$$

and so on. Causal power, W_c, is the complement of this aggregate disabling probability, which means that it expresses the likelihood that the cause will bring about the effect when there are no disablers to prevent it:

$$W_c = 1 - A'$$

To summarize, according to Cummins (2010) (a) causal power likelihood estimates diminish as the number of disablers retrieved increases; and (b) earlier retrieved disablers have greater impact than later ones. According to Fernbach and Erb (2013), causal power likelihood can be captured by aggregate disabler impact, a value not affected by order of disabler retrieval.

Fernbach and Erb (2013) found that their model constituted a reasonably good fit for causal arguments but not for non-causal ones, despite similarity in their conditional probabilities. These results constitute strong support for the inclusion of believability parameter when modeling disabler impact. Cummins (2014) found that aggregate impact scores did not fully capture final likelihood judgments well, and the disparity was due to the fact that order of disabler retrieval mattered. Stronger disablers are retrieved first, but, contrary to Cummins' model, the ultimate judgment is more strongly influenced by later retrieved items than by earlier ones.

Recent research has successfully identified the neurocorrelates of disabler retrieval during causal reasoning. Of particular interest are two specific event-related potentials: N2 and P3b. N2 is a frontal negative deflection observed between 200 ms and 300 ms after stimulus onset while P3b is a centroparietal positive deflection observed 250–450 ms after stimulus onset. N2 is typically observed when causal expectations are violated while P3b is typically observed when such expectations are satisfied (Verleger, 1988; Folstein and VanPetten, 2008). Causal arguments that admit of many disablers elicit more pronounced N2 and less pronounced P3b responses than do causal arguments that admit of few disablers (Bonnefond et al., 2014). This pattern of response is interpreted to mean that disabler retrieval lowers reasoners' expectations that an effect will in fact be elicited by a particular cause.

In a related fMRI study (Fenker et al., 2010), a task cue prompted people to evaluate either the causal or the non-causal associative relationship between pairs of words. Causally related pairs elicited higher activity than non-causal associates in orbitofrontal cortex, amygdala, striatum, and substantia nigra/ventral tegmental area. Importantly, this network overlaps with the mesolimbic and mesocortical dopaminergic network known to code prediction errors (O'Doherty et al., 2003, 2007). Because the study context did not explicitly require people to make predictions, activity in this network suggests that that prediction error processing might be automatically recruited in assessments of causality.

The take-home message of this work is that human causal inference cannot be adequately modeled without taking into consideration the ways in which knowledge is activated and weighted in the decision process. Current popular models of causal inference (e.g., Fernbach et al., 2011; Fernbach and Erb, 2013) analyze it as a type of Bayesian inference, yet such models do not constitute adequate *descriptive* models of human predictive inference because they abstract away from these crucially important variables. This implies that human predictive inference is not purely Bayesian. As was well-documented by Kahneman (2011), the source of the discrepancy seems to lie in the way knowledge retrieval transacts with probability estimations. Automatic (e.g., Cummins, 1995, 2010) activation of relevant alternatives is a hallmark of human reasoning, and this characteristic must be accommodated in descriptive models of causal inference if human causal judgments are to be adequately predicted.

AUTHOR NOTES

Dr. Cummins is retired from the University of Illinois at Urbana-Champaign. Correspondence regarding this research should be directed to her at denise.cummins87@gmail.com.

REFERENCES

Ahn, W. K., Kalish, C. W., Medin, D. L., and Gelman, S. A. (1995). The role of covariation vs. mechanism information in causal attribution. *Cognition* 54, 299–352. doi: 10.1016/0010-0277(94)00640-7

Bonnefond, M., Kaliuzhna, M., Van der Henst, J.-B., and De Neys, W. (2014). Disabling conditional inferences: an EEG study. *Neuropsychologia* 56, 255–262. doi: 10.1016/j.neuropsychologia.2014.01.022

Cheng, P. W. (1997). From covariation to causation: a causal power theory. *Psychol. Rev.* 104, 367–405. doi: 10.1037//0033-295x.104.2.367

Corner, A., Harris, A. J. L., and Hahn, U. (2010). "Conservatism in belief revision and participant skepticism," in *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, eds S. Ohlsson and R. Catrambone (Austin, TX: Cognitive Science Society), 1625–1630.

Cummins, D. D. (1995). Naive theories and causal deduction. *Mem. Cognit.* 23, 646–658. doi: 10.3758/bf03197265

Cummins, D. D. (1997). Reply to fairley and Manktelow's comment on "Naïve theories and causal deduction". *Mem. Cognit.* 25, 415–416.

Cummins, D. D. (2010). "How memory processes temper causal inferences," in *Cognition and Conditionals*, eds M. Oaksford and N. Chater (Oxford: Oxford University Press), 207–218.

Cummins, D. D. (2014). The impact of disablers on predictive inference. *J. Exp. Psychol. Learn. Mem. Cogn.* 40, 1638–1655. doi: 10.1037/xlm0000024

Cummins, D. D., Lubart, T., Alksnis, O., and Rist, R. (1991). Conditional reasoning and causation. *Mem. Cognit.* 19, 274–282. doi: 10.3758/bf03211151

De Neys, W., Schaeken, W., and d'Ydewalle, G. (2002). Causal conditional reasoning and semantic memory retrieval: a test of the 'semantic memory framework'. *Mem. Cognit.* 30, 908–920. doi: 10.3758/bf03195776

De Neys, W., Schaeken, W., and d'Ydewalle, G. (2003). Inference suppression and semantic memory retrieval: every counterexample counts. *Mem. Cognit.* 31, 581–595. doi: 10.3758/bf03196099

Doyle, J. (1979). A truth maintenance system. *Artif. Intell.* 12, 251–272. doi: 10.1016/0004-3702(79)90008-0

Fenker, D. B., Schoenfeld, M. A., Waldmann, M. R., Schuetze, H., Heinze, H.-J., and Duezel, E. (2010). "Virus and epidemic": causal knowledge activates prediction error circuitry. *J. Cogn. Neurosci.* 22, 2151–2163. doi: 10.1162/jocn.2009.21387

Fernbach, P. M., Darlow, A., and Sloman, S. A. (2011). Asymmetries in predictive and diagnostic reasoning. *J. Exp. Psychol. Gen.* 140, 168–185. doi: 10.1037/a0022100

Fernbach, P. M., and Erb, C. D. (2013). A quantitative theory of conditional reasoning. *J. Exp. Psychol. Learn. Mem. Cogn.* 39, 1327–1343. doi: 10.1037/a0031851

Folstein, J. R., and VanPetten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology* 45, 152–170. doi: 10.1111/j.1469-8986.2007.00602.x

Fonlupt, P. (2003). Perception and judgement of physical causality involve different brain structures. *Brain Res. Cogn. Brain Res.* 17, 248–254. doi: 10.1016/s0926-6410(03)00112-5

Fugelsang, J., and Dunbar, K. (2005). Brain-based mechanisms underlying complex causal thinking. *Neuropsychologia* 43, 1204–1213. doi: 10.1016/j.neuropsychologia.2004.10.012

Fugelsang, J. A., Roser, M. E., Corballis, P. M., Gazzaniga, M. S., and Dunbar, K. N. (2005). Brain mechanisms underlying perceptual causality. *Brain Res. Cogn. Brain Res.* 24, 41–47. doi: 10.1016/j.cogbrainres.2004.12.001

Janveau-Brennan, G., and Markovits, H. (1999). The development of reasoning with causal conditionals. *Dev. Psychol.* 35, 904–911. doi: 10.1037//0012-1649.35.4.904

Kahneman, D. (2011). *Thinking: Fast and Slow.* New York: Penguin Books.

Kelly, K., Schulte, O., and Hendricks, V. (1997). Reliable belief revision. *Log. Sci. Methods* 259, 383–398. doi: 10.1007/978-94-017-0487-8_20

Markovits, H., Fleury, M., Quinn, S., and Venet, M. (1998). The development of conditional reasoning and the structure of semantic memory. *Child Dev.* 69, 742–755. doi: 10.1111/j.1467-8624.1998.tb06240.x

O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., and Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron* 38, 329–337. doi: 10.1016/s0896-6273(03)00169-7

O'Doherty, J. P., Hampton, A., and Kim, H. (2007). Model-based fMRI and its application to reward learning and decision making. *Ann. N Y Acad. Sci.* 1104, 35–53. doi: 10.1196/annals.1390.022

Parris, B. A., Kuhn, G., Mizon, G. A., Benattayallah, A., and Hodgson, T. L. (2009). Imaging the impossible: an fMRI study of impossible causal relationships in magic tricks. *Neuroimage* 45, 1033–1039. doi: 10.1016/j.neuroimage.2008.12.036

Pearl, J. (2000). *Causality.* Cambridge: Cambridge University Press.

Roser, M. E., Fugelsang, J. A., Dunbar, K. N., Corballis, P. M., and Gazzaniga, M. S. (2005). Dissociating processes supporting causal perception and causal inference in the brain. *Neuropsychology* 19, 591–602. doi: 10.1037/0894-4105.19.5.591

Satpute, A. B., Fenker, D. B., Waldmann, M. R., Tabibnia, G., Holyoak, K. J., and Lieberman, M. D.. (2005). An fMRI study of causal judgments. *Eur. J. Neurosci.* 22, 1233–1238. doi: 10.1111/j.1460-9568.2005.04292.x

Straube, B., and Chatterjee, A. (2010). Space and time in perceptual causality. *Front. Hum. Neurosci.* 4:28. doi: 10.3389/fnhum.2010.00028

Straube, B., Wolk, D., and Chatterjee, A. (2011). The role of the right parietal lobe in the perception of causality: a tDCS study. *Exp. Brain Res.* 215, 315–325. doi: 10.1007/s00221-011-2899-1

Verleger, R. (1988). Event-related potentials and cognition: a critique of the context-updating hypothesis and an alternative interpretation of P3. *Behav. Brain Sci*. 11, 343–356.

Vershueren, N., Schaeken, W., De Neys, W., and d'Ydewalle, G. (2004). The difference between generating counterexamples and using them during reasoning. *Q. J. Exp. Psychol. A* 57A, 1285–1308. doi: 10.1080/02724980343000774

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: 08 October 2014; accepted: 30 November 2014; published online: 22 December 2014.

Citation: Cummins DD (2014) Neural correlates of causal power judgments. Front. Hum. Neurosci. ***8****:1014. doi: 10.3389/fnhum.2014.01014*

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

frontiers in
HUMAN NEUROSCIENCE

GENERAL COMMENTARY
published: 06 February 2015
doi: 10.3389/fnhum.2015.00056

The prospects of working memory training for improving deductive reasoning

Erin L. Beatty[1]* and Oshin Vartanian[1,2]

[1] Defence Research and Development Canada, Toronto Research Centre, Toronto, Canada
[2] Department of Psychology, University of Toronto Scarborough, Toronto, Canada
*Correspondence: erin.beatty@drdc-rddc.gc.ca

Edited by:
Jérôme Prado, Centre National de la Recherche Scientifique, France

Reviewed by:
Allyson P. Mackey, Massachusetts Institute of Technology, USA

Keywords: reasoning, working memory training, fluid intelligence, n-back, education intervention

A commentary on

Improving reasoning skills in secondary history education by working memory training
by Ariës, R. J., Groot, W., and van den Brink, H. M. (2014). Br. Educ. Res. J. doi: 10.1002/berj.3142. [Epub ahead of print].

Cognitive (brain) training has been a major focus of study in recent years. In applied settings, the excitement regarding this research programme emanates from its prospects for *far transfer*—defined as observing performance benefits in outcome measures that are contextually, structurally or superficially dissimilar to the trained task (Perkins and Salomon, 1994). By and large, researchers have focused on training working memory (WM). This is not surprising, given the ubiquity of WM requirements for thinking (Baddeley, 2003). Currently, much evidence suggests that adaptive training on WM tasks can increase WM skills. In contrast, consistent evidence regarding far transfer is lacking (see Melby-Lervåg and Hulme, 2013), although there is evidence to suggest that when the training modality is visuospatial, the likelihood of transfer and the long-term stability of its benefits are enhanced (Melby-Lervåg and Hulme, 2013; Stephenson and Halpern, 2013).

Theoretically, there is reason to suspect that interventions that increase WM skills and/or capacity could improve deductive reasoning. This prediction stems from the observation that individual differences in WM capacity predict deductive reasoning performance on conflict problems where the believability of conclusions conflicts with logical validity (e.g., Newstead et al., 2004). Conflict problems require WM resources because their correct solution depends on the suppression of the heuristic system (System I) in favor of responding in accordance with the analytic system (System II). Evidence for this interpretation was provided by De Neys (2006), who presented participants with conflict and non-conflict syllogisms while also burdening their executive resources with a secondary task. Specifically, the between-subjects manipulation of WM load consisted of presenting a 3 × 3 matrix prior to each syllogism, wherein the matrix was filled with a complex four-dot pattern (high load) or with three dots on a horizontal line (low load)[1]. After making a validity judgment, participants reproduced the matrix pattern. This experimental design required them to maintain the matrix pattern in WM while reasoning. Whereas the high load condition impaired performance on conflict problems, there was no effect of load on non-conflict problems. This demonstrates that overcoming belief-logic conflict is limited by WM capacity.

WM training could also lead to improvement in deductive reasoning via its effect on fluid intelligence—typically measured using matrix reasoning tasks. Specifically, much evidence suggests that general cognitive ability and deductive reasoning are positively correlated (Stanovich and West, 2000). In addition, a recent meta-analysis demonstrated that training specifically on the *n*-back family of WM tasks leads to a small but positive effect on fluid intelligence (Au et al., 2014). Therefore, theoretically, increases in fluid intelligence could mediate the link between *n*-back training and deductive reasoning, offering an indirect route for improving the latter (**Figure 1**).

Recently, Ariës et al. (2014) investigated the combined effect of reasoning strategy and WM training on school performance. The participants for Experiment 1 were enrolled in *lower-level* Higher Secondary Education history classes. During the 6-week intervention period, participants in the control condition were taught using a "conservative" method that involved the introduction of new subjects in new paragraphs, and the answering of reasoning questions from the textbook. In contrast, for participants in the experimental condition the same material was embedded within two WM training tasks: *n*-back and the Odd One Out. This approach ensured that training was contextualized within the subject matter of the history class. For example, on each trial of the Odd One Out four historical words or pictures were presented successively on the screen, three of which were related (e.g., were drawn from agrarian civilizations) whereas the fourth was not (i.e., was a depiction of hunter-gatherer civilization). The participant had to maintain all four stimuli in WM to select the odd one out. In the *n*-back task, nouns (e.g., farming) and pictures (e.g., hieroglyphics) drawn from the content of the history class were used as stimuli.

In addition, the experimenters trained reasoning strategies using a modification of the IMPROVE method (see Mevarech and Kramarski, 2003). This intervention is designed to teach the structure of reasoning, and works by testing understanding of the problems, highlighting similarities between problems, applying strategies

[1] There was also a third no-load condition—not pertinent to the present discussion.

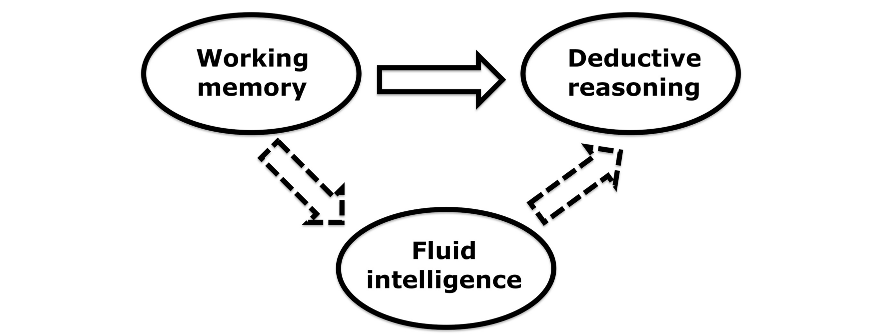


FIGURE 1 | Two possible routes for improving deductive reasoning by working memory training. The solid arrow depicts a direct effect. The dashed arrows depict an indirect effect.

for solving problems, and prompting reflection on the reasoning process. Compared to the control condition, students in the experimental condition exhibited significant gains in performance on reasoning questions in official school tests that necessitate inference making—a difference that remained significant 16 weeks after the termination of training. Subsequently, participants in Experiment 2 who were enrolled in *higher-level* Higher Secondary Education history classes received *either* WM *or* reasoning strategy training. On its own, reasoning strategy but not WM training improved school test performance.

The results of Ariës et al. (2014) suggest that for students of relatively lower ability, the combination of WM and reasoning strategy training can be a successful recipe for improving reasoning. This is likely because whereas the former enhances WM skills, the latter facilitates the acquisition of the cognitive tools for logic. For students of higher ability there might be less room for improving WM (i.e., a ceiling effect), such that learning the structure of reasoning becomes a relatively more important factor for improving performance. Although the results of the two experiments are not directly comparable because of differences in the composition of the samples and intervention strategies, they do suggest that differences in baseline ability must be taken into account while assessing transfer effects (see Jaeggi et al., 2014).

In conclusion, it appears useful to pursue the possibility that WM training could benefit deductive reasoning directly by increasing WM skills, or indirectly by increasing fluid intelligence. Critically, Ariës et al.'s successful intervention consisted of embedding WM training with domain-relevant material. It has yet to be demonstrated whether a domain-general intervention to train WM will exhibit a similar transfer profile in the context of deductive reasoning. In addition, the extent to which successful transfer to deductive reasoning will require supplementing WM training with strategy training remains an open question.

REFERENCES

Ariës, R. J., Groot, W., and van den Brink, H. M. (2014). Improving reasoning skills in secondary history education by working memory training. *Br. Educ. Res. J.* doi: 10.1002/berj.3142. [Epub ahead of print].

Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuehl, M., and Jaeggi, S. M. (2014). Improving fluid intelligence with training on working memory: a meta-analysis. *Psychon. Bull. Rev.* doi: 10.3758/s13423-014-0699-x. [Epub ahead of print].

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nat. Rev. Neurosci.* 4, 829–839. doi: 10.1038/nrn1201

De Neys, W. (2006). Dual processing in reasoning: Two systems but one reasoner. *Psychol. Sci.* 17, 428–433. doi: 10.1111/j.1467-9280.2006.01723.x

Jaeggi, S. M., Buschkuehl, M., Shah, P., and Jonides, J. (2014). The role of individual differences in cognitive training and transfer. *Mem. Cognit.* 42, 464–480. doi: 10.3758/s13421-013-0364-z

Melby-Lervåg, M., and Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Dev. Psychol.* 49, 270–291. doi: 10.1037/a0028228

Mevarech, Z. R., and Kramarski, B. (2003). The effects of metacognitive training versus worked-out examples on students' mathematical reasoning. *Brit. J. Educ. Psychol.* 73, 449–471. doi: 10.1348/000709903322591181

Newstead, S. E., Handley, S. J., Harley, C., Wright, H., and Farrelly, D. (2004). Individual differences in deductive reasoning. *Q. J. Exp. Psychol. A* 57, 33–60. doi: 10.1080/02724980343000116

Perkins, D. N., and Salomon, G. (1994). "Transfer of learning," in *International Handbook of Educational Research, 2nd Edn.*, eds T. Husen and T. N. Postelwhite (Oxford: Pergamon Press), 6452–6457.

Stanovich, K. E., and West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate. *Behav. Brain Sci.* 23, 645–726. doi: 10.1017/S0140525X00003435

Stephenson, C. L., and Halpern, D. F. (2013). Improved matrix reasoning is limited to training on tasks with a visuospatial component. *Intelligence* 41, 341–357. doi: 10.1016/j.intell.2013.05.006

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: 31 October 2014; accepted: 20 January 2015; published online: 06 February 2015.

Citation: Beatty EL and Vartanian O (2015) The prospects of working memory training for improving deductive reasoning. Front. Hum. Neurosci. ***9****:56. doi: 10.3389/fnhum.2015.00056*

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

Frontiers
EPFL Innovation Park, Building I • 1015 Lausanne • Switzerland
Tel +41 21 510 17 00 • Fax +41 21 510 17 01 • info@frontiersin.org
www.frontiersin.org

Find us on