Title
The Roots of Deductive Reasoning: Neuroimaging and Behavioral Investigations

Author
Coetzee, John Philip

Publication Date
2018

UNIVERSITY OF CALIFORNIA

Los Angeles

The Roots of Deductive Reasoning:

Neuroimaging and Behavioral Investigations

A dissertation submitted in partial satisfaction of the requirements

for the degree of Doctor of Philosophy in Psychology

by

John Philip Coetzee

2018

2018

ABSTRACT OF THE DISSERTATION

The Roots of Deductive Reasoning:

Neuroimaging and Behavioral Investigations

by

John Philip Coetzee

Doctor of Philosophy in Psychology

University of California, Los Angeles, 2018

Professor Martin M. Monti, Chair

Deductive reasoning has been an object of investigation in psychology for almost a century now. Yet, key questions remain unanswered. These include (but are not limited to) the relationship between deductive reasoning and other psychological processes (such as language and memory), the identity of the neurological structures that are core to the deductive process, the source of the hierarchical structures on which deduction depends, whether deduction is a modular or domain general process, and what the source is of the facilitation that is frequently observed when deductive problems are framed in different ways. In order to address these questions, this thesis will present three experiments that shed light on different aspects of deduction. Study 1 is an fMRI study which replicates prior findings regarding the relationship between language and deduction and reveals a new dissociation between deduction and working memory. Study 2, a neuromodulation study, is a logical extension of the findings regarding deduction and language from Study 1 and other prior studies. Here, a form of transcranial magnetic stimulation (TMS) is used to manipulate brain function in order to establish a causal dissociation between brain areas that support language and those believed to support deduction, specifically with regard to the hierarchical frameworks on which both language and deduction depend. Finally, in Study 3, a large online study is used to test the Social Exchange Theory of facilitated performance on the Wason Card selection Task on a more diverse sample than has been the case in the past, especially with regard to the cues that have been suggested to trigger the underlying cognitive "modules." This study also tests the relationship between facilitation on the Wason Task and a number of individual differences, revealing novel associations with personality traits and with psychopathology. Together, these three studies provide a clearer picture of how deductive reasoning, one of our most distinctively human capacities, is situated amongst our other cognitive abilities.

The dissertation of John Philip Coetzee is approved.

Marco Iacoboni

Barbara Knowlton

Keith Holyoak

Martin M. Monti, Committee Chair

University of California, Los Angeles

2018

DEDICATION

To everyone who is without language (but still has thoughts).

TABLE OF CONTENTS

LIST OF FIGURES

Figures for Study 1

Figures for Study 2

Figures for Study 3

ACKNOWLEDGMENTS

I would like to acknowledge the generous assistance and support of Martin M. Monti, Marco Iacoboni, Barbara Knowlton, Keith Holyoak, Allan D. Wu, Micah A. Johnson, Leo Christov-Moore, Evan Lutkenhoff, Jeff Chiang, Amy Zhong Sheng Zheng, Nicco Reggente, John Dell'Italia, Andrew Westphal, Matthew Rosenberg, Keela Thomson, Mikey Garcia, Youngzie Lee, Elliot Kim, Risha Sanikommu, Joan Kim, Sydney Kim, Nina Standing, Madeline Gavin, Ariana Taghaddos, Darin Williams, and Alice Wong. I would also like to acknowledge the support and assistance of Elizabeth Coetzee, Jan Coetzee, Mark Coker, Stephen Long, Youngji Kim, Evan Sperber, Michael Loebs, D Coetzee, Krishni Burns, Katie Allen, Corinna Richards, James Hill, Erica Gutierrez, and the LA Metro System. This is by no means an exhaustive list.

VITA

EDUCATION Ph.D. Candidate Psychology, University of California, Los Angeles
M.A. Psychology, University of California, Los Angeles, 2013
B.A. Psychology, University of California, Berkeley, 2006

PUBLICATIONS

In preparation Coetzee, J. P., Johnson, M., Wu, A., Iacoboni, M., Monti, M. M. (in preparation) Dissociating language and thought in human reasoning.

2018 **Coetzee, J. P.**, & Monti, M. M. (2018). At the core of reasoning: Dissociating deductive and non-deductive load. *Human brain mapping*.

2016 Hough, C. M., Bersani, F. S., Mellon, S.H., Epel, E. S., Reus, V.I., Lindqvist, D., Lin, J., Mahan, L., Rosser, R., Burke, H., **Coetzee, J.**, Nelson, J. C., Blackburn, E. H., Wolkowitz, O. M. (2016). Leukocyte telomere length predicts SSRI response in major depressive disorder: A preliminary report. *Molecular neuropsychiatry*, 2(2), 88-96.

2006 Gruber, Eidelman, Talbot, **Coetzee** & Harvey. Emotion Reactivity and Regulation Using Cognitive Reappraisal in Bipolar Disorder. UC Berkeley. 2006. (Research poster)

RESEARCH EXPERIENCE

Winter 2018 **GSR for Professor Martin M. Monti.**

2011 – **Graduate Student – University of California, Los Angeles**

present **Research Advisor: Professor Martin M. Monti**

2010 - 2011 **Study Coordinator - University of California, San Francisco**
Supervisor: Professor Owen Wolkowitz, MD

2009 - 2010 **Research Assistant - University of California, San Francisco**
Supervisor: Professor Steven Batki, MD

2009 - 2010 **Study Coordinator - University of California, San Francisco**
Supervisor: Professor Angela Waldrop, PhD
“Impulsivity Related to Cocaine Dependence and Trauma”

2007 - 2009 **Study Coordinator - University of California, San Francisco**
Supervisor: Professor Owen Wolkowitz, MD

2005 - 2006 **Research Assistant - University of California, Berkeley**
Supervisor: June Gruber (under Professor Allison Harvey)

HONORS & AWARDS

2012 Received a UCLA graduate student research mentorship (GSRM) to do research over the summer.

2011 Received a UCLA award of $30K to cover the first year of my PhD program.

2006 Received a competitive summer employment award of $2500 on the basis of recommendations from my supervisors in the Sleep Lab at UC Berkeley.
2005 Made the Academic Honor Roll at UC Berkeley in the Spring and Fall semesters.

UNIVERSITY TEACHING EXPERIENCE

2012 - 2017 Served as a Teaching Assistant (and later Teaching Associate) for a variety of upper division classes in the UCLA Dept. of Psychology, including:
- Psych 100B, Research Methods
- Psych 115, Principles of Behavioral Neuroscience
- Psych 116, Behavioral Neuroscience Lab
- Psych 120A, Intro to Cognitive Psychology
- Psych 120B, Sensation & Perception
- Psych 121, Laboratory in Cognitive Psychology
- Psych/Neurosci M119L, Human Neuropsychology

Background

Deductive Reasoning

Deductive reasoning is that psychological process (or set of processes) by which we attempt to use the rules of logic to infer a conclusion that is necessarily true from some set of premises that are assumed to be true (Rips, 1994; Simon, 1996). For example, consider the following syllogism:

1) If the building is retrofitted it will survive an earthquake.

2) The building has been retrofitted.

3) The building will survive an earthquake.

So long as premise 1 and 2 are assumed to be true, premise 3 is necessarily true, meaning that the argument is *valid*. This validity depends only on the logical structure of the argument, and not on the contents themselves. Although deriving such an implied conclusion from some preexisting set of statements adds no new information, and only makes clear what was already there, it can often reveal important relationships of which we were previously unaware.

We must be careful, however, not to conflate deductive reasoning with logic itself, since there may not be anything distinctively “deductive” about the mental steps by which human minds actually work through such arguments (Rips, 1994; Schechter, 2013). It is the nature of these underlying processes which have been the subject of almost a century of psychological research (Wilkins, 1929), and which will be the subject of the experiments described here. Given the central role that deductive reasoning plays in many of our most important cultural enterprises, such as science, law, and philosophy, a more precise naturalistic understanding of deductive reasoning is of intrinsic scientific interest and can shed much light on the nature of human thought.

Deductive reasoning and working memory

Ever since Wason's work in the mid-20th century (1966), we have known that naive reasoners have much more difficulty with solving or evaluating modus tollens ($P \rightarrow Q, \neg Q, \therefore \neg P$) than they do with modus ponens ($P \rightarrow Q, P, \therefore Q$). There have been many attempts at explaining why. One of the most influential theories has been that evaluating modus tollens requires us to build a kind of "model" of the syllogism using spatial working memory, which is necessarily much more taxing and error prone than evaluating modus ponens, for which the necessary model is much simpler (Johnson-Laird, 2010). To understand why this might be, it is necessary to take a brief diversion into two theoretical perspectives that have dominated much of the past half century's attempts to explain this discrepancy, as well as a third more recent one: mental logic, mental models, and dual process theory.

Mental logic

For those who endorse a version of mental logic (M. D. Braine, 1994; Osherson, 1975; Rips, 1983) we possess a set of innate mental "rules" that allow us to perform some logical inferences quickly, accurately, and effortlessly, while inferences for which we lack such a rule must be performed using more effortful and error prone processes. With regard specifically to observed performance differences between modus ponens and modus tollens, the explanation given by Rips (1990) is that while we have an innate mental rule for solving modus ponens, we lack such a rule for modus tollens, and so for the latter must use the limited rules that we have at our disposal to derive a proof through a more complex and difficult process. This process, however, is not thought to place any special demands on working memory, since at each step of the proof what has to be represented in the mind is fairly simple (Gilhooly, Logie, Wetherick, & Wynn, 1993).

Mental models

On the other hand, those who endorse some version of mental models (Johnson-Laird, 2010; Knauff, 2009; Thagard, 2010) assume that deductive reasoning relies not on the application of sets of innate rules, but on the application of schematic spatial representations. Under this view, the gap in performance between modus ponens and modus tollens can be explained by referring to the number of spatial models that must be constructed in order to arrive at the answer. For modus ponens, only one model must be constructed. For modus tollens, on the other hand, we must conduct a whole set of models as we search for a *reductio ad absurdum* to falsify one or more of the syllogism’s premises (Johnson-Laird, 2010). Unlike mental logic, mental models has often been explicitly linked to the use of working memory, especially spatial working memory (Fangmeier & Knauff, 2009; Fangmeier, Knauff, Ruff, & Sloutsky, 2006; Johnson-Laird, 2010; Philip Nicholas Johnson-Laird, 1983; Knauff, Mulack, Kassubek, Salih, & Greenlee, 2002).

Dual process

Increased reliance on working memory during modus tollens as opposed to modus ponens is also compatible with dual process theory as advanced by Evans (Evans, Over, & Manktelow, 1993) which holds that reasoning should be partitioned into two types of rationality, which Evans labels rationality$_1$ and rationality$_2$, with the former consisting of automatic or overlearned cognitive behaviors, while the latter consists of cognitive behaviors that require the careful use of logic, abstraction, and hypothetical thinking. It is worth noting that rationality$_1$ and rationality$_2$ bear substantial similarity to System 1 and System 2 as introduced by Stanovich (1999) and later taken up by Kahneman (2011). In addition to being slower and more effortful than rationality$_{1,}$ rationality$_2$ is also proposed to require more cognitive resources, especially with regard to working memory, and so differences in working memory capacity would be expected to largely explain differences in ability to do deductive reasoning (Evans, 2003). Evans argues that differential demands on rationality$_1$ and rationality$_2$ can be used to explain not only differences in performance on modus ponens and modus tollens, but also the belief bias effect. The dual process theory proposed by Evans appears to be very similar to mental models, but Evans claims to be agnostic regarding mental logic and mental models, and believes that his theory is compatible with both (Evans, 2003). It is also worth noting that although he claims that rationality$_1$ and rationality$_2$ place differential demands on working memory, those demands are not specifically on spatial working memory, as is the case with mental models (Johnson-Laird, 2010; Knauff, 2009).

Associational evidence

If our ability to engage in complex deductive reasoning, such as that necessary for modus tollens, is dependent on working memory, then it should be possible to correlate aspects of working memory with performance on such tasks. While substantial research exists showing strong correlations between working memory and fluid general intelligence, beginning with studies by Kyllonen and Christal (1990) (also see Unsworth, Fukuda, Awh, & Vogel, 2014) there has been less research on the more specific association between deductive reasoning and working memory. For example, Süß, Oberauer, Wittmann, Wilhelm, and Schulze (2002) gave tests of intelligence and working memory to 128 young German adults and determined that working memory capacity was strongly predictive of reasoning ability, but the category of “reasoning” included both inductive and deductive reasoning, as well as construction. Given that there is good reason to assume that inductive and deductive reason may draw on somewhat different cognitive abilities (Heit & Rotello, 2010), it is difficult to draw conclusions from studies such as this one about the specific relationship between deductive reasoning and working memory.

Behavioral evidence

A study by Barrouillet and Lecas (1999) did look for a specific relationship between working memory and deductive reasoning. They had three sets of thirty children from the 3rd, 6th, and 9th grades undergo tests of working memory and attempt to construct cases that obeyed a rule of conditional logic. The found that the working memory span was even more predictive of the number of correct cases generated than age was (Barrouillet & Lecas, 1999). However, the fact that working memory was measured using a simple test of number span makes it difficult to extend their results to spatial working memory.

Studies by Toms, Morris, and Ward (1993) and by Gilhooly et al. (1993) took a somewhat more sophisticated approach to testing working memory. Both made use of a dual task interference paradigm and tested various aspects of the Baddely and Hitch model of working memory (1974). Both studies reported evidence that loading the central executive impaired deduction (selectively for modus tollens in the Toms et al., (1993) study) but that loading the visuospatial sketch pad or the articulatory loop had no effect. Gilhooly et al., (1993) further pointed out that many of their participants appeared to have relied on an atmosphere heuristic, a common problem in such studies (Monti, Osherson, Martinez, & Parsons, 2007).

Neuroimaging

The earliest significant neuroimaging studies of deductive reasoning were PET studies conducted by Vinod Goel and colleagues (Goel, Gold, Kapur, & Houle, 1997; 1998). While there may have been early hope that such studies could resolve lingering debates about how deductive reasoning is implemented, in the ensuing years this kind of research has instead produced results that are highly diverse and conflicting, with virtually every part of the brain having been implicated at one time or another, and an persistent lack of replication (Monti & Osherson, 2012). There are different approaches to explaining this diversity. Monti and Osherson (2012) attribute it to differing study designs, control tasks, and stimuli content, while (Goel, 2007) has speculated that deduction itself might be a highly fractionated process. Prado, Chadha, and Booth (2011) conducted a quantitative meta-analysis on the basis of which they argued that deductive reasoning should be broken down by type of reasoning (categorical, relational, propositional) but given the aforementioned inconsistencies amongst these studies it is debatable how much sense it makes to include all or even most of them in one meta-analysis.

One imaging study that sought to directly test the relationship between deductive reasoning and spatial working memory is reported in (Knauff et al., 2002). Twelve participants judged the validity of conditional and relational syllogisms while undergoing a functional magnetic resonance imaging (fMRI) scan. One example of such a conditional syllogism is:

> If the man is in love, then he likes pizza.
> The man is in love.
> Does it follow:
>   The man likes pizza?

All stimuli were presented aurally via pneumatic headphones. Although (Knauff et al., 2002) claimed evidence of activation in an occipito-parietal-frontal network, consistent with reliance on some form of visuospatial working memory, the fact that their baseline condition was fixation on a cross makes it difficult to interpret their study, given that there is no way to know what cognitive activities participants were engaged in while at rest (Monti et al., 2007). Additionally, conditional syllogisms based on modus ponens and modus tollens, which may rely on different cognitive mechanisms given the performance differences between them, were analyzed together, further complicating the interpretation of their results.

A similar fMRI study by Fangmeier et al. (2006) attempted to separate deductive inference into three stages: premise processing, integration, and validation. Their study reported that this process appeared to begin in temporal cortex, moved to frontal cortex, and then finished in parietal cortex, with only parietal cortex showing a difference in activation from baseline. Although the baseline in this study (a memory task) was well-matched to the experimental stimuli, the experimental stimuli consisted of "linear syllogisms" such as the following:

Premise 1: A D
Premise 2: D F
Conclusion: A F

Inferences of this sort would very likely induce the participant to engage in spatial visualizing of some sort (thinking of A to the left of D and D to the left of F) and indeed Fangmeier et al. (2006) say that they designed it to be a form of spatial reasoning, because they felt this would be easily understood by naive reasoners. Unfortunately, when such inherently spatial tasks are used, it becomes impossible to say whether the parietal activation seen in the study has anything to do with deductive reasoning (Monti & Osherson, 2012).

A pair of experiments with special relevance to the first experiment reported in this thesis are those reported in Monti et al. (2007), in which participants were asked to judge the validity of conditional syllogisms. A two-level cognitive load design was used in which the activations resulting from simple inferences (modus ponens) were subtracted from the activation resulting from more complex and/or difficult inferences (modus tollens). Sentences were linguistically matched, with each condition including one set of syllogisms about blocks and another matched one about “abstract” syllogisms using invented meaningless words. It was assumed that complex syllogisms (those structured around modus tollens) would recruit the same operations as the simple ones, but to a greater degree. The second experiment was done as a replication study in which syllogisms about blocks and abstractions were replaced with syllogisms about houses and faces. Subtraction of activation associated with simple syllogisms from activation associated with complex syllogisms produced patterns of activation that were distinct from reading, and instead showed activity in specific areas of frontal cortex, such as left frontopolar cortex, several bilateral frontal regions, and in left parietal cortex, and insula (Monti et al., 2007). Activation in the second experiment was very similar, with the exception that the complex-simple contrast using house content also revealed activation in the parahippocampal place area (Monti et al., 2007).

Monti et al. (2007) interpreted the results of these two studies as revealing a set of content dependent regions which were not specifically associated with deductive reasoning (such as the fusiform face area and parahippocampal place area), as well as a content independent network that was actually responsible for carrying out the deduction. This network, which included left hemisphere prefrontal , inferior, superior frontal, and parietal regions, was further separable into a content-independent “core” regions (including left frontopolar BA10 and medial BA8) as well as a set of support regions, which although still relatively content-independent, have the function of representing information that the core regions are conducting deductive operations on (Monti et al., 2007). These included left ventrolateral prefrontal cortex for the maintenance of verbal information, and right parietal and left occipital cortex for those arguments that involved shapes like blocks and balls, suggesting that parietal activation was only necessary to the extent that it was needed to represent the mental imagery associated with these shapes. Since right parietal activation was not consistently revealed by the complex minus simple contrast, the results do not appear to support mental models theory, a point that is also made by the authors. The strong activation associated with BA10p and medial BA8 however, is compatible with mental logic, given what we know about the involvement of these regions in the representation and implementation of abstract and hierarchical rules.

Results similar to those obtained by Monti et al. (2007) were reported by Rodriguez-Moreno and Hirsch (2009). In this study, categorical syllogisms were presented to 12 participants in both the visual and aural sensory domains as they underwent an fMRI. Such a design made it possible, by using a conjunction technique, to ignore brain regions associated with low level sensory processes and reveal the higher-level “supramodal” regions which implemented cognitive processes that were common to evaluating syllogisms experienced in either sense domain. The control task was to view the same syllogisms but merely search for a specific word. Activations associated with the control task were later subtracted from activations associated with reasoning and the conjunction of the visual and aural sense domains. Analysis revealed a supramodal pattern of activation that excluded low level sensory areas and which emerged during the display of the second premise and remained active through the display of the conclusion (Rodriguez-Moreno & Hirsch, 2009). Included in this pattern were left supplementary motor area, dorsolateral prefrontal cortex, middle frontal BA10, inferior parietal lobules, bilateral caudate nuclei, and several others. Active areas were considered “core” if they were active during presentation of premises and conclusion during reasoning trials but not during control trials, and were considered “support” regions if they were active during both control and reasoning tasks. These “core” regions included left superior frontal cortex (BA6/8), right medial frontal gyrus (BA8), and bilateral parietal lobule (BA39/40/7). “Support” regions, on the other hand, included left superior frontal gyrus (BA6), middle frontal gyrus (BA 8/6, BA9&10), and left inferior frontal gyrus (BA47). Importantly, the regions included by Rodriguez-Moreno and Hirsch (2009) as core regions are very similar to those included as such in Monti et al. (2007). The only significant difference between the set of core regions described in these two papers is parietal lobule, which Rodriguez-Moreno and Hirsch (2009) include as a core region while Monti et al. (2007) does not.

Deductive reasoning and language

Deductive reasoning appears to be a form of rule-based thought, one which bears a family resemblance to several other forms of rule-based thought, such as inductive reasoning, language, moral cognition, and mathematical cognition (while many more forms of cognition are perhaps *describable* by some rule, I am here only considering those forms of cognition which are, at least in part, *constituted* by such rules, such that they would be incoherent without them). Knowledge of the relationship between deductive reasoning and these other forms of thought, and the extent to which it is a distinct mode of cognition, is crucial for us to have a clear picture of what deductive reasoning is and of the role that it plays in human cognitive life.

Amongst all the forms of rule-based thought we are capable of, language appears to be the oldest, and has the strongest case for being an inborn product of evolution (Hauser, Chomsky, & Fitch, 2002). This being the case, it is important to ask whether deductive reasoning (and other forms of rule-based cognition) are outgrowths of our ability to comprehend linguistic rules (Clark, 1969; Falmagne, 1990), or are instead independent forms of cognition, unrelated to language (Monti & Osherson, 2012). This has strong parallels with debates about whether language is constitutive of thought in general, or whether thought precedes language and has some underlying structure of its own (Fodor, 1975).

Theoretical perspectives

The approach taken in this thesis to the relationship between deductive reasoning and language falls within the framework outlined in Monti and Osherson (2012), in which the authors surveyed over a decade's worth of neuroimaging studies on deductive reasoning and concluded that the frequent appearance of language centers in the results of such research (especially BA44/45 in the inferior frontal gyrus (IFG), also known as Broca's area) was most likely an artifact caused by the failure of these studies adequately control for the activity of reading . It is easy to see why it would be tempting to view BA44/45 as central to deductive reasoning, given the importance of this region for linguistic syntax (Friederici, 2011), and the apparent resemblance between the rules of linguistic syntax and the rules of deductive logic (in the sense that both are hierarchical) (Tettamanti & Weniger, 2006). Monti and Osherson (2012) argued that if we consider only studies that have carefully controlled for activation related to reading we see that activation in the linguistic areas of the IFG drop out after an initial brief encoding stage. After this encoding stage, the information is passed to a different set of regions that operate on the underlying logical framework of the argument.

Neuroimaging studies

Monti et al. (2007) was discussed above in the context of deductive reasoning and working memory, but it also highly relevant to questions about the relationship between deductive reasoning and language. Because the stimuli for this study were linguistically matched, and because the control task (a set of inferences organized around modus ponens) were likely to elicit just as much reading as the comparison task (a set of inferences organized around modus tollens) it was possible to say with some clarity what the role of language in deductive reasoning might be. As described previously, the pattern of activation revealed when simple reasoning tasks were subtracted from complex reasoning tasks included no activation near the inferior frontal gyrus, but did reveal activation in a set of other left lateralized frontal content-independent “core” regions, and a somewhat more widely dispersed set of content-independent and content dependent support regions (Monti et al., 2007).

The Rodriguez-Moreno and Hirsch study (2009), also mentioned in the previous section, is also relevant to the question of language and deduction. The authors were very clear that once activation associated with their control task (finding a word in the same stimuli) was subtracted off of the conjunction of activation areas from reasoning during visual and aural presentation of the stimuli, activation in the inferior frontal gyrus disappeared almost entirely. The sole exception was some activation in BA47/45 during the audio presentation of conclusions, but the authors point out that BA47/45 is not within the area traditionally thought of as Broca’s area (Rodriguez-Moreno & Hirsch, 2009).

In Monti, Parsons, and Osherson (2009) the relationship between language and deductive reasoning was addressed more directly. In this fMRI study, 15 participants were presented with pairs of statements involving the items X, Y, and Z, and were asked to judge either whether the statements formed a valid argument or whether they both had the same linguistic meaning. In addition to these deductive and linguistic inferences, there was also a baseline task in which participants were asked to judge the grammatical correctness of the same set of arguments (with the addition of 16 grammatically incorrect statements). In subsequent subtractions of activations associated with the baseline grammar task from activations associated with the linguistic inference task, areas in the left hemisphere traditionally associated with language were revealed, such as BA44/45 and the posterior superior temporal gyrus (Friederici, 2011; Pulvermüller, 2010). Such areas were, however, absent from the parallel subtraction involving deductive inference. There, the pattern of activation revealed locations familiar from previous studies by this group, including left frontopolar (BA10p) and medial superior (BA8) prefrontal cortex. There was some overlap between the two patterns in regions known to scale with cognitive demand, such as parietal cortex and parietal lobule (Carpenter, Just, Keller, Eddy, & Thulborn, 1999; Klingberg, O’Sullivan, & Roland, 1997).

When considering the argument that the linguistic centers might play a central role in deductive reasoning, the strongest evidence is that presented by Reverberi and his collaborators (Reverberi et al., 2007, 2010). In Reverberi et al. (2007), an fMRI study was reported in which 14 participants were scanned while evaluating very simple syllogisms that were either conditional or disjunctive. Contents of either syllogism referred to simple abstract shapes, and the logical form of the conditional syllogisms was always based on modus ponens. Premises were presented one at a time, and then the participant was asked to select the correct conclusion from a list. Furthermore, the premises of some syllogisms were unrelated to each other, rendering the argument non-integrable. A disjunction analysis was used such that areas of activation had to show greater activation for integrable than non-integrable premises, an effect of integration for both types of syllogisms, and greater activation in response to the disjunctive than to the conditional syllogisms. Only two areas of activity met all three requirements: left inferior frontal gyrus and left precentral gyrus (BA44 & 6), with the second being the left inferior parietal lobule (BA40). The authors also found that activity in these areas was correlated with the amount of time spent reasoning about the stimuli.

Another study by Reverberi et al. (2010) used fMRI to investigate the processing of conditional and categorical syllogisms in a group of 25 participants. As a baseline, participants viewed the same syllogisms but were only asked to try and memorize them, not evaluate them for validity. Analysis was similar to Reverberi et al. (2007), involving a conjunction and then a contrast with baseline. Once again, the syllogisms used were simple, with one of the conditional structures being:

P1 If a thing is x then it is y
P2 If a thing is y then it is z
Correct conclusion: If a thing is x then it is z

Results of the analyses once again implicated left BA44/45 (as well as left BA6 and BA7), which Reverberi and colleagues argued supports their view that the language centers of the left IFG are crucial for deductive reasoning.

It is possible that the results obtained by Reverberi et al. (2007 & 2010) are a consequence of the relatively simple stimuli used. Given the substantial differences in response time and accuracy that have long been evident in behavioral studies for modus ponens and modus tollens (Wason & Johnson-Laird, 1972), it is entirely plausible that relatively simple and relatively complex forms of reasoning rely on substantially different neural mechanisms, with modus ponens being akin to mere reading while more complex forms of reasoning engage an entirely different set of processes.

Patient studies

In Reverberi, Shallice, D'Agostini, Skrap, and Bonatti (2009) neurological patients were tested and results obtained that were somewhat at odds with Reverberi et al. (2007 & 2010). Thirty six patients with focal brain lesions in either left, right, or medial frontal cortex were given tests of working memory and verbal ability, and were then asked to read a set of 30 short stories (ranging from 24 to 68 words each) in which an abstract problem was embedded that could be solved via a few reasoning steps (the direct reasoning routine described in Braine (1990). Participants were asked to judge whether the conclusions of these stories were valid, and to judge how difficult the problem had been. Results indicated that ability to judge the validity of the story conclusions was impaired by lesions to left and medial frontal cortex, but not by lesions to right frontal cortex. However, all patients in the study performed at ceiling on a verbal abilities test, so whatever damage they had in left frontal cortex did not appear to impair verbal abilities. Additionally, the effect of working memory deficits depended on where the lesion was located. Patients with left frontal lesions were impaired at the deductive task if they also had impaired working memory, but suffered no deductive impairment if their working memory was unharmed. Medial frontal lesion patients, by contrast, were impaired at both deductive accuracy and judgments of how complex (difficult) the tasks were, and this impairment was independent of working memory function. The apparent importance of medial frontal cortex for deductive reasoning revealed in this study is compatible with the results of other studies (Canessa et al., 2005; Monti et al., 2007; Rodriguez-Moreno & Hirsch, 2009).

TMS studies

One disadvantage of fMRI studies of cognitive function is that, because of the correlative nature of the technique, it can be difficult to draw causal connections from structure to function (Ramsey et al., 2010). Lesion studies appear to avoid this difficulty, but brain damage is rarely as focal as researchers would like, and the plasticity of the brain after damage also creates obstacles (Rorden & Karnath, 2004). Transcranial magnetic stimulation (TMS) holds the promise of making it possible to transiently disrupt (or facilitate) localized brain regions during a cognitive task, and thereby to gain clearer evidence of the causal necessity of that region for a given task (Schutter, Van Honk, & Panksepp, 2004). In a field which has seen as much conflicted neuroimaging evidence as the study of deductive reasoning has (Goel, 2007; Monti & Osherson, 2012; Prado et al., 2011) this is an especially appealing advantage.

Two studies by Tsujii and colleagues are of particular interest. In Tsujii, Sakatani, Masuda, Akiyama, & Watanabe (2011), repetitive TMS (rTMS) was used on two groups with 24 participants each to investigate the role of three regions in deductive reasoning: left IFG, right IFG, and bilateral superior lobule (SL). The task involved categorical syllogisms with content that was either abstract, congruent (with the participant's knowledge about the world), or incongruent. After administration of rTMS for 10 minutes at a frequency of 1Hz to bilateral parietal lobule impaired performance was observed on both abstract and incongruent (but not congruent) syllogisms (Tsujii et al., 2011). However, application of the same treatment to left IFG only impaired performance on congruent syllogisms, while facilitating performance on incongruent syllogisms (i.e. reducing the belief bias effect) (Evans, Barston, & Pollard, 2013). Conversely, application of rTMS to right IFG impaired performance on incongruent syllogisms, but had no other effect.

With regard to the IFG, these results are mirrored in an earlier study by (Tsujii, Masuda, Akiyama, & Watanabe, 2010) in which the participants either received rTMS at a frequency of 1Hz for 10 minutes, or a sham treatment, to either left or right IFG in 72 participants. As in Tsujii et al. (2011), application of rTMS to right IFG impaired the ability to evaluate incongruent syllogisms, but not congruent ones, while applying the treatment to left IFG impaired the ability to evaluate congruent but not incongruent syllogisms. Additionally, also as in Tsujii et al. (2011), applying rTMS to left IFC also resulted in an attenuation of the belief bias effect.

The two studies by Tsujii, when considered alongside the other studies reviewed above, indicate that the language centers in left IFG are unlikely to be sufficient for performing processes that we would recognize as deductive reasoning. Rather, it is more plausible that the language centers become engaged for processes that are so simple that they are little more than reading, or perhaps constitute something like the rationality$_1$ described by Evans (2003). For deductions that are even moderately complex, it appears other frontal areas are just as, if not more, important (Monti et al., 2007, 2009; Reverberi et al., 2009; Rodriguez-Moreno & Hirsch, 2009). It is true that Tsujii et al. (2011) implicates the parietal lobule in more abstract or difficult reasoning, but given the results from Reverberi et al. (2009), this finding is consistent with a view of the parietal lobule as necessary for representing the premises of a syllogism but not being “core” to the deductive process (Monti & Osherson, 2012).

Deductive Reasoning as a modular process

One of the most prominent theories advanced in recent decades in favor of a distinctive character for deductive reasoning is social exchange theory (Cosmides, 1985; Cosmides & Tooby, 1992, 2005). Under this theory, the performance facilitation seen in untrained participants when a deductive argument (specifically a form of the Wason selection task (Wason, 1968; Wason, 1966)) is framed in terms of the enforcement of a social rule is explained as being due to a specialized social exchange “module” that evolution has bequeathed to us for the purpose of dealing with the complex and important social obligations that our species is so good at generating (Cosmides & Tooby, 1992, 2005). A competing theory, the permission schema theory advanced by Cheng & Holyoak (1985), explains such facilitation as having less to do with the possession of a specialized module, and more to do with the activity of more general mechanisms, such as our ability to understand under what circumstances we have permission to do something.

Theoretical perspectives

Social exchange theory was designed to explain the “facilitation effect” which had long been observed in the Wason selection task (Wason, 1968; Wason, 1966), in which participants displayed extremely low rates of accuracy at solving the abstract version of the task, but much higher rates of accuracy when the task was contextualized as a familiar social rule (Cosmides & Tooby, 1992). Cosmides put forward a proposal that this effect was the result of a special psychological “module” within our reasoning capacities which was an evolutionary product of our need to detect those who violated social rules during the long span of human prehistory that was spent in small hunter gatherer tribes (Cosmides, 1985). Specifically, this module was created by natural selection to detect violations that possessed two structural properties: 1) the taking of some benefit, while 2) failing to pay some associated cost (Cosmides, 1985). In more recent publications, Cosmides and associates seem to have dropped the idea that social exchange facilitation requires a cost, instead saying that the necessary elements are a benefit to be had and a requirement to be met (Cosmides, Barrett, & Tooby, 2010). The structural elements are also distinct from the “cues” that activate the presumed model, which are described as including an intentional act, a benefit to be had, and the possibility of cheating (Cosmides et al., 2010).

In later years this theory was expanded to include a second “hazard management” module, which natural selection is presumed to have devised for purposes of facilitating the processing of those scenarios possessing two properties: 1) taking some risk, 2) taking some prescribed precaution (Fiddick, Cosmides, & Tooby, 2000). Once again, two cues have been identified that are distinct from the structural elements of the task and are thought to activate the module in question (Cosmides et al., 2010).

Behavioral evidence for social exchange theory

The earliest studies to test social exchange theory were those reported in Cosmides (1985). In a series of 6 experiments conducted on a total of 165 Harvard undergraduates, Cosmides demonstrated a consistent tendency of participants to choose the logically correct cards (P, ¬Q) on the Wason selection task when the task was framed by accompanying text as a social exchange involving a cost and a benefit, except for those trials in which the positions of the cost and benefit were switched, in which case participants chose the logically fallacious cards (¬P, Q) at a high rate, despite those cards rarely being chosen under abstract or descriptive circumstances (Cosmides, 1985).

In another series of behavioral experiments Cosmides (1989) sought to test her theory against the more abstract pragmatic reasoning schemas theory of Cheng and Holyoak (1985). There were five experiments, with four of them using forty undergraduates each, while one used twenty. All participants were Stanford undergraduates. One specific purpose of these experiments was to demonstrate that permission schemas, a type of pragmatic reasoning schema, was unable to explain facilitated performance on the Wason selection task without the benefit-cost machinery that is integral to social contract theory, a goal that appeared to borne out by the higher rate of facilitation for rules framed as a cost-benefit social exchange as opposed to those framed as a permission schema (Cosmides, 1989). A second purpose was to show that social contract rules were capable of inducing high rates of social exchange conforming answers even in cases of “switched” rules, in comparison to permission schemas, and once again appeared to have shown this (criticisms of the these experiments and their interpretation will be addressed later).

The two groups of studies described above appear to be the largest direct tests of social exchange theory by Cosmides and colleagues, although her results were replicated by Gigerenzer and Hug (1992) in a study involving 93 undergraduates. Subsequently, Cosmides (usually with her collaborator Tooby) has elaborated social exchange theory and tested it in other ways, including applying it to the hunter-horticulturist Shiwiar of the Ecuadorian Amazon (Sugiyama, Tooby, & Cosmides, 2002), who exhibited a tendency to turn over the same “cheater relevant” cards as the Harvard undergraduates, but who also turned over “cheater irrelevant” cards (cards in which no benefit was accepted, or in which the requirement was met) at a high rate, a behavior which was attributed to “curiosity” (Sugiyama et al., 2002). Another novel application of social contract theory involved a patient, RM, with bilateral limbic damage who was impaired on social contract scenarios but not on those involving a hazard (although given a sample size of one it is difficult to know how much weight to give this study) (Stone, Cosmides, Tooby, Kroll, & Knight, 2002).

Hazard management (precaution) theory

It later became apparent that there are other scenarios that appear to create a facilitation effect on the Wason selection task, those involving a hazard for which some precaution must be taken (Cosmides & Tooby, 1997; Fiddick et al., 2000). In one experiment from Fiddick et al. (2000) 60 undergraduates were divided into two equal groups, and each was presented with a Wason selection task paired with the rule “if you go hunting, then you wear these orange jackets to avoid being shot.” For the standard group, this rule was contextualized by instructions to merely test the truth of the claim by observing what other hunters were doing, while in the hazard/precaution condition the rule was contextualized by instructions to see which hunters were “endangering themselves” (Fiddick et al., 2000). A significantly larger number of participants were facilitated to choose the correct cards (corresponding to P and ¬Q) in the hazard/precaution condition. Cosmides and Tooby have described hazard/precaution management and social exchange theory as constituting two independent facilitatory modules for reasoning about such rules (Cosmides & Tooby, 2005).

Criticism of social exchange theory

As Evans et al. (1993) have pointed out, Cosmides’ theory has undergone significant criticism. Cheng and Holyoak (1989) have pointed out that there is much equivocation in Cosmides’ theory regarding the importance of “paying a cost” in return for a benefit, as opposed to merely meeting a requirement. For example, in such well known cases as the facilitation induced by enforcing a drinking age rule (Griggs & Cox, 1982) there does not appear to be any “cost” to be paid in return for the benefit of drinking alcohol, apart perhaps from the cost of being older. Also, in such circumstances, it is hard to argue that any “exchange” is taking place (Cheng & Holyoak, 1989). A similar problem arises with the cholera scenario (Cheng & Holyoak, 1985) in which someone passing through an immigration office at the Manila International Airport can only enter if they possess a stamp indicating that they have the necessary inoculations. Despite high degrees of facilitation, there is no apparent cost and thus no exchange, only the meeting of a permission requirement. It is possible that the decision to drop the necessity of a cost from social exchange theory (it is absent in Cosmides et al., (2010)) was a response to such criticism.

Another line of criticism comes from Sperber and Girotto (2003) in which they argue that even if a cheater detection module does exist, the Wason selection task is an inappropriate tool to test it with, since participants are very likely responding not to anything like the logical modus tollens structure of the Wason selection task, but instead to the particular details of the social scenario that is created to accompany it. Cosmides' own success at inducing participants to violate the logical structure of the task in the "switched" version of her task appears to support this view (Cosmides, 1985).

Overview of Studies

Study 1. We conducted an fMRI experiment employing a fast, event related design. Twenty healthy volunteers were asked to evaluate matched simple and complex deductive and non-deductive arguments in a 2×2 design. The contrast of complex versus simple deductive trials resulted in a pattern of activation closely matching previous work, including frontopolar and frontomedial "core" areas of deduction as well as other "cognitive support" areas in frontoparietal cortices. Conversely, the contrast of complex and simple non-deductive trials (designed to instead load working memory) resulted in a pattern of activation that does not include any of the aforementioned "core" areas. Direct comparison of the load effect across deductive and non-deductive trials further supports the view that activity in the regions previously interpreted as "core" to deductive reasoning cannot merely reflect non-deductive load, but instead might reflect processes specific to the deductive calculus. Additionally, language areas in left inferior frontal gyrus and posterior temporal cortex do not appear to participate in deductive inference beyond their role in encoding stimuli presented in linguistic format.

Study 2. Designed as a follow up to and an extension of Study 1, in Study 2 we sought to verify our findings regarding the non-involvement of language in deductive reasoning, by employing TMS. This approach allowed us to directly manipulate brain function, and therefore made it possible for us to draw causal conclusions regarding the role (or lack thereof) of language in deductive reasoning. To this end, we designed our study to test two competing hypotheses. According to the first, the structure-dependent operations required for deductive reasoning are linguistic in nature and are based on the neural mechanisms of natural language (Reverberi et al., 2007, 2010). According to the second, deductive reasoning is supported by language-independent processes primarily implemented in frontomedial and left frontopolar cortices (BA8 and 10, respectively) (Coetzee & Monti, 2018; Monti et al., 2007, 2009; Rodriguez-Moreno & Hirsch, 2009). We employed continuous Theta Burst Stimulation (cTBS (Huang, Edwards, Rounis, Bhatia, & Rothwell, 2005)), a type of patterned TMS, which has been demonstrated to have a transient inhibitory effect on neural function (Christov-Moore, Sugiyama, Grigaityte, & Iacoboni, 2017).Using this method we found that transient cTBS inhibition of Broca’s area impairs accuracy for linguistic problems but not for logic ones. Conversely, cTBS to frontomedial cortices produces the opposite pattern.

Study 3. We sought here to test two other aspects of deductive reasoning. First, whether the evolved cognitive modules proposed by Cosmides and Tooby (1989) under Social Exchange Theory are accurate descriptions of the performance facilitation that is classically observed when the Wason Card Selection Task is framed using certain types of content. Additionally, we sought to test the association of facilitation on this task with several measures of individual differences. We were especially interested in a prediction made by Scott and Baron-Cohen (1996) that autistic individuals should fail to show the kind of facilitation in response to social exchange scenarios that has been demonstrated with normal participants. To this end, we administered a series of tests to online participants in the United States and India, with the tests being hosted on the Qualtrics platform and a total of 417 participants being recruited through the Amazon Mechanical Turk system. We found that although scenarios framed in terms of social exchange or hazard management. The presence or absence of the cues which Cosmides has suggested are necessary to activate the modules did not have any detectable effect. Additionally, the individual scores on an autism measure did have an association with performance on the Wason task, but in the opposite direction from what had been predicted. This was also true for psychopathy.

References

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. *The Psychology of Learning and Motivation*, *8*, 47–89.

Barrouillet, P., & Lecas, J.-F. (1999). Mental Models in Conditional Reasoning and Working Memory. *Thinking & Reasoning*, *5*(4), 289–302. https://doi.org/10.1080/135467899393940

Braine, M. D. (1994). Mental logic and how to discover it. In J. Macnamara & G. E. Reyes (Eds.) (pp. 241–263). New York, NY, US: Oxford University Press. Retrieved from http://psycnet.apa.org/psycinfo/1995-97494-009

Braine, M. D. S. (1990). The natural logic approach to reasoning. *Reasoning, Necessity, and Logic: Developmental Perspectives*, *133157*.

Canessa, N., Gorini, A., Cappa, S. F., Piattelli-Palmarini, M., Danna, M., Fazio, F., & Perani, D. (2005). The effect of social content on deductive reasoning: An fMRI study. *Human Brain Mapping*, *26*(1), 30–43. https://doi.org/10.1002/hbm.20114

Carpenter, P. A., Just, M. A., Keller, T. A., Eddy, W., & Thulborn, K. (1999). Graded Functional Activation in the Visuospatial System with the Amount of Task Demand. *Journal of Cognitive Neuroscience*, *11*(1), 9–24. https://doi.org/10.1162/089892999563210

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, *17*(4), 391–416. https://doi.org/10.1016/0010-0285(85)90014-3

Cheng, P. W., & Holyoak, K. J. (1989). On the natural selection of reasoning theories.

Cognition, *33*(3), 285–313. https://doi.org/10.1016/0010-0277(89)90031-0

Christov-Moore, L., Sugiyama, T., Grigaityte, K., & Iacoboni, M. (2017). Increasing generosity by disrupting prefrontal cortex. *Social Neuroscience*, *12*(2), 174–181. https://doi.org/10.1080/17470919.2016.1154105

Clark, H. H. (1969). Linguistic processes in deductive reasoning.

Coetzee, J. P., & Monti, M. M. (2018). At the core of reasoning: Dissociating deductive and non-deductive load. *Human Brain Mapping*, *39*(4), 1850–1861. https://doi.org/10.1002/hbm.23979

Cosmides, L. (1985). Deduction or Darwinian Algorithms? An Explanation of the "Elusive" Content Effect on the Wason Selection Task. Retrieved from http://www.researchgate.net/publication/33882408_Deduction_or_Darwinian_Algorithms_An_Explanation_of_the_Elusive_Content_Effect_on_the_Wason_Selection_Task

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, *31*(3), 187–276. https://doi.org/10.1016/0010-0277(89)90023-1

Cosmides, L., Barrett, H. C., & Tooby, J. (2010). Colloquium paper: adaptive specializations, social exchange, and the evolution of human intelligence. *Proceedings of the National Academy of Sciences of the United States of America*, *107 Suppl*(Supplement_2), 9007–14. https://doi.org/10.1073/pnas.0914623107

Cosmides, L., & Tooby, J. (1989). Evolutionary psychology and the generation of culture, part II. *Ethology and Sociobiology*, *10*(1–3), 51–97. https://doi.org/10.1016/0162-3095(89)90013-7

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange (pp. 163–228).

Cosmides, L., & Tooby, J. (1997). Dissecting the computational architecture of social inference mechanisms. *Characterizing Human Psychological Adaptations*, 132–156.

Cosmides, L., & Tooby, J. (2005). Neurocognitive adaptations designed for social exchange (pp. 584–627).

Evans, J. S. B. T. (2003). In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*(10), 454–459. https://doi.org/10.1016/j.tics.2003.08.012

Evans, J. S. B. T. B., Barston, J. L., & Pollard, P. (2013). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, *11*(3), 295–306. https://doi.org/10.3758/BF03196976

Evans, J. S. B. T., Over, D. E., & Manktelow, K. I. (1993). Reasoning, decision making and rationality. *Cognition*, *49*(1–2), 165–187. https://doi.org/10.1016/0010-0277(93)90039-X

Falmagne, R. J. (1990). Language and the acquisition of logical knowledge. *Reasoning, Necessity, and Logic: Developmental Perspectives*, 111–131.

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

Fangmeier, T., Knauff, M., Ruff, C. C., & Sloutsky, V. (2006). FMRI evidence for a three-stage model of deductive reasoning. *Journal of Cognitive Neuroscience*, *18*(3), 320–34. https://doi.org/10.1162/089892906775990651

Fiddick, L., Cosmides, L., & Tooby, J. (2000). No interpretation without representation: the role of domain-specific representations and inferences in the Wason selection task. *Cognition*, *77*(1), 1–79. https://doi.org/10.1016/S0010-0277(00)00085-8

Fodor, J. A. (1975). *The language of thought* (Vol. 5). Harvard University Press.

Friederici, A. D. (2011). The Brain Basis of Language Processing: From Structure to Function. *Physiological Reviews*, *91*(4), 1357–1392. https://doi.org/10.1152/physrev.00006.2011

Gigerenzer, G., & Hug, K. (1992). Domain-specific reasoning: Social contracts, cheating, and perspective change. *Cognition*, *43*(2), 127–171. https://doi.org/10.1016/0010-0277(92)90060-U

Gilhooly, K. J., Logie, R. H., Wetherick, N. E., & Wynn, V. (1993). Working memory and strategies in syllogistic-reasoning tasks. *Memory & Cognition*, *21*(1), 115–24. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8433642

Goel, V. (2007). Anatomy of deductive reasoning. *Trends in Cognitive Sciences*, *11*(10), 435–441. https://doi.org/10.1016/j.tics.2007.09.003

Goel, V., Gold, B., Kapur, S., & Houle, S. (1997). The seats of reason? An imaging study of deductive and inductive reasoning. [Miscellaneous Article]. *Neuroreport March 24, 1997*, *8*(5), 1305–1310. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=ovftc&AN=00001756-199703240-00049

Goel, V., Gold, B., Kapur, S., & Houle, S. (1998). Neuroanatomical correlates of human reasoning. *Journal of Cognitive Neuroscience*, *10*(3), 293–302. https://doi.org/10.1162/089892998562744

Griggs, R. A., & Cox, J. R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology*, *73*(3), 407–420. https://doi.org/10.1111/j.2044-8295.1982.tb01823.x

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science*, *298*(5598), 1569–1579. https://doi.org/10.1126/science.298.5598.1569

Heit, E., & Rotello, C. M. (2010). Relations between inductive reasoning and deductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(3), 805.

Huang, Y. Z., Edwards, M. J., Rounis, E., Bhatia, K. P., & Rothwell, J. C. (2005). Theta burst stimulation of the human motor cortex. *Neuron*, *45*(2), 201–206. https://doi.org/10.1016/j.neuron.2004.12.033

Johnson-Laird, P. N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Harvard University Press. Retrieved from http://books.google.com/books?id=FS3zSKAfLGMC

Johnson-Laird, P. N. (2010). Mental models and human reasoning. *Proceedings of the*

National Academy of Sciences, *107*(43), 18243–18250. https://doi.org/10.1073/pnas.1012933107

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Klingberg, T., O'Sullivan, B. T., & Roland, P. E. (1997). Bilateral activation of fronto-parietal networks by incrementing demand in a working memory task. *Cerebral Cortex*, *7*(5), 465–471. https://doi.org/10.1093/cercor/7.5.465

Knauff, M. (2009). A Neuro-Cognitive Theory of Deductive Relational Reasoning with Mental Models and Visual Images. *Spatial Cognition & Computation*, *9*(2), 109–137. https://doi.org/10.1080/13875860902887605

Knauff, M., Mulack, T., Kassubek, J., Salih, H. R., & Greenlee, M. W. (2002). Spatial imagery in deductive reasoning: A functional MRI study. *Cognitive Brain Research*, *13*(2), 203–212. https://doi.org/10.1016/S0926-6410(01)00116-1

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, *14*(4), 389–433. https://doi.org/10.1016/S0160-2896(05)80012-1

Monti, M. M., & Osherson, D. N. Logic, language and the brain, 1428 Brain Research § (2012). https://doi.org/10.1016/j.brainres.2011.05.061

Monti, M. M., Osherson, D. N., Martinez, M. J., & Parsons, L. M. (2007). Functional neuroanatomy of deductive inference: A language-independent distributed network. *NeuroImage*, *37*(3), 1005–1016. https://doi.org/10.1016/j.neuroimage.2007.04.069

Monti, M. M., Parsons, L. M., & Osherson, D. N. (2009). The boundaries of language and thought in deductive inference. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(30), 12554–9. https://doi.org/10.1073/pnas.0902422106

Osherson, D. N. (1975). Logic and models of logical thinking. *Reasoning: Representation and Process in Children and Adults*, 81–91.

Prado, J., Chadha, A., & Booth, J. R. (2011). The Brain Network for Deductive Reasoning: A Quantitative Meta-analysis of 28 Neuroimaging Studies. *Journal of Cognitive Neuroscience*, *23*(11), 3483–3497. https://doi.org/10.1162/jocn_a_00063

Pulvermüller, F. (2010). Brain embodiment of syntax and grammar: Discrete combinatorial mechanisms spelt out in neuronal circuits. *Brain and Language*, *112*(3), 167–179. https://doi.org/10.1016/j.bandl.2009.08.002

Ramsey, J. D., Hanson, S. J., Hanson, C., Halchenko, Y. O., Poldrack, R. A., & Glymour, C. (2010). Six problems for causal inference from fMRI. *NeuroImage*, *49*(2), 1545–58. https://doi.org/10.1016/j.neuroimage.2009.08.065

Reverberi, C., Cherubini, P., Frackowiak, R. S. J., Caltagirone, C., Paulesu, E., & Macaluso, E. (2010a). Conditional and syllogistic deductive tasks dissociate functionally during premise integration. *Human Brain Mapping*, *31*(9), 1430–1445. https://doi.org/10.1002/hbm.20947

Reverberi, C., Cherubini, P., Frackowiak, R. S. J., Caltagirone, C., Paulesu, E., & Macaluso, E. (2010b). Conditional and syllogistic deductive tasks dissociate functionally during premise integration. *Human Brain Mapping*, *31*(9), 1430–1445. https://doi.org/10.1002/hbm.20947

Reverberi, C., Cherubini, P., Rapisarda, A., Rigamonti, E., Caltagirone, C., Frackowiak, R. S. J. J., … Paulesu, E. (2007a). Neural basis of generation of conclusions in elementary deduction. *NeuroImage*, *38*(4), 752–762. https://doi.org/10.1016/j.neuroimage.2007.07.060

Reverberi, C., Cherubini, P., Rapisarda, A., Rigamonti, E., Caltagirone, C., Frackowiak, R. S. J., … Paulesu, E. (2007b). Neural basis of generation of conclusions in elementary deduction. *NeuroImage*, *38*(4), 752–762. https://doi.org/10.1016/j.neuroimage.2007.07.060

Reverberi, C., Shallice, T., D’Agostini, S., Skrap, M., & Bonatti, L. L. (2009). Cortical bases of elementary deductive reasoning: Inference, memory, and metadeduction. *Neuropsychologia*, *47*(4), 1107–1116. https://doi.org/10.1016/j.neuropsychologia.2009.01.004

Rips, L. J. (1983). Cognitive processes in propositional reasoning. *Psychological Review*, *90*(1), 38–71. https://doi.org/10.1037/0033-295X.90.1.38

Rips, L. J. (1990). Reasoning. *Annual Review of Psychology*, *41*(1), 321–353. https://doi.org/10.1146/annurev.ps.41.020190.001541

Rips, L. J. (1994). *The psychology of proof : deductive reasoning in human thinking*. MIT Press. Retrieved from https://mitpress.mit.edu/books/psychology-proof

Rodriguez-Moreno, D., & Hirsch, J. (2009). The dynamics of deductive reasoning: An fMRI investigation. *Neuropsychologia*, *47*(4), 949–961. https://doi.org/10.1016/j.neuropsychologia.2008.08.030

Rorden, C., & Karnath, H.-O. (2004). Using human brain lesions to infer function: a relic from a past era in the fMRI age? *Nature Reviews. Neuroscience*, *5*(10), 813–9. https://doi.org/10.1038/nrn1521

Schechter, J. (2013). Deductive reasoning. *Encyclopedia of the Mind. Los Angeles, CA: SAGE Publications*.

Schutter, D. J. L. G., Van Honk, J., & Panksepp, J. (2004). Introducing Transcranial Magnetic Stimulation (TMS) and its Property of Causal Inference in Investigating Brain-Function Relationships. *Synthese*, *141*(2), 155–173. https://doi.org/10.1023/B:SYNT.0000042951.25087.16

Scott, F. J., & Baron-Cohen, S. (1996). Logical, analogical, and psychological reasoning in autism: A test of the Cosmides theory. *Development and Psychopathology*, *8*(01), 235–245. https://doi.org/10.1017/S0954579400007069

Simon, M. A. (1996). Beyond inductive and deductive reasoning: The search for a sense of knowing. *Educational Studies in Mathematics*, *30*(2), 197–210. https://doi.org/10.1007/BF00302630

Sperber, D., & Girotto, V. (2003). Does the selection task detect cheater-detection? (pp. 197–226).

Stanovich, K. E. (1999). *Who is rational?: Studies of individual differences in reasoning*. Psychology Press.

Stone, V. E., Cosmides, L., Tooby, J., Kroll, N., & Knight, R. T. (2002). Selective impairment of reasoning about social exchange in a patient with bilateral limbic system damage. *Proceedings of the National Academy of Sciences*, *99*(17), 11531–11536. https://doi.org/10.1073/pnas.122352699

Sugiyama, L. S., Tooby, J., & Cosmides, L. (2002). Cross-cultural evidence of cognitive adaptations for social exchange among the Shiwiar of Ecuadorian Amazonia. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(17), 11537–42. https://doi.org/10.1073/pnas.122352999

Süß, H.-M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability—and a little bit more. *Intelligence*, *30*(3), 261–288. https://doi.org/10.1016/S0160-2896(01)00100-3

Tettamanti, M., & Weniger, D. (2006). Broca's area: A supramodal hierarchical processor? *Cortex*, *42*(4), 491–494. https://doi.org/10.1016/S0010-9452(08)70384-8

Thagard, P. (2010). How Brains Make Mental Models. In L. Magnani, W. Carnielli, & C. Pizzi (Eds.) (pp. 447–461). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-15223-8_25

Toms, M., Morris, N., & Ward, D. (1993). Working memory and conditional reasoning. *The Quarterly Journal of Experimental Psychology Section A*, *46*(4), 679–699. https://doi.org/10.1080/14640749308401033

Tsujii, T., Masuda, S., Akiyama, T., & Watanabe, S. (2010). The role of inferior frontal cortex in belief-bias reasoning: An rTMS study. *Neuropsychologia*, *48*(7), 2005–2008. https://doi.org/10.1016/j.neuropsychologia.2010.03.021

Tsujii, T., Sakatani, K., Masuda, S., Akiyama, T., & Watanabe, S. (2011). Evaluating the roles of the inferior frontal gyrus and superior parietal lobule in deductive reasoning: An rTMS study. *NeuroImage*, *58*(2), 640–646. https://doi.org/10.1016/j.neuroimage.2011.06.076

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: capacity, attention control, and secondary memory retrieval. *Cognitive Psychology*, *71*, 1–26. https://doi.org/10.1016/j.cogpsych.2014.01.003

Wason, P. C. (1966). Reasoning (Vol. 1, pp. 135–151).

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, *20*(3), 273–281. https://doi.org/10.1080/14640746808400161

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content* (Vol. 86). Harvard University Press.

Wilkins, M. C. (1929). The effect of changed material on ability to do formal syllogistic reasoning. *Archives of Psychology*, *102*, 83. Retrieved from http://psycnet.apa.org/psycinfo/1929-04403-001

Study 1: Dissociating Deductive and Non-Deductive Load

Introduction

The mental representations and processes underlying deductive reasoning in humans have long been discussed in the psychological literature (e.g., Braine & O'Brien, 1998; Johnson-Laird, 1999; Osherson, 1975; Rips, 1994). Over the past 18 years, non-invasive neuroimaging methods have played a growing role in investigating how the human mind achieves deductive inferences (see Monti & Osherson, 2012; Prado, Chadha, & Booth, 2011 for a review). At least in the context of propositional and categorical problems, two main hypotheses have emerged concerning the localization of the neural substrate of this ability. On the one hand, some have proposed that the centers of language, and in particular regions within what has traditionally been referred to as Broca's area, in the left inferior frontal gyrus, are critical to deductive reasoning. Reverberi and colleagues, for example, have shown that in the context of simple propositional and categorical inferences, premise integration consistently activates the left inferior frontal gyrus (particularly within Brodmann areas (BA) 44 and 45; see Reverberi et al., 2007, 2010). This region has thus been suggested to be involved in the extraction and representation of the superficial and formal structure of a problem, with other frontal regions, such as the orbital section of the inferior frontal gyrus (i.e., BA47), also contributing to representing the full logical meaning of an argument (Baggio et al., 2016; Reverberi et al., 2012).

On the other hand, a number of studies have failed to uncover any significant activation within Broca's area for propositional and categorical deductive inferences (Canessa et al., 2005; Kroger, Nystrom, Cohen, & Johnson-Laird, 2008; Monti, Osherson, Martinez, & Parsons, 2007; Monti, Parsons, & Osherson, 2009; Noveck, Goel, & Smith, 2004; Parsons & Osherson, 2001; Rodriguez-Moreno & Hirsch, 2009), even under much more naturalistic experimental conditions (Prado et al., 2015). An alternative hypothesis has thus been proposed under which logic is subserved by a set of language-independent regions within frontopolar (i.e., BA10) and frontomedial (i.e., BA8) cortices, among others (Monti & Osherson, 2012; Monti et al., 2007, 2009). Consistent with this proposal, neuropsychological investigations have shown that lesions extending at through medial BA8 are sufficient to impair deductive inference-making, as well as meta-cognitive assessments of inference complexity, despite an anatomically intact Broca's Area and ceiling performance on neuropsychological assessments of language function (Reverberi, Shallice, D'Agostini, Skrap, & Bonatti, 2009).

As discussed elsewhere, a number of experimental factors might help reconcile the divergence of results, including (i) the complexity of the deductive problems (typically much greater in studies failing to uncover activation in Broca's area), (ii) the specific task employed to elicit deductive reasoning, such as argument generation, which is correlated with detecting activation in the left inferior frontal gyrus, versus argument evaluation, which correlates with failing to uncover such activation, and (iii) the degree to which participants are trained prior to the experimental session (see Monti & Osherson, 2012 for a review). In the present work we will mainly address the first point with respect to the possibility that activity in the so-called "core" regions of deductive inference (and particularly left frontopolar cortex) might in fact reflect, partially or entirely, increased load on non-deductive processes (e.g., greater working memory demands) imposed by hard deductions (cf., Kroger et al., 2008; Prado, Mutreja, & Booth, 2013). For, inasmuch as complex deductions impose greater load on deductive processes, and thereby increase the need for branching, goal-subgoal processing, and the simultaneous consideration of multiple interacting variables, as is likely in the load design used in Monti et al. (2007), the “workspace” of working memory must also undergo increased load (Halford, Wilson, & Phillips, 2010).

Of course, there is broad agreement that deductive inference is best understood as relying on a broader “cascade of cognitive processes requiring the concerted operation of several, functionally distinct, brain areas” (Reverberi et al., 2012, p. 1752; see also discussion in Prado et al., 2015) beyond those we considered above. Posterior parietal cortices, for example, are often observed across propositional, categorical, and relational deductive syllogisms, although with a preponderance of unilateral activations in propositional and categorical deductions and bilateral activations for relational problems (e.g., Knauff, Fangmeier, Ruff, & Johnson-Laird, 2003; Knauff, Mulack, Kassubek, Salih, & Greenlee, 2002; Kroger et al., 2008; Monti et al., 2009; Noveck et al., 2004; Prado, Der Henst, & Noveck, 2010; Reverberi et al., 2007, 2010; Rodriguez-Moreno & Hirsch, 2009). Dorsolateral frontal regions also appear to be recruited across different types of deductive problems (e.g., Fangmeier, Knauff, Ruff, & Sloutsky, 2006; Knauff et al., 2002; Monti et al., 2007, 2009; Reverberi et al., 2012; Rodriguez-Moreno & Hirsch, 2009) and might possess interesting hemispheric asymmetries in their functional contributions to deductive inference making (see discussion in Prado et al., 2015). Nonetheless, the present study is specifically meant to address the significance of regions that have been previously proposed to encapsulate processes which lie at the heart of deductive reasoning (Monti et al., 2007, 2009) vis-à-vis the concern that it is "[…] not clear whether the results found by Monti et al. [2007] reflect deduction or working memory demands" (Kroger et al., 2008, p. 90).

In what follows, we report on a 3T functional magnetic resonance imaging (fMRI) study in which we tested whether the pattern of activations that has been previously characterized as "core" to deductive inference (Monti et al., 2007, 2009; see Monti & Osherson, 2012 for a review) might instead reflect activity resulting from increased general non-deductive cognitive load (a view we will label the *general cognitive load hypothesis*). Specifically, using a 2 × 2 design (simple/complex, deductive/non-deductive), we compare the effect of non-deductive (e.g., working memory) versus deductive load on the putative "core" regions of deduction. Using a forward inference approach (Henson, 2006; see also Heit, 2015 for a discussion on the relevance of this approach to the field of the neural basis of human reasoning), under the general complexity hypothesis, activity in frontopolar and frontomedial cortices should be elicited equally by deductive and non-deductive load. Conversely, under our previous interpretation of these regions, deductive load alone should elicit activity within these areas. As we report below, contrary to the general cognitive load hypothesis, non-deductive load fails to elicit significant activation in the so-called "core" regions of deduction which, in fact, appear sensitive to the interaction of load and deductive problems, further establishing the genuine tie between these areas and cognitive processes that sit at the heart of the deductive (specifically, propositional) inference.

Methods

Participants

Twenty (eight female) undergraduates from the University of California, Los Angeles (UCLA), participated in the study for monetary compensation after giving written informed consent, in accordance with the Declaration of Helsinki and with the rules and standards established by the UCLA Office of the Human Research Protection Program. All participants were right-handed native English speakers with no prior history of neurological disorders and had no prior formal training in logic. Ages ranged from 18 to 23 years old (*M* = 20.4 years, *SD* = 1.4).

Before being enrolled in the fMRI component of the study participants underwent a screening procedure similar to that used in Canessa et al. (2005) in which they had to achieve an accuracy level of 60% or better on each of the four task types to be performed in the neuroimaging session (i.e. the simple and complex deductive and non-deductive tasks, see below). Prospective participants were recruited via the UCLA Psychology Department’s SONA system, which provides a means for undergraduates to participate in psychology experiments for academic credit. The pre-test, in which participants received no feedback on either individual problems or overall performance, utilized the same experimental stimuli as those employed in the neuroimaging session, although presented in a different random order, and occurred on average 5.3 months (*SD* = 4.6 mo) prior to the fMRI session, to minimize the possibility of practice effects. A correlational analysis showed that there was no relationship between the amount of time that elapsed between the screening and imaging visit and the participants’ accuracy at the imaging session (both overall and individually for any subtask).

Importantly, participants did not undergo training on the task (e.g., of the sort used, for example, in Reverberi et al., 2007).

Task

In each trial, participants were presented with a triplet of sentences (henceforth “arguments”), asked to engage in a deductive or non-deductive task, and then, after a brief (3 s) delay, prompted to respond (see Figure 1 for a depiction of the task timeline).

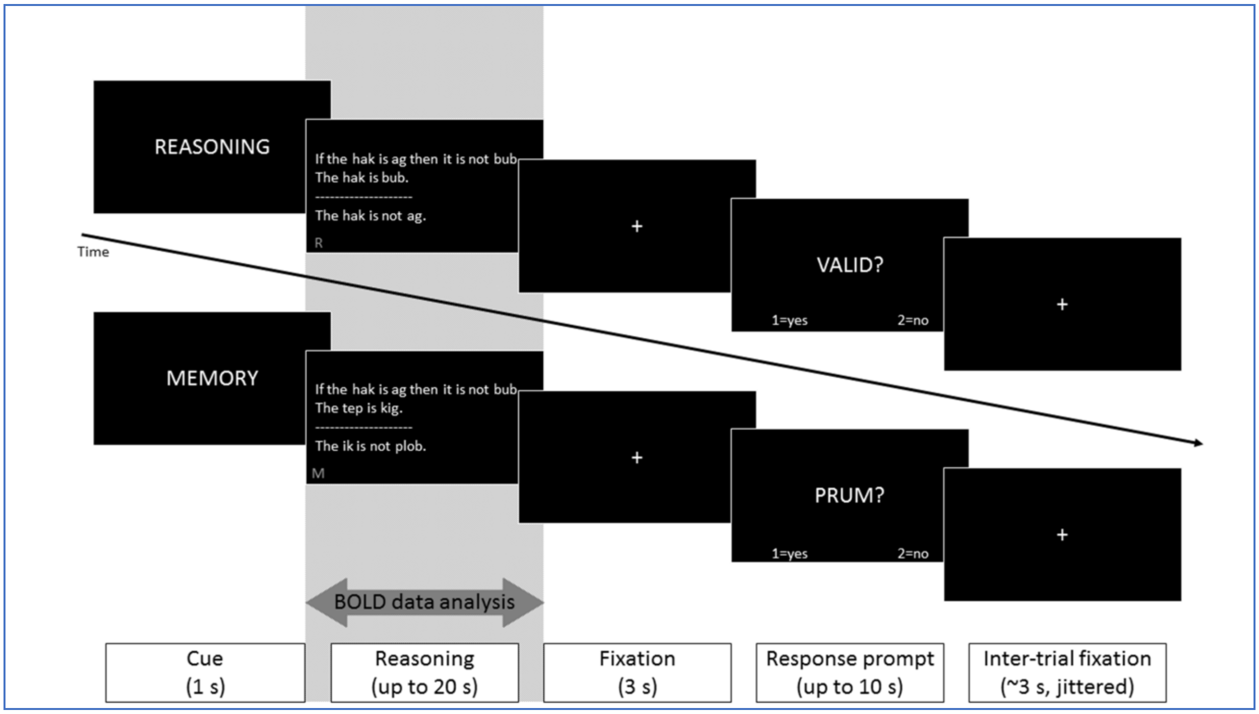


Figure 1. Experimental design. Sample deductive (complex; top) and non-deductive (simple; bottom) trials. The gray area highlights the period relevant for the BOLD data analysis (In the original displays the bottom left reminder cue appeared in red).

In the deductive condition, participants were asked to assess the logical status (i.e., (in)validity) of the argument, which they were probed on at the response prompt. Following previous work (Monti et al., 2007), half the arguments were subjectively simple to assess while the remainder were subjectively complex (henceforth simple and complex deductive conditions, respectively; see below for further details). While we adopted an empirical definition of deductive load, it is clear that the simple deductive arguments, which were structured around the logical form of *modus ponens* (which refers to inferences that conform to the valid propositional rule, “If P then Q; P; therefore Q”) were low in both general cognitive load and relational complexity (defined as the “the number of interacting variables represented in parallel” Halford, Wilson, & Phillips, 1998, p. 851). Conversely, the complex deductive arguments, which were structured around the logical form of *modus tollens* (inferences that conform to the valid propositional rule, “If P then Q; not Q; therefore not P”), were higher in both relational complexity and in general cognitive load than the simple deductive arguments. (We take it to be uncontroversial that, under the view that working memory serves “as the workspace where relational representations are constructed,” Halford et al., 2010, increased relational complexity must be accompanied by increased working memory demands.) In the non-deductive trials, participants were asked to retain in memory the presented sentence triplet or a mentally manipulated version of the triplet (*simple* and *complex non-deductive* conditions, respectively) for a later recognition test. The response prompt contained a fragment which participants had to respond to by determining whether it was part of the presented argument (in simple non-deductive trials) or the *mentally manipulated* argument (in complex non-deductive trials). The rule for manipulating the argument in the complex non-deductive trials required participants to mentally replace the first term of the first statement with the last term of the third statement, and the last term of the first statement with the last term of the second statement (see Figure 2 for an example). It is this mentally manipulated version of the argument that participants were then to retain for the ensuing recognition test. This complex non-deductive task was specifically designed in order to elicit abstract, but non-deductive, rule-based manipulation of the stimuli and tax participants' cognitive resources (as confirmed by the response time and accuracy analysis reported below).

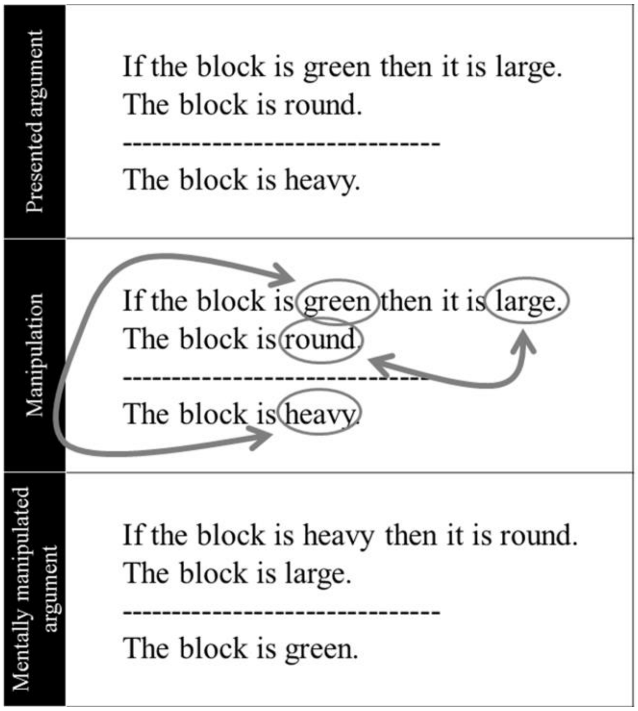


Figure 2. Example mental manipulation trial. Top: presented argument; middle: rules for manipulating the triplet; bottom: argument participants had to retain in memory for the later response prompt.

Stimuli

The stimuli for this experiment consisted of 128 arguments. The 64 arguments employed for the deductive trials were obtained from 8 logical forms validated in our previous work (Monti et al., 2007; see also Table 1). The eight logical forms give rise, when instantiated into English (see below), to four pairs of logical arguments of which two are deductively valid and two are deductively invalid. Within each pair, arguments were linguistically matched but of varying deductive difficulty, with one argument appearing consistently more difficult than the other in terms of response time, subjective complexity judgment, and binary forced choice (see Monti et al., 2007). Arguments were instantiated into natural language by replacing the logical connectives (i.e., $\rightarrow$, $\neg$, $\wedge$, $\vee$) with their standard English translation (i.e., “If … then …”, “not”, “and”, “or”) and by replacing logical variables (i.e., `p, q, r’) with features of an imaginary block (e.g., “large”, “round”, “green”; henceforth *concrete* instantiation) or meaningless pseudo-words (e.g., “kig”, “frek”, “teg”; henceforth *abstract* instantiation) modeled on the items of the Nonsense Word Fluency subtest of the DIBELS assessment (Good, Kaminski, Smith, Laimon, & Dill, 2004). Each logic form was instantiated four times with concrete materials and four times with abstract materials. In each instantiation, the logical variables were replaced with a different object feature or nonsense word so that each specific instantiation was unique. To illustrate, consider the following logic form:

$$p \rightarrow \neg q$$

$$p$$

$$\neg q$$

By systematically replacing the logic variables (i.e., `p, q’) with propositions describing the features of an imaginary block, we can obtain a concrete instantiation of the above logic form:

If the block is green then it is not small.

The block is green.

The block is not small.

Replacing the features of an imaginary block with nonsense pseudo-words results in an abstract instantiation of the same logic form:

If the kig is teg then it is not frek.

The kig is teg.

The kig is not freg.

The remaining 64 stimuli, employed in the non-deductive trials, were obtained by taking the deductive arguments described above and shuffling the sentences across triplets (without mixing concrete and abstract instantiations) so that, within each argument, no common variable was shared amongst all three statements. The absence of common variables across the three statements ensured that, despite being presented with stimuli superficially analogous to those featured in deductive trials, participants could not engage in any deductive inference-making. To illustrate:

If the block is large then it is not green.

The block is square.

The block is not heavy.

Abstract non-deductive instantiations were obtained analogously to the abstract deductive ones (see Figure 1 for one such example).

Table 1.

Logical forms used in stimuli, with sample arguments

Validity	**Cognitive Load**	**Formal Structure**	**Block Argument Examples**	**Nonsense Phoneme Argument Examples**
Valid	Low	$p \rightarrow \neg q$ p $\neg q$	If the block is green then it is not small. The block is green. ------------------------------ The block is not small.	If the hak is kig then it is not gop. The hak is kig. ------------------------------ The hak is not gop.
Valid	Low	$(p \vee q) \rightarrow \neg r$ p $\neg r$	If the block is either heavy or large then it is not blue. The block is heavy. ------------------------------ The block is not blue.	If the hak is either kig or ik then it is not gop. The hak is kig. ------------------------------ The hak is not gop.
Valid	High	$(p \vee q) \rightarrow \neg r$ r $\neg p$	If the block is either red or square then it is not large. The block is large. ------------------------------ The block is not square.	If the hak is either ag or bub then it is not frek. The hak is ag. ------------------------------ The hak is not frek.
Valid	High	$p \rightarrow \neg q$ q $\neg p$	If the block is large then it is not blue. The block is blue. ------------------------------ The block is not large.	If the hak is ag then it is not bub. The hak is bub. ------------------------------ The hak is not ag.
Invalid	Low	$(p \wedge q) \rightarrow \neg r$ p $\neg r$	If the block is both blue and square then it is not large. The block is blue. ------------------------------ The block is not large.	If the tep is both kig and ob then it is not plob. The tep is kig. ------------------------------ The tep is not plob.
Invalid	Low	$\neg p \rightarrow (q \vee r)$ $\neg p$ q	If the block is not red then it is either square or small. The block is not red. ------------------------------ The block is square.	If the hak is not trum then it is either stas or plob. The hak is not trum. ------------------------------ The hak is stas.
Invalid	High	$(p \wedge q) \rightarrow \neg r$ r $\neg p$	If the block is both square and small then it is not blue. The block is blue. ----------------------------- The block is not square.	If the tep is both grix and ag then it is not ik. The tep is ik. ------------------------------ The tep is not grix.
Invalid	High	$\neg p \rightarrow (q \vee r)$ $\neg q$ p	If the block is not blue then it is either small or light. The block is not small. ------------------------------ The block is blue.	If the tep is not trum then it is either frek or ob. The tep is not frek. ------------------------------ The tep is trum.

Design and Procedure

The 128 trials were distributed across four runs of 32 trials each. Within each run participants performed 8 trials per each condition (simple/complex, deductive/non-deductive). Within each run, trials were presented (pseudo-)randomly, under the sole constraint that no two consecutive trials were of the same condition. Each run featured a different pseudo-random ordering of the trials, and participants were randomly allocated to one of two different orderings of the runs.

As depicted in Figure 1, each trial started with a one second verbal cue, presented visually, informing the participant which task they were about to perform. Deductive conditions (i.e., simple and complex) were prompted with the verbal cue “REASONING,” while non-deductive conditions were prompted with the verbal cue “MEMORY,” for simple trials, and “SWITCH,” for complex trials. Next, the full argument was presented, with all three premises being displayed simultaneously. Given the randomized task order, to help participants keep track of which condition they were supposed to be performing on each trial a small red reminder cue (“R,” “M,” or “S”) was displayed in the lower left corner of the screen, alongside the argument. Each triplet was available, on-screen, for a maximum of 20 seconds, or until participants pressed the “continue” button (with their right little finger) to signal that they were ready to move to the response prompt. In the deductive trials participants would press the “continue” button upon reaching a decision as to the logical status of the argument, whereas in the non-deductive trials they would press the “continue” button upon having sufficiently encoded the presented triplet (easy condition) or the mentally manipulated triplet (complex condition). At this stage, the button-press was only employed to indicate that the participant was ready to move to the response phase, and was thus the same in all trials regardless of condition. Upon the button press, or the elapsing of the allotted 20 seconds, a fixation cross appeared, for three seconds, followed by a response prompt. In deductive trials, the response prompt displayed the question “VALID?” to which participants had to respond yes or no. In non-deductive trials, the response prompt displayed a probe to which participants had to respond yes or no as to whether it was part of the presented argument (in simple non-deductive trials) or the mentally manipulated argument (in complex non-deductive trials). In the simple non-deductive trials the probe was a single word (e.g., “GREEN?”) whereas in the complex non-deductive trials the probe was a fragment (e.g., “…is not blop. The ik …”). For example, given the argument shown in Figure 2, a correct response screen might read “…is large. The block is green.”, which is indeed part of the mentally manipulated argument (Figure 2, bottom), whereas an incorrect response screen might read “…is round. The block is heavy.”, which is not part of the mentally manipulated argument. This design allowed us to ensure that participants could not simply recognize that terms in the response prompt were present in the argument, as they could do in the simple non-deductive trials. (As we will show in the results section, the behavioral data confirm the difficulty of the task.) The response phase lasted up to a maximum of 10 seconds, and participants were instructed to answer (for all trial types equally) with the index finger to signal a ‘yes’ response and the middle finger to signal a ‘no’ response. Upon their response (or elapsing of the 10 seconds), a fixation cross signaled the end of the trial and remained on screen for a randomly jittered amount of time (on average 3 s), prior to the beginning on the following trial. The average time needed to complete each of the four runs was

12.5 minutes. The full session, inclusive of the four functional runs, as well as structural data acquisition and brief resting periods in between runs, lasted between 75 and 105 minutes.

fMRI Data Acquisition

All imaging data were acquired on a 3 Tesla Siemens Tim Trio MRI system at the Staglin IMHRO Center for Cognitive Neuroscience (CCN) in the Semel Institute for Neuroscience and Human Behavior at the University of California, Los Angeles. First, T2* sensitive images were acquired using a gradient echo sequence (TR = 2,500 ms, TE = 30 ms, FA = 81 degrees, FoV = 220 × 220) in 38 oblique interleaved slices with a distance factor of 20%, resulting in a resolution of 3 × 3 × 3.6 mm. Structural images were acquired using a T1-sensitive MPRAGE (magnetization-prepared 180 degree radio-frequency pulses with rapid gradient-echo; TR = 1,900 ms, TE = 2.26 ms, FA = 9º) sequence, acquired in 176 axial slices, at a 1-mm isovoxel resolution.

fMRI Data Analysis

Functional data were analyzed using FSL (FMRI Software Library, Oxford University; Smith et al., 2004). The first four volumes of each functional dataset were discarded to allow for the blood oxygenation level dependent (BOLD) signal to stabilize. Next, each individual echo planar imaging time series (EPI) was brain-extracted, motion-corrected to the middle time-point, smoothed with an 8 mm FWHM kernel, and corrected for autocorrelation using pre-whitening (as implemented in FSL). Data were analyzed using a general linear model inclusive of four main regressors focusing, as done in previous work, on the data acquired during the reasoning periods of each trial (cf., Christoff et al., 2001; Kroger et al., 2008). The four regressors marked the onset and duration of each condition separately (i.e., simple non-deductive, complex non-deductive, simple deductive, complex deductive). Specifically, each regressor marked the time from which an argument appeared on screen up to the button press indicating that the participant was ready to proceed to the answer prompt screen (i.e., the "Reasoning" period show in Figure 1 and shaded in gray). In other words, these four regressors captured – for each trial type – the time during which the participants were encoding and reasoning over the stimuli up to when they felt ready to go to the answer prompt screen. It is important to stress that, in order to account for response-time differences across tasks, we explicitly factor the length of each trial in our regressors, thereby adopting what is known as a "variable epoch" GLM model (Henson, 2007). This approach has been previously shown to be physiologically plausible and to have higher power and reliability for detecting brain activation (Grinband, Wager, Lindquist, Ferrera, & Hirsch, 2008), and is conventionally used in tasks where RTs are likely to vary across trials and/or conditions (see, for example, Christoff et al., 2001; Crittenden & Duncan, 2014; Strand, Forssberg, Klingberg, & Norrelgen, 2008). A conventional double gamma response function was convolved with each of the regressors in order to account for the known lag of the hemodynamic mechanisms upon which the blood oxygenation level dependent (BOLD) signal is predicated (Buxton, Uludag, Dubowitz, & Liu, 2004). Finally, to moderate the effects of motion, the six motion parameter estimates from the rigid body motion correction, as well as their first and second derivative and their difference, were also included in the model as nuisance regressors. For each run, we performed four contrasts. First, we assessed the simple effect of load for deductive (D) and non-deductive (ND) trials (i.e., [complex > simple]$_{ND}$ and [complex > simple]$_{D}$), and, following, we assessed the interaction effect of load and task type (i.e., [complex – simple]$_{ND}$ > [complex – simple]$_{D}$ and [complex – simple]$_{D}$ > [complex – simple]$_{ND}$). In order to avoid reverse activations (Morcom & Fletcher, 2007), each contrast was masked to only include voxels for which the sum of the z-score statistic of the minuend and subtrahend was equal to or greater than zero (cf., DeWolf, Chiang, Bassok, Holyoak, & Monti, 2016). For the simple effect of load in deduction, for example, the mask was created by only including voxels for which $[z_{(Complex-Fix)_D} + z_{(Simple-Fix)_D} \geq 0]$ while, for the interaction effect the mask was created by only including voxels for which $[z_{(Complex-Simple)_D} + z_{(Complex-Simple)_{ND}} \geq 0]$. This procedure ensures that a voxel cannot be found active merely because the subtrahend is negative (i.e., less active than baseline/fixation) while the minuend is not greater than zero (i.e., not more active than baseline/fixation; see DeWolf et al., 2016).

Prior to group analysis, single subject contrast parameter statistical images were coregistered to the Montreal Neurological Institute (MNI) template with a two-step process using 7 and 12 degrees of freedom. Group mean statistics were generated with a mixed effects model accounting for both the within-session variance (fixed-effects) as well as the between-session variance (random-effects) with automatic outlier de-weighting. Group statistical parameteric maps were thresholded using a conventional cluster correction, based on random field theory, determined by $Z > 2.7$ and a (corrected) cluster significance of $P < 0.05$ (Brett, Penny, & Kiebel, 2003; Worsley, 2001, 2007). Considering the recent debate on the validity of using cluster-based correction methods in fMRI analyses, it is important to point out that, given the existing published data on the topic (e.g., Woo, Krishnan & Wager, 2014; Eklund, Nichols & Knuttson, 2016), our specific approach is not expected to suffer from greater than nominal (.05) familywise error rate (FWE). Specifically, (i) we are using a conservative cluster determining threshold (CDT; $Z > 2.7$, i.e., ~ $p < 0.003$) which falls below the envelope of liberal primary thresholds which might result in an “anti-conservative bias” (i.e., 0.01 < CDT < 0.005; see Woo et al., 2014, p. 417) and, consistent with the previous point, (ii) the data presented in Eklund et al. (2016) suggest that the use of cluster correction with a CDT of $p < 0.003$, with FSL’s FLAME1 algorithm (Beckman, Jenkinson, and Smith, 2003; Woolrich et al., 2004) and with an event related design, should yield a valid analysis (i.e., within the nominal 5% error rate; cf., Eklund et al 2016, Figure 1, p. 7901).

Results

Behavioral results

Across all participants and conditions, the mean accuracy was 90.2% (*SD* = 4.94%), and the mean response time was 10.48 s (*SD* = 3.1 s). Two separate 2 × 2 repeated-measures ANOVA were conducted to test the effects of load (simple vs. complex) and task (deductive vs. non-deductive) on accuracy and response time (respectively).

With regard to accuracy (see Supplementary Figure 1a), we found a significant interaction between load and task type ($F(1, 19) = 9.03$, $p = 0.007$, $\hat{\omega}_p^2 = 0.05$) mainly driven by a larger load effect for non-deductive trials (*M* = 98%, *SD* = 2.7%, and *M* = 78.8%, *SD* = 8.7%, for simple and complex trials, respectively) as compared to that observed for deductive trials (*M* = 96.25%, *SD* = 7.56% and *M*= 87.81%, *SD* = 10.96%, respectively). (Effect sizes reported here, based on the unbiased $\hat{\omega}_p^2$, as opposed to the conventional $\hat{\eta}_p^2$, were calculated using the formulae provided by Maxwell and Delaney 2004). We also found a significant main effect of both load and task type ($F(1, 19) = 66.95$, $p < 0.001$, $\hat{\omega}_p^2 = 0.39$, and $F(1, 19) = 6.547$, $p = 0.02$, $\hat{\omega}_p^2 = 0.02$, respectively). Specifically, participants exhibited lower mean accuracy in complex trials (*M* = 83.3%), as compared to simple ones (*M* = 97.1%), and in non-deductive trials (*M* = 88.4%), as compared to deductive ones (*M* = 92.0%).

With regard to response time (see Supplementary Figure 1b), we again found a significant interaction between the load and task type ($F(1, 19) = 5.97$, $p = 0.02$, $\hat{\omega}_p^2 = 0.01$), which was mainly driven by a larger difference in response times across simple and complex trials for non-deductive materials (*M* = 8.85 s, *SD* = 4.41 s, and *M* = 15.12 s, *SD* = 3.83 s, respectively) as compared to deductive ones (*M* = 6.79 s, *SD* = 2.11 s, and *M* = 11.18 s, *SD* = 3.45 s, respectively). We also found a significant main effect for both load and task type with regard to response time ($F(1, 19) = 170.30$, $p < 0.001$, $\hat{\omega}_p^2 = 0.32$; and $F(1, 19) = 30.76$, $p < 0.001$, $\hat{\omega}_p^2 = 0.1$, respectively). Specifically, participants exhibited longer response times in complex trials (*M* = 13.15 s), as compared to simples trials (*M* = 7.82 s), and in non-deductive trials (*M* = 11.98 s), as compared to reasoning trials (*M* = 8.99 s). The response time analysis included both correct and incorrect trials (as done in previous work; e.g., Prado, Mutreja, & Booth, 2013; Reverberi et al., 2007, 2010).

Neuroimaging results

Subtraction of simple non-deductive trials from complex ones uncovered a number of significant activations across posterior parietal and frontal regions. As depicted in Figure 3a (in blue; see Table 2 for a full list of activation maxima), significant activations were uncovered in bilateral middle frontal gyri (MFG; spanning BA8 and BA6), precuneus (Prec; BA7), left lateralized superior and inferior parietal lobuli (SPL and IPL; in BA7 and BA40, respectively), and angular gyrus (AG; BA39). The subtraction of simple deductive trials from complex ones resulted in a very different pattern of activations (see Figure 3a, green areas, and Table 3 for full list of activation maxima). Consistent with previous work (Monti et al., 2007, 2009), maxima were uncovered in regions previously reported as being "core" to deductive reasoning, such as the left middle frontal gyrus (in BA10), and the bilateral medial frontal gyrus (MeFG; in BA8; also referred to as the pre-supplementary motor area, pre-SMA), as well as "support" regions of deductive inference, including bilateral MFG (spanning BA8 and BA6 in the right hemisphere, and BA9 and BA10/47 in the left), superior and inferior parietal lobuli (BA7 and BA40, respectively), left superior frontal gyrus (SFG; in BA8), and right lateralized cerebellum (in crus I). The overlap between the two load subtractions was minimal (see Figure 3a, in purple), and mostly confined to "bleed-over" at the junction of the more dorsal MFG cluster uncovered in the non-deductive comparison, and the more ventral MFG cluster observed in the deductive comparison.

Table 2.

Activations for memory complexity subtraction (i.e., complex memory minus simple memory)

MNI Coordinates					
x	y	z	Hem	Region Label (BA)	Z-score
Frontal Lobe					
-24	12	46	L	Middle Frontal Gyrus (8)	5.35
-26	2	54	L	Middle Frontal Gyrus (6)	5.24
24	12	44	R	Middle Frontal Gyrus (8)	4.89
30	2	68	R	Middle Frontal Gyrus (6)	2.93
Parietal Lobe					
-6	-62	54	L	Precuneus (7)	7.52
-6	-64	50	L	Precuneus (7)	6.92
10	-66	52	R	Precuneus (7)	6.81
-10	-68	58	L	Superior Parietal Lobule (7)	6.69
-36	-80	30	L	Angular Gyrus (39)	6.55
-36	-50	42	L	Inferior Parietal Lobule (40)	5.79

As shown in Figure 3b (see Tables 4 and 5 for full list of activation maxima), the interaction analysis revealed that non-deductive load, as compared to deductive load, specifically recruited foci in bilateral SPL, as well as left lateralized Prec (BA7), AG at the junction with the superior occipital gyrus (BA19/39), and MFG (BA6). Conversely, deductive load, as compared to non-deductive load, uncovered significant activations within bilateral MeFG (BA8 bilaterally, and BA9 in the left hemisphere), left junction of the SFG and MFG (BA10), left MFG (in BA10, BA10/47, and BA47), left SFG (BA6), left SPL and IPL (BA7, BA40, respectively), and right cerebellum (crus I).

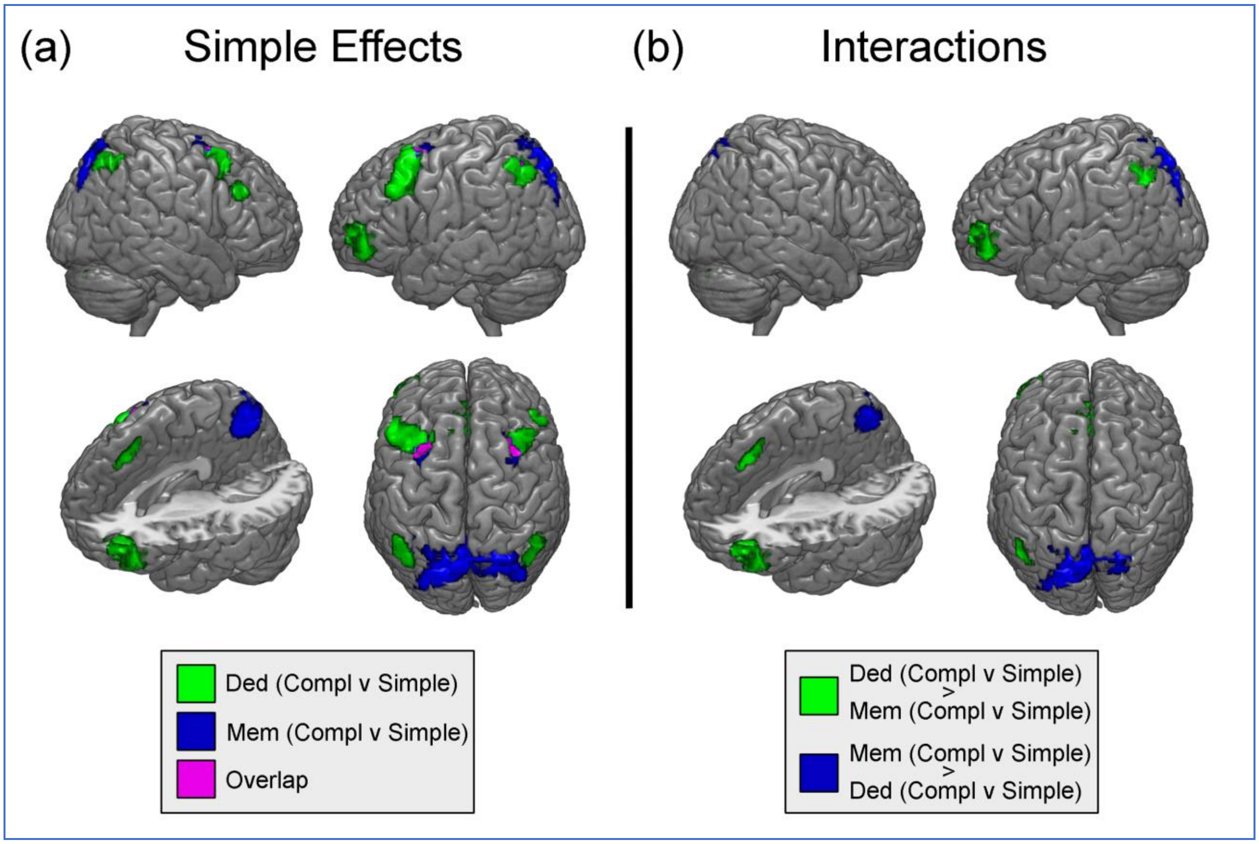


Figure 3. Imaging results: (a) Activations for the complexity subtraction for memory (in blue) and logic (in green) materials (overlap shown in purple); (b) Activations for the interaction of complexity by materials.

Discussion

The present study reports two main findings. First, we have shown that deductive reasoning is supported by a distributed set of frontoparietal regions closely matching our previous reports (Monti et al., 2007, 2009). Second, we have shown that the putative "core" regions for deductive inference cannot be interpreted as merely responding to general (i.e., non-deductive) cognitive load and/or increased working memory demands (Kroger et al., 2008; Prado et al., 2013). We discuss each point in turn.

Consistent with previous reports, our results favor the idea that deductive reasoning is supported by a distributed network of regions encompassing frontal and parietal areas (Monti et al., 2007, 2009; Rodriguez-Moreno & Hirsch, 2009; Prado et al.,

Table 3.

Activations for reasoning complexity subtraction (i.e., complex minus simple reasoning)

MNI Coordinates					
x	y	z	Hem	Region Label (BA)	Z-score
Frontal Lobe					
-42	22	36	L	Middle Frontal Gyrus (9)	5.39
-48	46	-2	L	Middle Frontal Gyrus (10/47)	4.75
-52	24	36	L	Middle Frontal Gyrus (9)	4.66
-34	16	54	L	Superior Frontal Gyrus (8)	4.38
-38	14	54	L	Superior Frontal Gyrus (8)	4.29
-8	24	50	L	Medial Frontal Gyrus (8)	4.19
38	14	50	R	Middle Frontal Gyrus (8)	4.13
-6	28	46	L	Medial Frontal Gyrus (8)	4.13
-36	54	-2	L	Middle Frontal Gyrus (10)	4.05
36	20	52	R	Middle Frontal Gyrus (8)	3.87
-6	38	38	L	Medial Frontal Gyrus (8)	3.63
-6	38	42	L	Medial Frontal Gyrus (6)	3.62
42	22	52	R	Middle Frontal Gyrus (8)	3.42
40	20	38	R	Precentral Gyrus (9)	3.33
34	18	62	R	Middle Frontal Gyrus (6)	3.32
12	26	42	R	Medial Frontal Gyrus (6)	3.22
Parietal Lobe					
-46	-54	40	L	Inferior Parietal Lobule (40)	4.03
44	-60	40	R	Inferior Parietal Lobule (40)	3.83
-44	-58	42	L	Inferior Parietal Lobule (40)	3.69
-44	-58	46	L	Inferior Parietal Lobule (40)	3.67

-44	-62	52	L	Superior Parietal Lobule (7)	3.63
-42	-64	44	L	Inferior Parietal Lobule (40)	3.60
50	-60	50	R	Inferior Parietal Lobule (40)	3.55
44	-60	58	R	Superior Parietal Lobule (7)	3.49
42	-68	46	R	Superior Parietal Lobule (7)	3.39
50	-60	46	R	Inferior Parietal Lobule (40)	3.39
Cerebellum					
36	-74	-28	R	Crus I	3.95
32	-70	-32	R	Crus I	3.69
32	-76	-28	R	Crus I	3.65
38	-68	-40	R	Crus I	3.56

2015). In light of recent discussions concerning the replicability of psychological and neuroimaging findings (Barch & Yarkoni, 2013; Pashler & Wagenmakers, 2012) it is particularly noteworthy that coordinates of the activation foci reported here match closely those reported in three previous experiments by our group (Monti et al., 2007, 2009; see Supplementary Figure 2 for a visual comparison) despite different samples, stimuli, MR systems, and analysis methods. In particular, we find that complex deductive inferences, as compared to simpler ones, consistently recruit regions previously described as “core” to deductive reasoning (including left rostrolateral (in BA10) and medial (BA8) prefrontal cortices) which are proposed to be implicated in the construction of a derivational path between premises and conclusions. In addition, complex deductive inferences also recruited a number of (content-independent) “cognitive support” regions across frontal and parietal areas which were previously described as subserving non-deduction specific processes (cf., Monti et al., 2007, 2009). The dissociation between “core” and “support” regions is also supported by a neuropsychological study (Reverberi et al., 2009) in which patients with lesions extending to medial frontal cortex (encompassing “core” BA8) demonstrated specific deficits in deductive and meta-deductive ability, while patients with lesions to left dorsolateral frontal cortex (encompassing “support” regions) only displayed deductive deficits in the presence of working memory impairments (while retaining meta-deductive abilities; i.e., the ability to judge the complexity of an inference). This pattern of impairment can be seen as important independent and cross-methodological support for the view that medial frontal cortex plays a central role in representing the overall structure of sufficiently complex deductive arguments, while left dorsolateral frontal cortex (likely in conjunction with parietal regions) only serves a supporting role, perhaps as a kind of “memory space” for the representation of deductive arguments (Monti et al., 2007; Reverberi et al., 2009).

Our second finding relates to the alternative interpretation of the previously described “core” regions of deductive reasoning (left rostrolateral frontopolar cortex in particular; i.e., BA10) as reflecting a more general, non-deductive, increase in “cognitive difficulty” and/or working memory demands (Kroger et al., 2008; Prado, Mutreja, & Booth, 2013). The results of the present study, however, do not support this view. Indeed, in contrast to the results for deductive trials, when performed over the non-deductive trials, the load contrast failed to elicit any significant activity in the proposed “core” regions of deduction (left BA10 and medial BA8). Rather, it highlighted a different set of regions, including middle frontal gyrus (BA8, 6), precuneus (BA7), superior and inferior parietal lobule (BA7, 40), and angular gyrus (BA39), consistent with previous literature on working memory (Kirschen, Chen, Schraedley-Desmond, & Desmond, 2005; Koechlin, Basso, Pietrini, Panzer, & Grafman, 1999; Owen, McMillan, Laird, & Bullmore, 2005; Ranganath, Johnson, & D'Esposito, 2003; Ricciardi et al., 2006). This result further supports the view that the cognitive processes implemented within areas identified as core to deductive reasoning cannot be reduced to merely reflecting increased working memory demands, and might rather be integral to "identify[ing] and represent[ing] the overall structure of the proof necessary to solve a deductive problem" (Reverberi et al., 2009; p. 1113). Indeed, anterior prefrontal cortex has been previously associated with processes such as relational complexity (Halford et al., 1998, 2010; Robin & Holyoak, 1995; Waltz et al., 1999), and BA10 in particular has been found to activate in response to tasks requiring relational integration, keeping track of and integrating multiple related variables and sub-operations, and the handling of branching subtasks (Charron & Koechlin, 2010; Christoff et al., 2001; De Pisapia & Braver, 2008; Koechlin, Basso, Pietrini, Panzer, & Grafman, 1999b; Kroger et al., 2002; Ramnani & Owen, 2004), all of which are analogous to the kinds of processes that may be involved in evaluating deductive inferences. Similarly, previous research has associated medial BA8 with processes such as executive control, the choosing and coordinating of sub-goals, and resolving competition between rules to transform a problem from one state into another (Fletcher & Henson, 2001; Koechlin, Corrado, Pietrini, & Grafman, 2000; Posner & Dehaene, 1994; Volz, Schubotz, & Cramon, 2005). Of course, it should be noted that the current study, as well as our previous work, focuses on the propositional inferences (as defined in Garnham & Oakhill, 1994, p.60). Yet, previous work suggests that these results might well extend to propositional inferences in natural discourse (which, however might also depend on additional contributions from the right frontopolar cortex; Prado et al., 2015) and categorical syllogisms (Prado et al., 2013; Rodriguez-Moreno & Hirsch, 2009), further generalizing the idea that processes of relational integration might well explain the role of left rostrolateral PFC in deductive inference making (Prado et al., 2013).

Table 4.

Activations for the first interaction contrast, showing where the difference between complex memory and simple memory was greater than the difference between complex reasoning and simple reasoning

MNI Coordinates					
x	y	z	Hem	Region Label (BA)	Z-score
Frontal Lobe					
-24	0	52	L	Middle Frontal Gyrus (6)	4.01
-22	2	48	L	Middle Frontal Gyrus (6)	3.79
Parietal Lobe					
-36	-82	28	L	Angular Gurys/Superior Occipital Gyrus (19/39)	6.06
-10	-68	56	L	Superior Parietal Lobule (7)	5.82
-6	-62	54	L	Precuneus (7)	5.69
-18	-66	56	L	Superior Parietal Lobule (7)	5.29
-18	-70	52	L	Precuneus (7)	5.08
12	-66	54	R	Superior Parietal Lobule (7)	5.05

Finally, our results further add to a growing body of evidence failing to detect any involvement of Broca’s area in deductive inference-making across propositional, set-inclusion and relational problems (e.g., Canessa et al., 2005; Fangmeier et al., 2006; Knauff et al., 2002; Monti et al., 2007, 2009; Parsons & Osherson, 2001; Prado &

Noveck, 2007; Rodriguez-Moreno & Hirsch, 2009; Prado et al., 2015), and is consistent with neuropsychological work demonstrating that damage spanning at least one of the “core” regions (medial BA8) impairs deductive inference-making, as well as appreciation for the difficulty of an inference, despite the absence of any damage in the left IFG and ceiling performance in clinical tests of language (Reverberi et al., 2009).

In interpreting these findings, a number of important issues should be considered. First, although our behavioral results replicate the classic differential performance across modus ponens and modus tollens type inferences, accuracy in the modus tollens type inferences was higher than typically reported in classic studies (i.e., 87% in our work versus 63% in Taplin, 1971; 62% in Wildman & Fletcher, 1977). Nonetheless, our accuracy rates are in line with a number of behavioral and neuroimaging reports (e.g., 94% in Prado et al. 2010, 84% in Luo et al., 2011; between 80% and 88% in Blumfield & Rips 2003; 79% in Knauff et al., 2002; 78% in Trippas et al., 2017; 75% in Evans, 1977; and above 90% in the Wason Selection Task as implemented in Li et al., 2014, Qiu et al., 2007, Liu et al., 2012) as well as developmental work showing that by age 16 accuracy rates for modus tollens range between 78 and 87% (Daniel et al., 2006). We do stress, however, that although our participants did not undergo any overt training (e.g., training to criterion; see for example Reverberi et al., 2007) and reported no formal training in logic, our procedure selected high-performance individuals in the sense that they had to meet a 60% accuracy criterion across each of the four conditions (complex/simple, deductive/non-deductive). Thus, our approach also favored participants with good performance in working memory (as captured by the SWITCH task), which is known to be an important variable in deductive reasoning (e.g., Toms et al., 1993; Handley et al., 2002; Capon et al., 2003; Markovitz et al., 2002). In addition, it is also possible that self-selection operated, to some extent, among participants volunteering to take part in a study on "higher cognition." Second, as shown by our behavioral results, deductive and non-deductive tasks were not fully matched for difficulty (as captured by response time and accuracy). Indeed, we did find a significant difference between memory and logic difficulty (i.e., a significant interaction), with complex non-deductive trials being harder to evaluate than complex deductive trials, and with the difference between simple and complex non-deductive trials being larger than that for simple and complex deductive trials. This imbalance should make it all the more likely that, if the general cognitive load hypothesis were correct (i.e., that BA10 and BA8m are recruited by non-deductive cognitive load), the simple effect of load for non-deductive trials (i.e., $[\text{complex} > \text{simple}]_{ND}$) as well as the interaction effect of load and task type (i.e., $[\text{complex} - \text{simple}]_{ND} > [\text{complex} - \text{simple}]_{D}$) should uncover activity within the "core" areas. Yet, as we reported, this was not the case. Rather, it was the simple effect of load for deductive trials (as well as the interaction effect of load and task type (i.e., $[\text{complex} - \text{simple}]_{D} > [\text{complex} - \text{simple}]_{ND}$) that uncovered activity within left rostrolateral and mediofrontal cortex. In this sense, the lack of exact matching of difficulty across types of tasks works against our hypothesis and thus makes our test all the more stringent. Finally, our design contained a small asymmetry across deductive and non-deductive taks. Simple and complex deductive trials were signaled with the same cue (i.e., REASONING) whereas simple and complex non-deductive trials were signaled with different cues (i.e., MEMORY and SWITCH, respectively). Nonetheless, the close matching of the neuroimaging results presented above and previous work (cf., Wager & Smith, 2003; Monti et al., 2007, 2009) suggests that this might not have significantly affected our results.

In conclusion, this experiment shows that deductive inference making is based on a distributed network of regions including frontal and frontomedial “core” areas, the activation of which cannot be reduced to working memory demands or cognitive difficulty. Rather, these regions appear to be involved in processes that are at the heart of deductive inference (e.g., finding the derivational path uniting premises and conclusion; (cf., Monti & Osherson, 2012; Reverberi et al., 2009). Furthermore, our findings are also consistent with the idea that “[l]ogical reasoning goes beyond linguistic processing to the manipulation of non-linguistic representations” (Kroger et al., 2008, p. 99).

References

Baggio, G., Cherubini, P., Pischedda, D., Blumenthal, A., Haynes, J. D., & Reverberi, C. (2016). Multiple neural representations of elementary logical connectives. *NeuroImage*, *135*, 300–310. https://doi.org/10.1016/j.neuroimage.2016.04.061

Barch, D. M., & Yarkoni, T. (2013). Introduction to the special issue on reliability and replication in cognitive and affective neuroscience research. *Cognitive, Affective & Behavioral Neuroscience*, *13*(4), 687–9. https://doi.org/10.3758/s13415-013-0201-7

Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2003). General multilevel linear modeling for group analysis in FMRI. *Neuroimage*, 20(2), 1052-1063.

Bloomfield, A. N., & Rips, L. J. (2003). Content Effects in Conditional Reasoning: Evaluating the Container Schema. In Proceedings of the Cognitive Science Society (Vol. 25).

Braine, M. D. S., & O’Brien, D. P. (1998). *Mental logic*. Mahwah, N.J. : L. Erlbaum Associates.

Brett, M., Penny, W., & Kiebel, S. (2003). Introduction to Random Field Theory. In R. S. J. Frackowiak, K. J. Friston, C. D. Frith, R. J. Dolan, C. J. Price, S. Zeki, … W. D. Penny (Eds.), *Human Brain Function* (2nd ed., pp. 1–23). Academic Press-Elsevier. https://doi.org/10.1049/sqj.1969.0076

Buxton, R. B., Uludag, K., Dubowitz, D. J., & Liu, T. T. (2004). Modeling the hemodynamic response to brain activation. *Neuroimage*, *23*(Supplement 1), S220–S233. https://doi.org/DOI:10.1016/j.neuroimage.2004.07.013

Canessa, N., Gorini, A., Cappa, S. F., Piattelli-Palmarini, M., Danna, M., Fazio, F., & Perani, D. (2005). The effect of social content on deductive reasoning: An fMRI study. *Human Brain Mapping*, *26*(1), 30–43. https://doi.org/10.1002/hbm.20114

Capon, A., Handley, S., & Dennis, I. (2003). Working memory and reasoning: An individual differences perspective. Thinking & Reasoning, 9(3), 203–244.

Charron, S., & Koechlin, E. (2010). Divided representation of concurrent goals in the human frontal lobes. *Science*, *328*(5976), 360–363. https://doi.org/10.1126/science.1183614

Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J. K., Holyoak, K. J., & Gabrieli, J. D. E. (2001). Rostrolateral Prefrontal Cortex Involvement in Relational Integration during Reasoning. *NeuroImage*, *14*(5), 1136–1149. https://doi.org/10.1006/nimg.2001.0922

Crittenden, B. M., & Duncan, J. (2014). Task difficulty manipulation reveals multiple demand activity but no frontal lobe hierarchy. *Cerebral Cortex*, *24*(2), 532–540. https://doi.org/10.1093/cercor/bhs333

De Pisapia, N., & Braver, T. S. (2008). Preparation for integration: the role of anterior prefrontal cortex in working memory. *Neuroreport*, 2008, 19(1), 15–19. https://doi.org/10.1097/WNR.0b013e3282f31530

DeWolf, M., Chiang, J. N., Bassok, M., Holyoak, K. J., & Monti, M. M. (2016). Neural representations of magnitude for natural and rational numbers. *NeuroImage*, *141*, 304–312. https://doi.org/10.1016/j.neuroimage.2016.07.052

Fangmeier, T., Knauff, M., Ruff, C. C., & Sloutsky, V. (2006). FMRI evidence for a three-stage model of deductive reasoning. *Journal of Cognitive Neuroscience*, *18*(3), 320–34. https://doi.org/10.1162/089892906775990651

Fletcher, P. C. C., & Henson, R. N. (2001). Frontal lobes and human memory: insights from functional neuroimaging. *Brain*, *124*(5), 849–881. https://doi.org/10.1093/brain/124.5.849

Garnham, A., & Oakhill, J. (1994). *Thinking and reasoning*. Blackwell.

Good, R. H., Kaminski, R. A., Smith, S., Laimon, D., & Dill, S. (2004). *Dynamic indicators of basic early literacy skills*. Sopris West Educational Services.

Grinband, J., Wager, T. D., Lindquist, M., Ferrera, V. P., & Hirsch, J. (2008). Detection of time-varying signals in event-related fMRI designs. *NeuroImage*, *43*(3), 509–520. https://doi.org/10.1016/j.neuroimage.2008.07.065

Halford, G., Wilson, W., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences*, *21*(6), 803–865.

Halford, G., Wilson, W., & Phillips, S. (2010). Relational knowledge: the foundation of higher cognition. *Trends in Cognitive Sciences*, *14*(11), 497–505. https://doi.org/10.1016/J.TICS.2010.08.005

Handley, S. J., Capon, A., Copp, C., & Harper, C. (2002). Conditional reasoning and the Tower of Hanoi: The role of spatial and verbal working memory. *British Journal of Psychology*, 93(4), 501–518. https://doi.org/10.1348/000712602761381376

Heit, E. (2015). Brain Imaging, Forward Inference, and Theories of Reasoning. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.01056

Henson, R. (2006). Forward inference using functional neuroimaging: Dissociations versus associations. *Trends in Cognitive Sciences*, 10(2), 64–69. https://doi.org/10.1016/j.tics.2005.12.005

Henson, R. (2007). Efficient experimental design for fMRI. In K. J. Friston, J. Ashburner, S. J. Kiebel, T. E. Nichols, & W. D. Penny (Eds.), *Statistical Parametric Mapping: The Analysis of Functional Brain Images* (pp. 193–210). London: Academic Press.

Johnson-Laird, P. N. (1999). Deductive reasoning. *Annual Review of Psychology*, *50*(1), 109–135. https://doi.org/10.1146/annurev.psych.50.1.109

Kirschen, M. P., Chen, S. H. A., Schraedley-Desmond, P., & Desmond, J. E. (2005). Load- and practice-dependent increases in cerebro-cerebellar activation in verbal working memory: an fMRI study. *NeuroImage*, *24*(2), 462–72. https://doi.org/10.1016/j.neuroimage.2004.08.036

Knauff, M., Fangmeier, T., Ruff, C. C., & Johnson-Laird, P. N. (2003). Reasoning, models, and images: behavioral measures and cortical activity. *Journal of Cognitive Neuroscience*, *15*(4), 559–73. https://doi.org/10.1162/089892903321662949

Knauff, M., Mulack, T., Kassubek, J., Salih, H. R., & Greenlee, M. W. (2002). Spatial imagery in deductive reasoning: A functional MRI study. *Cognitive Brain Research*, *13*(2), 203–212. https://doi.org/10.1016/S0926-6410(01)00116-1

Koechlin, E., Basso, G., Pietrini, P., Panzer, S., & Grafman, J. (1999). The role of the anterior prefrontal cortex in human cognition, *399*(6732), 148–151. https://doi.org/10.1038/20178

Koechlin, E., Corrado, G., Pietrini, P., & Grafman, J. (2000). Dissociating the role of the medial and lateral anterior prefrontal cortex in human planning. *Proceedings of the National Academy of Sciences*, *97*(13), 7651–7656. https://doi.org/10.1073/pnas.130177397

Kroger, J. K., Nystrom, L. E., Cohen, J. D., & Johnson-Laird, P. N. (2008). Distinct neural substrates for deductive and mathematical processing. *Brain Research*, *1243*, 86–103. https://doi.org/10.1016/j.brainres.2008.07.128

Kroger, J. K., Sabb, F. W., Fales, C. L., Bookheimer, S. Y., Cohen, M. S., & Holyoak, K. J. (2002). Recruitment of Anterior Dorsolateral Prefrontal Cortex in Human Reasoning: a Parametric Study of Relational Complexity. *Cerebral Cortex*, *12*(5), 477–485. https://doi.org/10.1093/cercor/12.5.477

Li, B., Zhang, M., Luo, J., Qiu, J., & Liu, Y. (2014). The difference in spatiotemporal dynamics between modus ponens and modus tollens in the Wason selection task: An event-related potential study. Neuroscience, 270, 177–182. https://doi.org/10.1016/J.NEUROSCIENCE.2014.04.007

Liu, J., Zhang, M., Jou, J., Wu, X., Li, W., & Qiu, J. (2012). Neural bases of falsification in conditional proposition testing: Evidence from an fMRI study. International Journal of Psychophysiology, 85(2), 249–256. https://doi.org/10.1016/J.IJPSYCHO.2012.02.011

Luo, J., Yang, Q., Du, X., & Zhang, Q. (2011). Neural correlates of belief-laden reasoning during premise processing: an event-related potential study. Neuropsychobiology, 63(2), 112–8. https://doi.org/10.1159/000317846

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (Vol. 1). Psychology Press.

Markovits, H., Doyon, C., & Simoneau, M. (2002). Individual differences in working memory and conditional reasoning with concrete and abstract content. Thinking & Reasoning, 8(2), 97–107. https://doi.org/10.1080/13546780143000143

Monti, M. M., & Osherson, D. N. (2012). Logic, language and the brain. *Brain research, 1428*, 33-42. https://doi.org/10.1016/j.brainres.2011.05.061

Monti, M. M., Osherson, D. N., Martinez, M. J., & Parsons, L. M. (2007). Functional neuroanatomy of deductive inference: A language-independent distributed network. *NeuroImage*, *37*(3), 1005–1016. https://doi.org/10.1016/j.neuroimage.2007.04.069

Monti, M. M., Parsons, L. M., & Osherson, D. N. (2009). The boundaries of language and thought in deductive inference. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(30), 12554–9. https://doi.org/10.1073/pnas.0902422106

Morcom, A. M., & Fletcher, P. C. (2007). Does the brain have a baseline? Why we should be resisting a rest. *NeuroImage*, *37*(4), 1073–1082. https://doi.org/10.1016/j.neuroimage.2006.09.013

Noveck, I. A., Goel, V., & Smith, K. W. (2004). The Neural Basis of Conditional Reasoning with Arbitrary Content. *Cortex*, *40*(4), 613–622. https://doi.org/10.1016/S0010-9452(08)70157-6

Osherson, D. N. (1975). Logic and models of logical thinking. *Reasoning: Representation and Process in Children and Adults*, 81–91.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, *25*(1), 46–59. https://doi.org/10.1002/hbm.20131

Parsons, L. M., & Osherson, D. (2001). New Evidence for Distinct Right and Left Brain Systems for Deductive versus Probabilistic Reasoning. *Cerebral Cortex (New York, N.Y. : 1991)*, *11*(10), 954–65. https://doi.org/10.1093/cercor/11.10.954

Pashler, H., & Wagenmakers, E.-J. (2012). Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence? Perspectives on Psychological Science, 7(6), 528–530. https://doi.org/10.1177/1745691612465253Posner, M. I., & Dehaene, S. (1994). Attentional networks. *Trends in Neurosciences*, *17*(2), 75–79. https://doi.org/10.1016/0166-2236(94)90078-7

Prado, J., Chadha, A., & Booth, J. R. (2011). The Brain Network for Deductive Reasoning: A Quantitative Meta-analysis of 28 Neuroimaging Studies. *Journal of Cognitive Neuroscience*, *23*(11), 3483–3497. https://doi.org/10.1162/jocn_a_00063

Prado, J., Der Henst, J. B. Van, & Noveck, I. A. (2010). Recomposing a fragmented literature: How conditional and relational arguments engage different neural systems for deductive reasoning. *NeuroImage*, *51*(3), 1213–1221. https://doi.org/10.1016/j.neuroimage.2010.03.026

Prado, J., Mutreja, R., & Booth, J. R. (2013). Fractionating the Neural Substrates of Transitive Reasoning: Task-Dependent Contributions of Spatial and Verbal Representations. *Cerebral Cortex*, *23*(3), 499–507. https://doi.org/10.1093/cercor/bhr389

Prado, J., & Noveck, I. A. (2007). Overcoming Perceptual Features in Logical Reasoning: A Parametric Functional Magnetic Resonance Imaging Study. *Journal of Cognitive Neuroscience*, *19*(4), 642–657.

Prado, J., Spotorno, N., Koun, E., Hewitt, E., Van der Henst, J.-B., Sperber, D., & Noveck, I. A. (2015). Neural interaction between logical reasoning and pragmatic processing in narrative discourse. *Journal of Cognitive Neuroscience*, *27*(4), 692–704. https://doi.org/10.1162/jocn_a_00744

Qiu, J., Li, H., Huang, X., Zhang, F., Chen, A., Luo, Y., … Yuan, H. (2007). The neural basis of conditional reasoning: An event-related potential study.

Neuropsychologia, 45(7), 1533–1539. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2006.11.014

Toms, M., Morris, N., & Ward, D. (1993). Working memory and conditional reasoning. The Quarterly Journal of Experimental Psychology Section A, 46(4), 679–699. https://doi.org/10.1080/14640749308401033

Trippas, D., Thompson, V. A., & Handley, S. J. (2017). When fast logic meets slow belief: Evidence for a parallel-processing model of belief bias. Memory & Cognition, 45(4), 539–552. https://doi.org/10.3758/s13421-016-0680-1

Ramnani, N., & Owen, A. M. M. (2004). Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. *Nature Review Neuroscience*, *5*(3), 184–194. https://doi.org/10.1038/nrn1343

Ranganath, C., Johnson, M. K., & D’Esposito, M. (2003). Prefrontal activity associated with working memory and episodic long-term memory. *Neuropsychologia*, *41*(3), 378–389. https://doi.org/10.1016/S0028-3932(02)00169-0

Reverberi, C., Bonatti, L. L., Frackowiak, R. S. J., Paulesu, E., Cherubini, P., & Macaluso, E. (2012). Large scale brain activations predict reasoning profiles. *NeuroImage*, *59*(2), 1752–1764. https://doi.org/10.1016/j.neuroimage.2011.08.027

Reverberi, C., Cherubini, P., Frackowiak, R. S. J., Caltagirone, C., Paulesu, E., & Macaluso, E. (2010). Conditional and syllogistic deductive tasks dissociate functionally during premise integration. *Human Brain Mapping*, *31*(9), 1430–1445. https://doi.org/10.1002/hbm.20947

Reverberi, C., Cherubini, P., Rapisarda, A., Rigamonti, E., Caltagirone, C., Frackowiak, R. S. J., … Paulesu, E. (2007). Neural basis of generation of conclusions in elementary deduction. *NeuroImage*, *38*(4), 752–762. https://doi.org/10.1016/j.neuroimage.2007.07.060

Reverberi, C., Shallice, T., D’Agostini, S., Skrap, M., & Bonatti, L. L. (2009). Cortical bases of elementary deductive reasoning: Inference, memory, and metadeduction. *Neuropsychologia*, *47*(4), 1107–1116. https://doi.org/10.1016/j.neuropsychologia.2009.01.004

Ricciardi, E., Bonino, D., Gentili, C., Sani, L., Pietrini, P., & Vecchi, T. (2006). Neural correlates of spatial working memory in humans: A functional magnetic resonance imaging study comparing visual and tactile processes. *Neuroscience*, *139*(1), 339–349. https://doi.org/10.1016/j.neuroscience.2005.08.045

Rips, L. J. (1994). *The psychology of proof : deductive reasoning in human thinking*. MIT Press.

Robin, N., & Holyoak, K. J. (1995). Relational complexity and the functions of prefrontal cortex. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 987-997). Cambridge, MA: The MIT Press.

Rodriguez-Moreno, D., & Hirsch, J. (2009). The dynamics of deductive reasoning: An fMRI investigation. *Neuropsychologia*, *47*(4), 949–961. https://doi.org/10.1016/j.neuropsychologia.2008.08.030

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., … Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, *23*, *Supple*, S208–S219. https://doi.org/10.1016/j.neuroimage.2004.07.051

Strand, F., Forssberg, H., Klingberg, T., & Norrelgen, F. (2008). Phonological working memory with auditory presentation of pseudo-words--An event related fMRI study. *Brain Research*, *1212*, 48–54. https://doi.org/10.1016/j.brainres.2008.02.097

Volz, K. G., Schubotz, R. I., & Cramon, D. Y. von. (2005). Variants of uncertainty in decision-making and their neural correlates. *Brain Research Bulletin*, *67*(5), 403–412. https://doi.org/10.1016/j.brainresbull.2005.06.011

Wager, T. D., & Smith, E. E. (2003). Neuroimaging studies of working memory: *Cognitive, Affective, & Behavioral Neuroscience*, *3*(4), 255–274. https://doi.org/10.3758/CABN.3.4.255

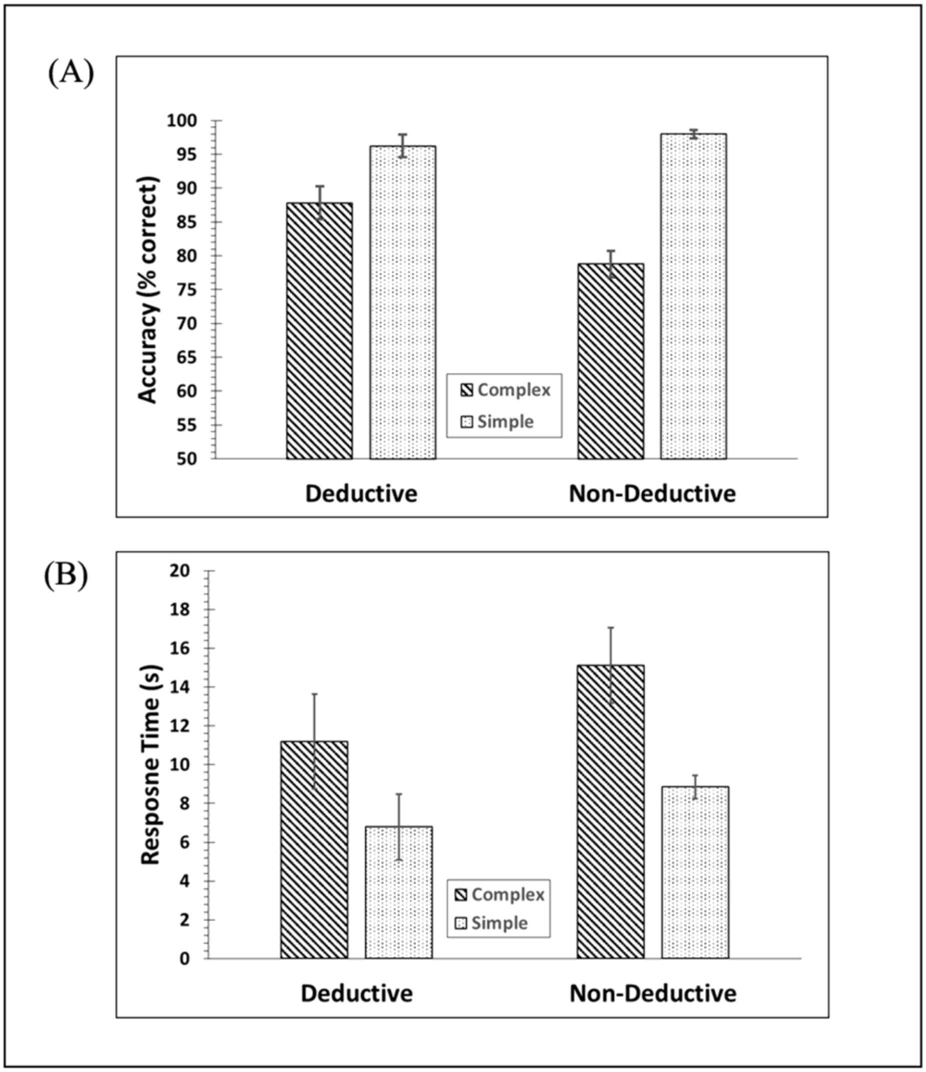
Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., Santos, M. de M., … Miller, B. L. (1999). A System for Relational Reasoning in Human Prefrontal Cortex. *Psychological Science*, *10*(2), 119–125. https://doi.org/10.1111/1467-9280.00118

Woolrich, M. W., Behrens, T. E., Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2004). Multilevel linear modelling for FMRI group analysis using Bayesian inference. *Neuroimage*, 21(4), 1732-1747.

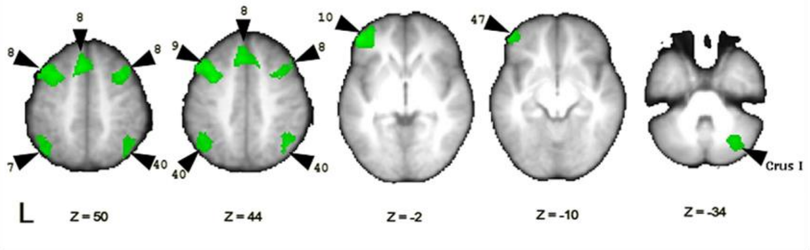
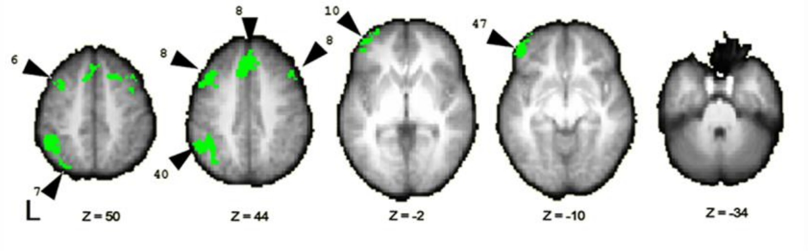
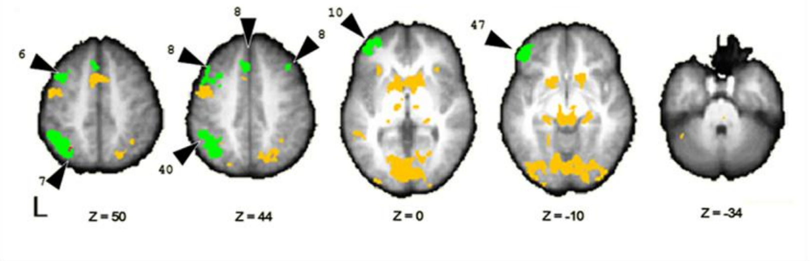
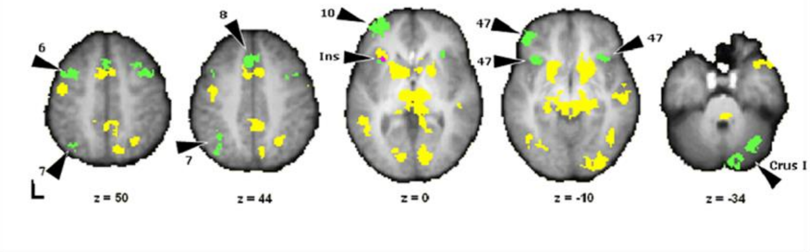
Worsley, K. J. (2001). Statistical analysis of activation images. In P. Jezzard, P. M. Matthews, & S. M. Smith (Eds.), *Functional MRI: an introduction to methods* (pp. 251–270). Oxford: Oxford University Press.

Worsley, K. J. (2007). Random Field Theory. In K. J. Friston, J. Ashburner, S. Kiebel, T. E. Nichols, & W. D. Penny (Eds.), *Statistical Parametric Mapping* (pp. 232–236). Academic Press-Elsevier. https://doi.org/http://dx.doi.org/10.1016/B978-012372560-8/50018-8

APPENDIX: SUPPLEMENTS



Supplement 1. Behavioral results. (A) Behavioral results for accuracy, across task types and levels of load. There was a significant interaction between load and task type (*F*(1, 19) = 9.03, *p* = 0.007, $\hat{\omega}_p^2$= 0.05). (B) Behavioral results for response time, across task types and levels of load. There was a significant interaction between load and task type (*F*(1, 19) = 5.97, *p* = 0.02, $\hat{\omega}_p^2$ = 0.01).

Contrast: Complex Logic Inference vs. Simple Logic Inference **Contents:** Features of imaginary shapes and pseudo-words **N = 20** (Coetzee & Monti, 2017)	
Contrast: Logic Inference vs. Logic Grammar **Contents:** Single letters (i.e. X,Y,Z) **N = 15** (Monti et al., 2009)	
Contrast: Complex Logic Inference vs. Simple Logic Inference **Contents:** Features of imaginary faces and houses **N = 12** (Monti et al., 2007)	
Contrast: Complex Logic Inference vs. Simple Logic Inference **Contents:** Features of imaginary shapes and pseudo-words **N = 10** (Monti et al., 2007)	

Supplement 2. Comparison with prior studies. Areas activated by deductive reasoning are displayed in green. From bottom to top, the first two studies are described in Monti et al. (2007), while the third is described in Monti et al. (2009), and the top row presents partial results from the current study.

Study 2: Dissociating Language and Thought in Human Reasoning

Introduction

Does language shape human cognition (Boeckx, 2010; Monti, 2017)? One aspect of this debate concerns whether the combinatorial operations of natural language also serve, supramodally (Fadiga, Craighero, & D'Ausilio, 2009; Fitch & Martins, 2014; Tettamanti & Weniger, 2006), as a basis for the combinatorial operations of other domains of human thought such as mathematics (Amalric & Dehaene, 2016; Monti, Parsons, & Osherson, 2012; R. A. Varley, Klessinger, Romanowski, & Siegal, 2005), music cognition (Chiang et al., 2017; Maess, Koelsch, Gunter, & Friederici, 2001; Patel, 2003), and theory of mind (R. Varley, 2001; Rosemary Varley & Siegal, 2000), among others. In the context of human reasoning, a growing body of research has generated two opposing views concerning the role of language in enabling deductive inference-making (Monti & Osherson, 2012). Under the first, the structure-dependent operations required for deductive reasoning are linguistic in nature and thus based on the neural mechanisms of natural language (Reverberi et al., 2007, 2010). Under the second, deductive reasoning is supported by language-independent processes principally implemented in frontomedial and left frontopolar cortices (in Brodmann areas [BA] 8 and 10, respectively) (Coetzee & Monti, 2018; Monti, Osherson, Martinez, & Parsons, 2007; Monti, Parsons, & Osherson, 2009; Rodriguez-Moreno & Hirsch, 2009). We tested the two competing views using continuous Theta Burst Stimulation (cTBS (Huang, Edwards, Rounis, Bhatia, & Rothwell, 2005)), an inhibitory neuromodulation approach. Here, we report a neural dissociation between the structure-dependent operations of natural language and those of deductive reasoning. Specifically, we found that transient cTBS inhibition of Broca's area, in the left inferior frontal gyrus (IFG), a region often associated with processing the hierarchical dependencies of natural language (Friederici, 2004, 2016), impairs accuracy for linguistic problems but not for logic ones. Conversely, cTBS to frontomedial cortices produces the opposite pattern. Our results demonstrate that, in the adult brain, the structure-dependent operations of reasoning are not parasitic to the neural substrate of natural language.

Continuous Theta Burst Stimulation (cTBS) has been shown capable of transiently inhibiting neural tissue, thereby modulating cognition and behavior (Christov-Moore, Sugiyama, Grigaityte, & Iacoboni, 2017; Huang et al., 2005). Here, we employ this technique to distinguish between the two opposing views of human reasoning. Specifically, if the mental operations of deductive reasoning depend on the mechanisms of natural language, as posited by the "language-centric" view of deduction (Reverberi et al., 2007, 2010), transient cTBS inhibition of the region traditionally referred to as Broca's area, in the left inferior frontal gyrus, ought to decrease performance (i.e., accuracy) for both linguistic and deductive reasoning to a comparable extent. Conversely, if the two mental domains do not share a common substrate in the structure-dependent mechanisms of natural language, as posited by the language-independent view of deduction (Coetzee & Monti, 2018; Monti et al., 2007, 2009; Rodriguez-Moreno & Hirsch, 2009), cTBS to Broca's area ought to impair linguistic reasoning while sparing logic reasoning. Furthermore, the language-independent view also predicts that cTBS to frontomedial cortex, in mesial BA8, a region previously described as "core" for deductive inference (Monti et al., 2009; Rodriguez-Moreno & Hirsch, 2009), should impair logic reasoning while sparing linguistic reasoning.

Methods

Participants

Fifteen participants took part in this study (twelve women, three men). The mean age was 21.1 and the age range was 18-30. Participants were recruited through flyers and from other (unrelated) studies. To be included, participants had to be right handed, native English speakers, between the ages of 18 and 50 years old, and have had no prior formal instruction in deductive reasoning. In addition, we only selected participants who had a recent structural MRI available (from previous participation in a neuroimaging experiment at UCLA) to allow for MR-guided targeting with the transcranial magnetic stimulation (TMS) coil on the basis of individual brain anatomy (see below). In keeping with TMS safety standards (Rossi, Hallett, Rossini, & Pascual-Leone, 2009), participants were excluded if they had metal implants in their head, if they engaged in regular alcohol use, were pregnant, had a family history of seizures, had been diagnosed with any significant medical, psychiatric or neurological conditions, or used any prescription medication that could lower their seizure threshold (i.e. bupropion). Participants were compensated $25 per hour for their time. Total compensation for each completing participant ranged from $125 to $175.

Each participant attended four study visits. The first was a screening visit, which took place in the UCLA Psychology Department, at which the participant was consented and, after viewing one example trial for each task, performed a set of problems analogous to those employed in the subsequent cTBS sessions (except for superficial differences in the stimuli). Participants never received any feedback on either individual problems or overall performance. To be included in the TMS sessions of the study participants had to perform at or above 50% accuracy on the overall task and each of the three primary subcomponents (i.e. linguistic problems, logic problems, and grammaticality judgments, described below). Seven participants were excluded for being unable to meet this criterion (five men and two women).

The three TMS sessions took place at the UCLA Ahmanson-Lovelace Brain Mapping Center. Visits took place at least one week apart. In each TMS session, one of three sites was targeted; namely, Broca's area, in the *pars opercularis* of the left inferior frontal gyrus (Brodmann area [BA] 44), mesial BA8, or left transverse occipital sulcus (LTOS) (see below for procedure and coordinates). The order in which target sites were stimulated was counterbalanced across participants. At each visit, participants first performed a ten minute baseline cognitive task. They then underwent the TMS procedure, which included a thresholding procedure followed by the administration of cTBS. Approximately 2 to 3 minutes after the TMS procedure ended they performed a 30 minute post-cTBS task. All participants who began the experimental phase of the experiment completed the experiment. All procedures were approved by the UCLA Institutional Review Board.

Task and Stimuli

Task and stimuli materials were adapted from previous work (Monti et al., 2009). For each of the three TMS sessions, participants were presented with 156 stimuli, in visual format. Each stimulus consisted of an argument, defined as a set of two sentences (one presented above a horizontal line and one below). Half the arguments were "linguistic" in that they described a subject-object-patient relationship (i.e., "who did what to whom"; e.g., "Y gave X to Z." and "X was given Z by Y."). The remaining were “logic” in that they described the logic implicature tying phrasal constituents together (i.e., “X,Y, Z”; e.g. “If Y or X then not Z.” and “If Z then not Y and not X.”).

For each argument, participants were asked to perform one of two tasks. In the reasoning task, they were asked to establish whether the two sentences of each argument matched in that they described the same state of affairs (that is, they had to decide whether the two sentences were transformations of one another). Half the arguments presented in the reasoning trials described the same state of affairs and half did not. In the grammaticality judgment task, participants were asked to evaluate whether both sentences of each argument were grammatical (with no need to relate the two sentences to each other). Half the arguments presented in the grammaticality trials were grammatical and half were not. As done in previous work, ungrammatical arguments were obtained by altering word order in either sentence (Monti et al., 2009, 2012). Half the ungrammatical sentences had an error in the sentence above the line, and half had the error in the sentence below the line. Overall, the 156 arguments that participants saw at each session included 104 reasoning trials (half with "linguistic" arguments and half with "logic arguments") and 52 grammaticality judgment trials (also evenly divided between types of arguments).

It should be noted that, in the context of the reasoning task, linguistic and logic arguments emphasize different types of structure-dependent relationships. When presented with linguistic arguments, the reasoning task required understanding the thematic relations of “X,Y, Z” with respect to the major verb of the sentence, across different syntactic constructs (e.g. X is a patient in “It was X that Y saw Z take.” but is an agent in “Z was seen by X taking Y.”). When presented with logic arguments, the reasoning task required understanding the logic relations tying phrasal constituents together across different statements (e.g. “If both X and Z then not Y.” and “If Y then either not X or not Z.”).

In order to manipulate the relational complexity (Halford, Wilson, & Phillips, 2010) of the arguments, for each type of problem, half the arguments contained three variables and half contained four variables. We also note that, for each task type, half the trials included statements concerning the relationships between three variables (e.g., “X was given Y by Z.” and “If either X or Y then not Z.”) and the remainder included statements concerning the relationship between 4 variables (e.g., “W heard that Z was seen by Y taking X.” and “If either Z or W then both X and not Y.”). For each type of problem, half the arguments featured sentences describing the same state of affairs (i.e., where the two sentences match in the circumstance they describe). Assignment of the variables W, X, Y, Z to elements/phrasal constituents was randomized across arguments.

In each session, the 156 arguments included 78 linguistic arguments and 78 logic arguments. For each type, 52 arguments were presented in reasoning trials, and 26 were presented in grammaticality judgment trials. Of the 156 trials, 36 (equally distributed across tasks) were presented prior to cTBS stimulation (i.e., baseline trials) and 120 (equally distributed across tasks) were presented after cTBS stimulation. The same 156 arguments were presented across the four sessions except for randomly allocating each argument to baseline or post-cTBS presentation and for different allocation of variables (i.e., W, X, Y, Z) to thematic roles/phrasal constituents. Within baseline and post-cTBS sequences, presentation order of each argument (and task) was randomized with the sole constraint that trials with identical parameters not occur consecutively.

Experimental Design

As shown in Fig. 1a, each trial began with a 1 second fixation cross followed by a 1 second cue signaling to the participant whether they were to perform a reasoning task (with either linguistic or logic materials), cued by the word “MEANING”, or the grammaticality judgment task (with either linguistic or logic materials), cued by the word “GRAMMAR”. The cue was followed by on screen presentation of the argument, with the two sentences arranged vertically, one above the other, separated by a horizontal line (cf., Fig. 1a). Given the randomized task order, a small “M” or “G” block letter at the top left of the screen served as a reminder of which tasks participants were expected to perform at each trial (as we have done in previous work (Coetzee & Monti, 2018)). Participants had up to a maximum of 15 seconds to press the A key for a positive answer (i.e., “the sentences describe the same state of affairs” and “both sentences are grammatical”, for the reasoning and grammaticality judgment task, respectively) and the L key for a negative answer (i.e., “the sentences do not describe the same state of affairs” and “one of the two sentences is grammatically incorrect”, for the reasoning and grammaticality judgment task, respectively). The trial terminated upon button-press or upon the elapsing of the allotted 15 s, after which a new trial would begin. Stimuli were delivered using Psychopy (Peirce, 2008) on a Toshiba Satellite laptop running Windows 7.

Transcranial Magnetic Stimulation

For each participant, the FMIRB Software Library (FSL) (Smith et al., 2004) was used to transform the individual T1-weighted structural MRI – which had been obtained, with consent, from previous studies they had taken part in – into standard space (MNI template space). Stimulation targets were defined on the basis of previous published work. These included a target in Broca's area (x = -50, y = 18, z = 18) (Monti et al., 2009), centered on the *pars opercularis* of the left inferior frontal gyrus, one in mesial BA8 (x = -6, y = 40, z = 38) (Coetzee & Monti, 2018; Monti et al., 2009), and a "hypothesis neutral" target in the LTOS (x = -25, y = -85, z = 25) (Iaria & Petrides, 2007). Two additional targets were used for the active motor thresholding (AMT) procedure. Coordinates for cortical stimulation of these two sites, the first dorsal interosseous (FDI) muscle in the right hand, and the tibialis anterior (TA) muscle of the right leg, were also marked in standard space based on prior literature (Mayka, Corcos, Leurgans, & Vaillancourt, 2006; Niskanen et al., 2010; Sarfeld et al., 2012). The targets, originally defined in MNI template space, were then projected back into the participant's native structural MRI space, to allow optimal TMS coil positioning for each target through the frameless stereotaxy Brainsight system (Rogue Research).

For TMS stimulation of the motor cortex representation of the right FDI muscle, Broca's area, and LTOS, a Magstim flat figure-eight (double 70 mm) coil was used. Because our mesial BA8 target and the motor cortex representation of the right TA muscle are located within the interhemispheric fissure, we used an angled figure-eight (110 mm double cone) coil to deliver pulses to these areas. This method is similar to that used in previous studies (Christov-Moore et al., 2017; Holbrook, Izuma, Deblieck, Fessler, & Iacoboni, 2016; Klucharev, Munneke, Smidts, & Fernandez, 2011).

After participants completed the baseline task, the AMT was measured for that session's target site using a two-step procedure (Christov-Moore et al., 2017; Deblieck, Thompson, Iacoboni, & Wu, 2008). For Broca's area and LTOS target sessions, "hot spot" coordinates were based on left motor cortex representation of the right FDI (flat figure-8 coil); for mesial BA8 target sessions, "hot spot" coordinates were based on left motor cortex representation of the right TA (double cone coil). Single TMS pulses were delivered while the target muscle was mildly activated. If single pulses from the coil did not produce motor evoked potentials (MEPs) of ≥200 μV at initial location, then the coil location was varied systematically around the initial target site until reliable MEP 's were evoked at a suprathreshold intensity. Once the motor cortex "hot spot" was determined, the AMT was determined as the minimum TMS intensity at which motor evoked potentials (MEPs) of ≥200 μV were obtained in at least five out of ten consecutive stimulations under active target muscle conditions.

Following the thresholding procedure, cTBS was applied to the target. In cTBS, triplets of TMS pulses at 50 Hz are delivered at 5 Hz, giving a total of 600 pulses over a period of 40 seconds. The intensity was set at 80% of the AMT, in accordance with prior studies (Christov-Moore et al., 2017; Fitzgerald, Fountain, & Daskalakis, 2006). For 12 out of 44 sessions (5 at which Broca's area was targeted, 1 at which mesial BA8 was targeted, and 6 at which TOS was targeted) the participant's AMT was too high for our TMS device to deliver cTBS without significant heating. For these sessions, instead of using 80% of AMT, we applied cTBS at the highest level allowed by the safety measures of our TMS device (43% of maximum stimulator output (MSO)). The cTBS pulse pattern was generated using a second generation Magstim Rapid2, and the average percentage of MSO used was 35.61% (with a range of 19%-43%).

Upon completion of the cTBS stimulation procedure, participants began the post-treatment task after a delay of approximately 2-3 minutes. The post cTBS portion of the experiment lasted for 30 minutes (the inhibitory effects of cTBS have previously been shown to last for 30 to 60 minutes) (Huang et al., 2005; Oberman, Edwards, Eldaief, & Pascual-Leone, 2011). Upon completion of all trials, participants filled out a brief questionnaire to assess how much pain and/or discomfort they experienced during the cTBS stimulation. Both the pain and discomfort scales asked the participant to rate, from 0 to 10, how much pain or discomfort they were in during the procedure, with 0 indicating no pain/discomfort and 10 indicating the worst pain/discomfort they had ever felt. Across all participants, the mean pain rating was 2.52 (*SD* = 1.76), while the mean discomfort rating was 3.25 (*SD* = 1.88). For each stimulation site, the mean pain ratings were as follows: 2.64 (*SD* = 1.67) for BA44, 3.33 (*SD* = 2.09) for BA8, and 1.60 (*SD* = 0.80) for TOS. For discomfort ratings at each stimulation site, the means were: 3.64 (*SD* = 2.12) for BA44, 3.87 (*SD* = 1.71) for BA8, and 2.27 (*SD* = 1.34) for TOS. It is worth noting that no participants who began the TMS component of the study failed to complete it.

Analysis

First, in order to remove accidental key presses from the results, all trials from all participants were ordered from fastest response time to slowest response time. Then, trials were binned into groups of ten, with accuracy and response time averaged within each bin to see if there was any response time threshold below which accuracy fell below 50%. There was no such threshold, but four individual trials had response times of less than one second which were deemed likely to have been accidental button presses and were thus removed from further consideration. Average accuracies for each combination of task and site for each participant were entered in the two following analyses.

Key analysis

The specific predictions of the language-centric and language-independent views of deductive reasoning described in the main text were tested in a 2 × 2 ANOVA with two within-participants factors, site (Broca's area vs mesial BA8) and task (linguistic reasoning vs logic reasoning). The analysis was followed-up with planned directional testing of the simple effect of site on each task individually with pairwise t-tests.

Full analysis

To report on the full set of sites and conditions tested, a 3 × 3 ANOVA with two within-participants factors, site (Broca's area vs mesial BA8 vs LTOS) and task (linguistic reasoning vs logic reasoning vs grammaticality judgments), was also performed. The analysis was followed up through testing of the simple effect of site on each task individually with a contrast analysis. Specifically, for each task, we created two contrasts conceived to identify the presence of systematic associations between post-cTBS accuracy percent change and site. Specifically, we identified two possible trends (of interest). The "linguistic trend" (T_{Lin}) contrast was specified in order to mark, where significant, tasks more sensitive to disruption of Broca's area than either mesial

BA8 or LTOS. T_{Lin} was thus obtained by setting contrast weights to -1 for Broca's area and 0.5 for mesial BA8 and LTOS. The "logic trend" (T_{Log}) contrast was specified in order to mark, where significant, tasks more sensitive to disruption of mesial BA8 than Broca's area and LTOS. T_{Log} was thus obtained by setting contrast weights to -1 for mesial BA8 and 0.5 for Broca's area and LTOS.As noted in the main text, a third manipulation was included in the experimental design whereby half the arguments concerned the relationships between 3 variables (i.e., thematic roles/phrasal constituents) and the remaining concerned 4 variables (evenly allocated across tasks). However, since no main effect of number of relations was found ($F_{1,14} = 1.301$, $p = 0.273$, $\eta p^2 = 0.085$), nor any two-way ($F_{2,28} = 2.215$, $p = 0.13$, $\eta_p^2 = 0.137$; and $F_{2,28} = 3.02$, $p = 0.062$, $\eta p^2 = 0.18$ for the task × level and site × level interactions, respectively) or three-way ($F_{4,56} = 0.651$, $p = 0.628$, $\eta_p^2 = 0.044$) interaction with task and site, for the purposes of the above analyses data were aggregated across this manipulation. (2,566 words)

Results and Discussion

As shown in Fig 1a,b, we tested the above predictions with an experimental design including two sites of interest (Broca's area, mesial BA8), as well as a "hypothesis neutral" site (left transversal occipital sulcus; LTOS), and two tasks of interest (linguistic reasoning, logic reasoning), as well as a third "control" task (all adapted from previous work; see Methods section(Monti et al., 2009)). We also note that, for each task, half the trials included statements concerning the relationships between three variables (e.g., "X was given Y by Z." and "If either X or Y then not Z.")

and the remainder included statements concerning the relationship between 4 variables (e.g., “W heard that Z was seen by Y taking X.” and “If either Z or W then both X and not Y.”) (see Supplementary Table 1 for examples of stimuli). However, since we found no significant main effect of number of variables, no significant 2-way interaction with either

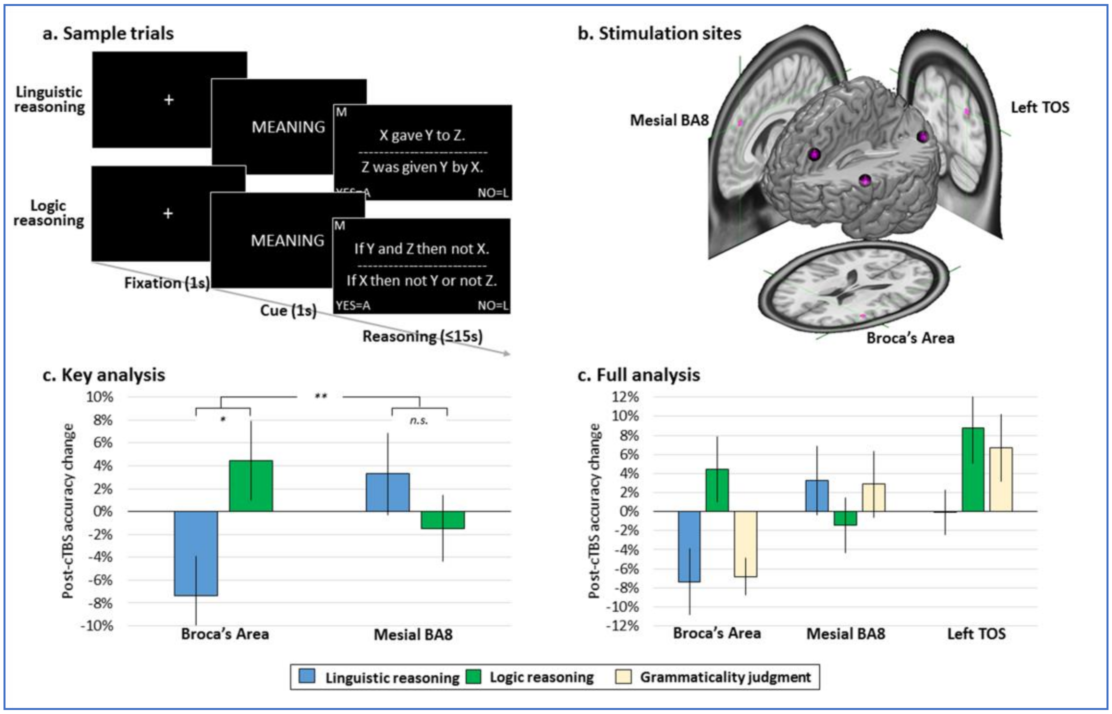


Fig. 1. Experimental design and results: (a) Timeline and sample trials for a linguistic reasoning and a logic reasoning trial (see Methods for sample grammaticality judgment trial). (b) cTBS target sites: Broca’s area (aimed at the *pars opercularis* of the inferior frontal gyrus; MNI coordinates: x = -50, y = 18, z = 18(Monti et al., 2009)), mesial BA8 (MNI coord: x = -6, y = 40, z = 38(Coetzee & Monti, 2018; Monti et al., 2009)), and LTOS (MNI coord: x = -25, y = -85, z = 25(Iaria & Petrides 2007)). (c) Key analysis result: Percent accuracy change for linguistic (blue) and logic (green) reasoning after cTBS to Broca’s area (left) and mesial BA8 (right). (d) Full analysis result: Percent accuracy change for linguistic (blue) and logic (green) reasoning, and grammaticality judgments (yellow) after cTBS to Broca’s area (left), mesial BA8 (middle), and LTOS (right). (Error bars indicate standard error; “**” indicates p < 0.005; “*” indicates p < 0.05; “n.s.” indicates non-significant effect; see text for details.)

site or task, and no significant 3-way interaction (all *p*s > 0.05; see discussion), we omitted this variable from the report below. Overall, collapsing across stimulation sites and across pre- and post-cTBS trials, we find accuracy for linguistic reasoning and grammaticality judgments to be higher than for logic reasoning (83%, 84%, and 73%, respectively; see Table 1).

In what follows, we first present the “key analysis” directly assessing the specific *ex ante* hypotheses outlined above (in a 2 × 2 within-subjects ANOVA framework) and we then expand the analysis to the hypothesis neutral site and control task (in a 3 × 3 framework; henceforth, “full analysis”).

A 2 × 2 within-subjects ANOVA over participants’ post-cTBS accuracy percent change relative to pre-cTBS baseline accuracy revealed a significant cross-over interaction ($F_{1,14} = 9.67$, $p = 0.008$, $\eta_p^2 = 0.41$) between stimulation site (Broca’s area versus mesial BA8) and task (linguistic reasoning vs. logic reasoning) (Fig. 1c). No main effect of stimulation site ($F_{1,14} = 0.85$, $p = 0.37$, $\eta_p^2 = 0.06$) or task ($F_{1,14} = 0.74$, $p = 0.40$, $\eta_p^2 = 0.05$) was observed. Planned directional testing of the *ex-ante* hypotheses described above revealed that transient cTBS inhibition of Broca’s area resulted in significantly different patterns of accuracy percent change across linguistic and logic reasoning ($t_{14} = -2.40$, $p = 0.015$). Specifically, as shown in Fig 1c, transient inhibition to Broca’s area decreased accuracy for linguistic problems by 7.4%, relative to pre-cTBS baseline accuracy, while sparing logic reasoning, for which accuracy increased by 4.4% relative to pre-cTBS baseline. This pattern is contrary to the prediction of the language-centric view and consistent with the prediction of the language-independent view. Conversely, transient inhibition of mesial BA8 resulted in the opposite pattern (Fig. 1b), with post-cTBS accuracy for logic reasoning decreasing by 1.5% and post-cTBS accuracy for linguistic reasoning increasing by 3.3%, compared to pre-cTBS baseline. Nonetheless, while this pattern is in line with the language-independent view, the difference was not statistically significant ($t_{14} = -0.99$, $p = 0.17$; see discussion below).

	Stimulation site					
	BROCA'S AREA		MESIAL BA8		LTOS	
	PRE	POST	PRE	POST	PRE	POST
LINGUISTIC REASONING	91%	83%	78%	81%	80%	80%
LOGIC REASONING	70%	75%	75%	73%	67%	76%
GRAMMATICALITY JUDGMENT	89%	82%	82%	84%	77%	84%

Table 1. Percent accuracy for each task before (PRE) and after (POST) transient inhibitory stimulation to each site.

Inclusion of the third, hypothesis neutral, site as well as the grammaticality judgment task (Monti et al., 2009, 2012) in a 3 × 3 within-subjects ANOVA confirmed the significant interaction of task and site ($F_{4,56} = 2.73$, $p = 0.038$, $\eta_p^2 = 0.16$) reported in the key analysis and uncovered a significant main effect of stimulation site ($F_{2,28} = 5.17$, $p = 0.012$, $\eta_p^2 = 0.27$), driven by the inclusion of the LTOS target (cf., Fig. 1d). Finally, in order to further test the relationship between site and task, we performed a trend analysis (see Supplementary Online Table 2 for details). If indeed Broca's area is specific to linguistic processes, we expect cTBS to this region to decrease accuracies for linguistic reasoning (but not for logic reasoning) more so than cTBS to either mesial

BA8 or LTOS (henceforth “linguistic trend” T_{Lin}; see Methods for detailed description and contrast weight settings). Conversely, if indeed mesial BA8 is specific to logic processes, we expect cTBS to this region to decrease accuracies for logic reasoning (but not for linguistic reasoning) more so than cTBS to either Broca’s area or LTOS (henceforth “logic trend”, T_{Log}). Consistent with the result described above, for linguistic reasoning T_{Lin} was significant ($F_{1,14} = 7.70$, $p = 0.015$) whereas T_{Log} was not ($F_{1,14} = 3.96$, $p = 0.066$; in fact, the marginal significance is due to a reverse pattern, with performance on linguistic problems after cTBS to mesial BA8 increasing by 3.3%, see Fig 1d). Conversely, for logic reasoning, T_{Log} was significant ($F_{1,14} = 6.626$, $p = 0.022$) whereas T_{Lin} was not ($F_{1,14} = 0.038$, $p = 0.849$). In addition, we also find that accuracy for the grammaticality judgments were affected by cTBS in a pattern similar to that observed for linguistic problems (and thus opposite to the pattern observed for logic problems; Fig.1d). Specifically, inhibition of Broca’s area leads to a decrease in accuracy, by 6.8%, compared to pre-cTBS baseline, whereas inhibition of mesial BA8 and LTOS both lead to increased accuracy (by 2.9% and 6.6%, respectively). The trend analysis thus returned a similar pattern to that obtained for linguistic problems (i.e., significant for T_{Lin} [$F_{1,14} = 11.221$, $p = 0.005$] and non-significant for T_{Log} [$F_{1,14} = 0.576$, $p = 0.460$]).

Overall, these data do not support the hypothesis that, in the adult brain, the structure-dependent operations of logic are parasitic on the mechanisms of language. For, it is possible to selectively impair the latter, by transient inhibition of Broca’s area, without affecting the former. This result is consistent with neuropsychological evidence demonstrating that patients with lesions spanning frontomedial cortices (including our cTBS site in mesial BA8) are impaired at deductive reasoning despite no observable structural damage in Broca's area and ceiling performance on standard neuropsychological tests of language(Reverberi, Shallice, D'Agostini, Skrap, & Bonatti, 2009). Although we tested the language-centric hypothesis of deduction in the context of a specific mode of deductive reasoning (i.e., propositional logic), previous work suggests that this conclusion might well extend to categorical syllogisms (Prado, Mutreja, & Booth, 2013; Rodriguez-Moreno & Hirsch, 2009; Tsujii, Sakatani, Masuda, Akiyama, & Watanabe, 2011) and relational problems (Knauff, Fangmeier, Ruff, & Johnson-Laird, 2003). The present data are still compatible with the Vygotskyan idea that language may serve, throughout development, as a "cognitive scaffolding"(Carruthers, 2002) enabling the acquisition of structure-dependent operations such as those of logic, to then become independent, in adulthood. Yet, recent evidence suggests that preverbal infants can already demonstrate elementary logic reasoning (Cesana-arlotti, Martín, & Téglás, 2018). Indeed, it is noteworthy that in our adult participants, logic reasoning appears unaffected by inhibitory stimulation to Broca's area despite decreased accuracy in both linguistic reasoning and simple grammaticality judgments, further supporting the idea that there is a fundamental difference between the representations and operations of logic and those of natural language.

While these results provide clear empirical evidence against the idea that the mechanisms of natural language participate in logic reasoning (beyond decoding verbally presented information into mental representations (Monti et al., 2007, 2009)), they only offer weak support for the hypothesis that medial prefrontal cortex includes the “core” substrate of deductive reasoning. For, while we numerically observe the pattern predicted by previous neuroimaging results (Coetzee & Monti, 2018; Monti et al., 2007, 2009; Rodriguez-Moreno & Hirsch, 2009) and the aforementioned neuropsychological data (Reverberi et al., 2009), it was not statistically meaningful. It remains to be understood whether this is a consequence of methodological limitations of the approach (e.g., the degree to which deep medial regions can be efficiently stimulated *via* transcranial magnetic stimulation) or whether it should have been expected given the general understanding that deduction might well rely on the “concerted operation of several, functionally distinct, brain areas” (Reverberi et al., 2012), thus making it a harder process to disrupt with single-location stimulation. Consistent with this understanding, we have previously voiced the view that “core” deductive processes might be implemented in multiple brain areas, including both the mesial BA8 target as well as left rostrolateral prefrontal cortex, in BA10 (Monti & Osherson, 2012).

Finally, the failure of cTBS to differentially affect problems with three versus four variables is relevant to two ongoing debates. With respect to logic reasoning, the fact that cTBS to mesial BA8 impaired equally three- and four-variable logic problems ($t_{14} = -1.18$, $p = 0.13$) is contrary to the idea that this region can be explained by non-deductive processes such as working memory demands imposed by complex deductions (Kroger, Nystrom, Cohen, & Johnson-Laird, 2008) or greater relational complexity (Halford et al., 2010), confirming recent neuroimaging data (Coetzee & Monti, 2018). With respect to linguistic reasoning, these results bear on the question of the role of Broca’s area in language processing (Friederici, 2004, 2016; Rogalsky & Hickok, 2011) and suggest that this region (and the *pars opercularis* target in particular) is key to processing the hierarchical, non-local, dependencies of natural language (Friederici, 2004, 2016) and not just a reflection of verbal working memory (Rogalsky & Hickok, 2011). For, not only does cTBS to this region impair manipulating long-distance relationship across non-canonical sentences, but it also fails to differentially affect three- versus four-variable problems (t_{14} = -0.197; p = 0.43), contrary to what a verbal working memory account would predict.

In conclusion, this work presents direct, causal, evidence from the adult healthy brain demonstrating that abstract logic reasoning is ontologically distinct (Lenartowicz, Kalar, Congdon, & Poldrack, 2010; Price & Friston, 2005) from the mechanisms of natural language, contrary to the popular idea that Broca's area serves as a supra-modal hierarchical parser across domains of human thought (Fadiga et al., 2009; Rosenbaum, Cohen, Jax, Weiss, & van der Wel, 2007; Tettamanti & Weniger, 2006).

References

Amalric, M., & Dehaene, S. (2016). Origins of the brain networks for advanced mathematics in expert mathematicians. *Proceedings of the National Academy of Sciences*, *113*(18), 4909–4917. https://doi.org/10.1073/pnas.1603205113

Boeck, C. (2010). *Language in cognition:Uncovering mental structures and the rules behind them* (Vol. 1). John Wiley & Sons.

Carruthers, P. (2002). The cognitive functions of language. *Behavioral and Brain Sciences*, *25*(06). https://doi.org/10.1017/S0140525X02000122

Cesana-arlotti, A. N., Martín, A., & Téglás, E. (2018). Title : Precursors of logical reasoning in preverbal infants, *1266*, 25–27.

Chiang, J. N., Rosenberg, M. H., Bufford, C. A., Stephens, D., Lysy, A., & Monti, M. M. (2017). The language of music: Common neural codes for structured sequences in music and natural language. *BioRxiv*. Retrieved from http://biorxiv.org/content/early/2017/10/12/202382.abstract

Christov-Moore, L., Sugiyama, T., Grigaityte, K., & Iacoboni, M. (2017). Increasing generosity by disrupting prefrontal cortex. *Social Neuroscience*, *12*(2), 174–181. https://doi.org/10.1080/17470919.2016.1154105

Coetzee, J. P., & Monti, M. M. (2018). At the core of reasoning: Dissociating deductive and non-deductive load. *Human Brain Mapping*, *39*(4), 1850–1861. https://doi.org/10.1002/hbm.23979

Deblieck, C., Thompson, B., Iacoboni, M., & Wu, A. D. (2008). Correlation between motor and phosphene thresholds: a transcranial magnetic stimulation study. *Human Brain Mapping*, *29*(6), 662–70. https://doi.org/10.1002/hbm.20427

Fadiga, L., Craighero, L., & D'Ausilio, A. (2009). Broca's area in language, action, and music. *Annals of the New York Academy of Sciences*, *1169*, 448–458. https://doi.org/10.1111/j.1749-6632.2009.04582.x

Fitch, W. T., & Martins, M. D. (2014). Hierarchical processing in music, language, and action: Lashley revisited. *Annals of the New York Academy of Sciences*, *1316*(1), 87–104. https://doi.org/10.1111/nyas.12406

Fitzgerald, P. B., Fountain, S., & Daskalakis, Z. J. (2006). A comprehensive review of the effects of rTMS on motor cortical excitability and inhibition. *Clinical Neurophysiology*, *117*(12), 2584–2596. https://doi.org/10.1016/j.clinph.2006.06.712

Friederici, A. D. (2004). Processing local transitions versus long-distance syntactic hierarchies. *Trends in Cognitive Sciences*, *8*(6), 245–247. https://doi.org/10.1016/j.tics.2004.04.013

Friederici, A. D. (2016). The Neuroanatomical Pathway Model of Language: Syntactic and Semantic Networks. In *Neurobiology of Language* (pp. 349–356). https://doi.org/10.1016/B978-0-12-407794-2.00029-8

Halford, G., Wilson, W., & Phillips, S. (2010). Relational knowledge: the foundation of higher cognition. *Trends in Cognitive Sciences*, *14*(11), 497–505. https://doi.org/10.1016/J.TICS.2010.08.005

Holbrook, C., Izuma, K., Deblieck, C., Fessler, D. M. T., & Iacoboni, M. (2016). Neuromodulation of group prejudice and religious belief. *Social Cognitive and Affective Neuroscience*, *11*(3), 387–394. https://doi.org/10.1093/scan/nsv107

Huang, Y. Z., Edwards, M. J., Rounis, E., Bhatia, K. P., & Rothwell, J. C. (2005). Theta burst stimulation of the human motor cortex. *Neuron*, *45*(2), 201–206.

https://doi.org/10.1016/j.neuron.2004.12.033

Iaria, G., & Petrides, M. (2007). Occipital sulci of the human brain: variability and probability maps. *Journal of Comparative Neurology*, *501*(2), 243–259.

Klucharev, V., Munneke, M. A. M., Smidts, A., & Fernandez, G. (2011). Downregulation of the Posterior Medial Frontal Cortex Prevents Social Conformity. *Journal of Neuroscience*, *31*(33), 11934–11940. https://doi.org/10.1523/JNEUROSCI.1869-11.2011

Knauff, M., Fangmeier, T., Ruff, C. C., & Johnson-Laird, P. N. (2003). Reasoning, models, and images: behavioral measures and cortical activity. *Journal of Cognitive Neuroscience*, *15*(4), 559–73. https://doi.org/10.1162/089892903321662949

Kroger, J. K., Nystrom, L. E., Cohen, J. D., & Johnson-Laird, P. N. (2008). Distinct neural substrates for deductive and mathematical processing. *Brain Research*, *1243*, 86–103. https://doi.org/10.1016/j.brainres.2008.07.128

Lenartowicz, A., Kalar, D. J., Congdon, E., & Poldrack, R. A. (2010). Towards an Ontology of Cognitive Control. *Topics in Cognitive Science*, *2*(4), 678–692. https://doi.org/10.1111/j.1756-8765.2010.01100.x

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: an MEG study. *Nature Neuroscience*, *4*(5), 540–545. https://doi.org/10.1038/87502

Mayka, M. A., Corcos, D. M., Leurgans, S. E., & Vaillancourt, D. E. (2006). Three-dimensional locations and boundaries of motor and premotor cortices as defined by functional brain imaging: A meta-analysis. *NeuroImage*, *31*(4), 1453–1474. https://doi.org/10.1016/j.neuroimage.2006.02.004

Monti, M. M. (2017). The role of language in structure-dependent cognition. In M. Moody (Ed.), *Neural Mechanisms of Language*. New York NY, NY, US: Springer.

Monti, M. M., & Osherson, D. N. Logic, language and the brain, 1428 Brain Research § (2012). https://doi.org/10.1016/j.brainres.2011.05.061

Monti, M. M., Osherson, D. N., Martinez, M. J., & Parsons, L. M. (2007). Functional neuroanatomy of deductive inference: A language-independent distributed network. *NeuroImage*, *37*(3), 1005–1016. https://doi.org/10.1016/j.neuroimage.2007.04.069

Monti, M. M., Parsons, L. M., & Osherson, D. N. (2009). The boundaries of language and thought in deductive inference. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(30), 12554–9. https://doi.org/10.1073/pnas.0902422106

Monti, M. M., Parsons, L. M., & Osherson, D. N. (2012). Thought Beyond Language: Neural Dissociation of Algebra and Natural Language. *Psychological Science*, *23*(8), 914–922. https://doi.org/10.1177/0956797612437427

Niskanen, E., Julkunen, P., Säisänen, L., Vanninen, R., Karjalainen, P., & Könönen, M. (2010). Group-level variations in motor representation areas of thenar and anterior tibial muscles: Navigated transcranial magnetic stimulation study. *Human Brain Mapping*, *31*(8), 1272–1280. https://doi.org/10.1002/hbm.20942

Oberman, L., Edwards, D., Eldaief, M., & Pascual-Leone, A. (2011). Safety of Theta Burst Transcranial Magnetic Stimulation: A systematic review of the literature. *Journal of Clinical Neurophysiology*, *28*(1), 67–74. https://doi.org/10.1097/WNP.0b013e318205135f

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*(7), 674–681. https://doi.org/10.1038/nn1082

Peirce, J. W. (2008). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, *2*(January), 1–8. https://doi.org/10.3389/neuro.11.010.2008

Prado, J., Mutreja, R., & Booth, J. R. (2013). Fractionating the Neural Substrates of Transitive Reasoning: Task-Dependent Contributions of Spatial and Verbal Representations. *Cerebral Cortex*, *23*(3), 499–507. https://doi.org/10.1093/cercor/bhr389

Price, C. J., & Friston, K. J. (2005). Functional ontologies for cognition: The systematic definition of structure and function. *Cognitive Neuropsychology*, *22*(3–4), 262–275. https://doi.org/10.1080/02643290442000095

Reverberi, C., Bonatti, L. L., Frackowiak, R. S. J., Paulesu, E., Cherubini, P., & Macaluso, E. (2012). Large scale brain activations predict reasoning profiles. *NeuroImage*, *59*(2), 1752–1764. https://doi.org/10.1016/j.neuroimage.2011.08.027

Reverberi, C., Cherubini, P., Frackowiak, R. S. J., Caltagirone, C., Paulesu, E., & Macaluso, E. (2010). Conditional and syllogistic deductive tasks dissociate functionally during premise integration. *Human Brain Mapping*, *31*(9), 1430–1445. https://doi.org/10.1002/hbm.20947

Reverberi, C., Cherubini, P., Rapisarda, A., Rigamonti, E., Caltagirone, C., Frackowiak, R. S. J., … Paulesu, E. (2007). Neural basis of generation of conclusions in elementary deduction. *NeuroImage*, *38*(4), 752–762. https://doi.org/10.1016/j.neuroimage.2007.07.060

Reverberi, C., Shallice, T., D’Agostini, S., Skrap, M., & Bonatti, L. L. (2009). Cortical bases of elementary deductive reasoning: Inference, memory, and metadeduction.

Neuropsychologia, *47*(4), 1107–1116. https://doi.org/10.1016/j.neuropsychologia.2009.01.004

Rodriguez-Moreno, D., & Hirsch, J. (2009). The dynamics of deductive reasoning: An fMRI investigation. *Neuropsychologia*, *47*(4), 949–961. https://doi.org/10.1016/j.neuropsychologia.2008.08.030

Rogalsky, C., & Hickok, G. (2011). The role of Broca's area in sentence comprehension. *J Cogn Neurosci*, *23*(7), 1664–1680. https://doi.org/10.1162/jocn.2010.21530

Rosenbaum, D. A., Cohen, R. G., Jax, S. A., Weiss, D. J., & van der Wel, R. (2007). The problem of serial order in behavior: Lashley's legacy. *Human Movement Science*, *26*(4), 525–554. https://doi.org/10.1016/j.humov.2007.04.001

Rossi, S., Hallett, M., Rossini, P. M., & Pascual-Leone, A. (2009). Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. *Clinical Neurophysiology*, *120*(12), 2008–2039. https://doi.org/10.1016/j.clinph.2009.08.016

Sarfeld, A. S., Diekhoff, S., Wang, L. E., Liuzzi, G., Uludağ, K., Eickhoff, S. B., … Grefkes, C. (2012). Convergence of human brain mapping tools: Neuronavigated TMS Parameters and fMRI activity in the hand motor area. *Human Brain Mapping*, *33*(5), 1107–1123. https://doi.org/10.1002/hbm.21272

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., … Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, *23*, *Supple*, S208–S219. https://doi.org/10.1016/j.neuroimage.2004.07.051

Tettamanti, M., & Weniger, D. (2006). Broca's area: A supramodal hierarchical processor? *Cortex*, *42*(4), 491–494. https://doi.org/10.1016/S0010-9452(08)70384-8

Tsujii, T., Sakatani, K., Masuda, S., Akiyama, T., & Watanabe, S. (2011). Evaluating the roles of the inferior frontal gyrus and superior parietal lobule in deductive reasoning: An rTMS study. *NeuroImage*, *58*(2), 640–646. https://doi.org/10.1016/j.neuroimage.2011.06.076

Varley, R. (2001). Severe Impairment in Grammar Does Not Preclude Theory of Mind. *Neurocase*, *7*(6), 489–493. https://doi.org/10.1093/neucas/7.6.489

Varley, R. A., Klessinger, N. J. C., Romanowski, C. A. J., & Siegal, M. (2005). Agrammatic but numerate. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(9), 3519–3524. https://doi.org/10.1073/pnas.0407470102

Varley, R., & Siegal, M. (2000). Evidence for cognition without grammar from causal reasoning and “theory of mind” in an agrammatic aphasic patient. *Current Biology*, *10*(12), 723–726. https://doi.org/10.1016/S0960-9822(00)00538-8

APPENDIX: SUPPLEMENTS

Reasoning task			
Type of argument	**Number of terms**	**Matching**	**Non-matching**
Logic	**3**	If both X and Z then not Y. --- If Y then either not X or not Z.	If either Y or Z then not X. --- If X then both Y and Z.
Logic	**4**	If both X and not Z then either Y or not W. --- If both W and not Y then either Z or not X.	If both not Y and not W then both Z and X. --- If both Z and X then both not Y and not W.
Linguistic	**3**	It was X that Y saw Z take. --- Z was seen by Y taking X.	It was Y that Z thought X said. --- Z was thought by Y to have said X.
Linguistic	**4**	It was X that W heard Y saw Z take. --- W heard that Z was seen by Y taking X.	What W knew that Y gave Z was X. --- It was X that W knew was given to Y by Z.
Grammar task			
Type of argument	**Number of terms**	**No errors**	**Errors**
Logic	**3**	If either Y or X then not Z. --- If Y then either X or Z.	If not Y then Z both and X. --- If either not Z or not X then not Y.
Logic	**4**	If either X or W then both Y and Z. --- If both not Y and not W then both Z and X.	If both Z and not Y then either X or not W. --- If both W and Y then either not X not or Z.
Linguistic	**3**	Z was thought by Y to have said X. --- It was Y that X thought Z said.	It was to Y that from Z told X. --- What Z told Y was X.
Linguistic	**4**	Z knows that X is given by Y to W. --- If either W or X then both Y and not Z.	Z will be seen by Y taking X is what W will hear. --- It was X that W heard Y take Z saw.

Supplementary Table 1. Examples of stimuli.

	Linear trend weights	F	p
Linguistic	**T_{lin} Broca (-1) vs LTOS (+.5) & mesial BA8 (+.5)**	**7.697**	**0.015**
	T_{log} Broca (+.5) & LTOS (+.5) vs mesial BA8 (-1)	3.966	0.066º
Logic	T_{lin} Broca (-1) vs LTOS (+.5) & mesial BA8 (+.5)	0.038	0.849
	T_{log} Broca (+.5) & LTOS (+.5) vs mesial BA8 (-1)	**6.626**	**0.022**
Grammar	**T_{lin} Broca (-1) vs LTOS (+.5) & mesial BA8 (+.5)**	**11.22**	**0.005**
	T_{log} Broca (+.5) & LTOS (+.5) vs mesial BA8 (-1)	0.576	0.46

Supplementary Table 2. Linear trend analysis. For each task, contrast weights per stimulation site are given, followed by F value and significance. Significant trends highlighted in bold. (º As mentioned in the main text, marginal significance is due to a "reverse" effect where linguistic reasoning appear to ameliorate after mesial BA8 cTBS.)

Study 3: Is Deductive Reasoning a Modular or Domain General Process?

Introduction

The relatively poor performance of untrained individuals on modus tollens relative to modus ponens has been known since the early 20th century, as has the improvement in performance observed when such problems are framed differently (Wilkins, 1929). This facilitation is shown especially clearly in the work of Wason (1966, 1968) who developed the eponymous card selection task that bears his name. In this task, participants are presented with a conditional rule and four cards, and are asked which cards must be turned over in order to test the rule. If the conditional rule corresponds to $(P \rightarrow Q, P, \therefore Q)$, then the four cards correspond to cases of P, $\neg P$, Q, and $\neg Q$, with the correct answers being those which could lead to a falsification of the rule, namely P and $\neg Q$. In a classic version of the task, participants are given the rule “If there is a vowel on one side of the card, then there is an even number on the other side.” The cards then show a vowel, a consonant, an even number, and an odd number. The correct cards to turn over are the vowel and the odd number, because those are the cards that could lead to falsifying the rule. However, participants tend to turn over the cards named in the rule, namely the vowel and the even number, thereby committing the logical fallacy of affirming the consequent. Performance is dramatically enhanced if the rule is given a social context, especially one involving the policing of social rules (Wason, 1968).

There have been various attempts to account for this facilitation. One of the most prominent attempts is social exchange theory, developed by Cosmides and Tooby (Cosmides, 1985; Cosmides & Tooby, 1992, 2005) and rooted in evolutionary psychology. Under this theory, framing the Wason task in terms of a social rule is thought to lead to facilitated performance by activating a distinctive evolved cognitive module which has the purpose of processing social contracts (i.e. catching cheaters) and which has developed out of our specie's pressing need to cooperate socially in an efficient manner (Cosmides & Tooby, 1987).

An alternative explanation for the facilitation observed in the Wason task is permission schemas, advanced by Cheng and Holyoak (Cheng & Holyoak, 1989; Cheng & Holyoak, 1985). Under this view, such facilitation can be explained without resorting to specially evolved cognitive modules. It can instead be explained by domain general mechanisms, such as our gradual accumulation of schemas or frameworks for understanding social life. One type of schema, permission schemas, are thought to be especially common because we encounter situations involving permission to perform an action throughout lives, from the time we are children. Under this view, the familiarity of a rule is crucially important, and experiments in which rule familiarity is systematically manipulated have provided strong evidence for permission schemas (Cheng & Holyoak,1985).

Social exchange theory has undergone significant empirical and theoretical criticism (Evans et al.,1993). In some of the most well-known cases of facilitation, such as when participants are asked to find out who is drinking underage, it is hard to argue that any "exchange" is taking place (Cheng & Holyoak,1989). Another critique has to do with the reliance by social exchange theory on the Wason task. Sperber and Girotto (2003) have argued that participants undergoing a Wason selection task are very likely responding not to anything like the logical modus tollens structure, but instead to the particular details of the social scenario that is created to accompany it. Cosmides' own success at inducing participants to violate the logical structure of the task in the "switched" version of her task appears to support this view (Cosmides, 1985).

Although Cosmides has conducted numerous empirical tests of Social Exchange Theory, these have almost all been conducted on college undergraduates, who may differ in significant ways from the general population (Henrich, Heine, & Norenzayan, 2010). The primary three goals of the current study are 1) to test Cosmides' theory about the importance of particular "cues" to the activation of the purported social exchange and the hazard management modules, 2) to see whether the observed facilitation can be better explained by familiarity with the rule than by the cues, and 3) to see if the facilitation effects observed by Cosmides can be found in a large online Mechanical Turk sample, one that would be more substantially more diverse than those which her theory has been tested on before.

An additional goal of the current study was inspired by a proposal from Scott and Baron-Cohen (1996). They suggested that if the social exchange module does exist, then it should have differential effects on autistic individuals, who should fail to show the kind of facilitation in response to social scenarios that normal individuals typically do. This was tested by including a self-report measure of autism in the online study (the ASQ, also developed by Baron-Cohen (2001)). We assumed that individuals with psychopathy would be likely to show a reverse pattern, showing facilitation on the social exchange task, but failing to show it on the hazard management task, because of the complementary deficits in cognitive and emotional empathy in these two groups of people (autistic persons are thought to be low in the former, while psychopaths are thought to be low in the latter (Smith, 2006). A number of other measures of individual differences were also included.

Methods

Participants

A total of 415 participants were recruited through Amazon’s Mechanical Turk system over a period of approximately six days, with the majority (400) being recruited in the first day. Participants were told that they would be completing a variety of cognitive tests for a university research study, and that the entire task would take approximately 40 minutes (based on pilot samples). Approximately half the participants (208) were nominally located in the United States, based on Mechanical Turk selective filters (“nominally” because it is Mechanical Turk determines the location of their participants through self report). The remainder were nominally located in India. United States participants were offered $5.00 for completing the task, which at 40 minutes meant that they were earning 12.5 cents per minute, or $7.50 per hour, which is above the US federal minimum wage of $7.25 an hour (see the Fair Labor Standards Act). Participants who were nominally located in India were offered $2.50 per hour, an amount which is comparable to what US residents were paid, when the lower cost of living in India is taken into account (Lansingh et al., 2015). Mechanical Turk filters were used to prevent participants from doing the task more than once, and they were not permitted to complete the task unless they read an online consent form and indicated that they understood and accepted it. They were asked to perform the task on a laptop or desktop computer, but not on a smartphone, because of concerns that individuals on smartphones were likely to be in more distracting environments. They were given 12 hours to complete the task, but were encouraged to do so in one sitting in an area where they would not be distracted. They were also encouraged to read all instructions carefully, and were told beforehand that there would be catch trials to test whether they were paying attention. The consent form and study procedures were approved ahead of time by the University of California, Los Angeles, IRB review board.

Once the initial data was collected, it was necessary to determine whether any participants needed to be removed. The Qualtrics platform (on which the task was hosted) automatically recorded the GPS coordinates of participants and these were used to determine whether participants were in the United States, India, or elsewhere. Seven participants were located in countries other than the United States or India, and so were excluded. Next, 92 participants who had answered catch trials at less than 50% accuracy were excluded. These constituted 22.2% of the initial respondents, and amount which was in line with expectations based on piloting. Another 14 participants had said they were familiar with the Wason task, and were removed for that reason. There were 74 participants who scored below 50% on the English Proficiency Test, most of whom were located in India, and these were removed as well.

Among the remaining participants, in the pool that was used for analysis, there were 228 individuals, 187 of which were located in the United States and 41 of which were located in India. With regard to sex, 144 participants were men and 84 were women. The average age was 34, and ages ranged from 19 to 71 years old. With regard to highest educational attainment, 19 were high school graduates, 51 had some college, 33 had either an associate’s degree or some technical certification, 90 had a Bachelor’s degree, 29 had Master’s degrees, and 2 had Ph.D.’s. With regard to language, 197 reported being Native English Speakers, with most of the remainder reporting either Hindi or Tamil as their first language. Participants were given an opportunity to provide feedback on the study, and many commented that they enjoyed taking it, although some complained about the length.

Stimuli

First, 108 different scenarios were constructed. In each, a rule was embedded in a verbally described scenario that gave context to the rule. At the end of each scenario, participants were asked to select those cards which needed to be turned over in order to test whether the rule was being followed (For examples see Supplements 1-3). Similar scenarios were organized into groups that were designed to either elicit the purported social exchange module, the hazard management module, or in the case of abstract scenarios, to elicit no facilitation. Social exchange scenarios were further organized into familiar and unfamiliar subgroups, and were systematically modified to yield sets or families of related scenarios which differed minimally and were designed to test the view that a special cognitive module for processing such scenarios is elicited by the presence of a distinct set of three cues: the presence of a benefit, the presence of intentional cheating, and the possibility of cheating. As such, each set or family of social exchange contained four versions, including a version with all cues, one in which there was no benefit (or minimal benefit) to be had, one in which rule violation could take place but would not be intentional, and one in which rule violation was either not possible or was very difficult. Hazard management scenarios were designed to test whether the elicitation of the purported module requires the presence of two cues. The first, suggested by Cosmides, is the presence of an effective precaution (Cosmides &

Tooby, 2005). The second, suggested by the current author, is the presence of a hazard. Hazard management scenarios were therefore organized into sets of three: a scenario with all cues, one with no hazard, and one with no precaution. Each set or family of related abstract scenarios contained only two members. Those which used the word “must” in their rule statement, and those which used some other word, such as “will.” This was in order to test whether the use of the word “must” had an effect of its own. All of the social exchange and hazard management modules used “must.”

Once the full set of 108 candidate stimuli had been created, participants were recruited from Mechanical Turk to rate the stimuli along ten dimensions. Participants were presented with the stimuli, and then with a series of statements about them, to which participants indicated their agreement or disagreement using a Likert scale. Examples include: “Someone who breaks the rule in this situation is a cheater,” and “Someone who follows the rule in this situation will be safe from harm.” Forty participants took part in the norming procedure, with half from India and half from the United States. The questions and norming procedure were based on procedures in Fiddick, Spampinato, and Grafman (2005). Once norming data was collected, a set of ideal responses was created for each class of scenarios. For example, the ideal response to the statement “Someone who breaks the rule in this situation is a cheater” would be “Strongly agree” (a one on our scale) for social exchange scenarios, and “Strongly disagree” for hazard management scenarios. Then, for each class of scenarios, the pattern of ideal responses was subtracted from average mTurk ratings for each candidate scenario, yielding a numerical rating of how distant, or different, each scenario was from the ideal for that category. Then, because we planned to use sets of similar scenarios, these difference scores were averaged across sets, and then we chose the scenarios within each category that scored highest and incorporated those into the primary experiment.

Individual differences

In addition to the rule-scenarios, the study included eight measures of individual differences, which were included to investigate whether such differences had meaningful effects on performance on the card selection task, and on how facilitated individuals are by the framing. The first individual differences measure included in the study was the Autism Spectrum Quotient (ASQ), a 50 item questionnaire developed by Baron-Cohen et al. (2001) as a brief but diagnostically valid measure of autism. Individuals who score above 32 on this scale can be classified as autistic with a high degree of confidence (Baron-Cohen et al. 2001). This was followed by the Levenson Self-Report Psychopathy Scale (Levenson, Kiehl, and Fitzpatrick 1995) a 26 item measure of psychopathic traits which includes two sub factors, and which has shown good validity and reliability in previous studies (Miller, Gaughan, and Pryor 2008; Walters et al. 2008). There are no cutoff thresholds reported for this measure, but it has shown to be usefully correlated with psychopathic traits in a variety of situations. Next came the Analytic-Holistic Scale (Choi, Koo, & Jong An Choi 2007), a scale which is designed to measure individuals' propensity to engage in different styles of thought by asking a series of questions that break down into four sub-factors: beliefs about causality, attitudes toward contradictions, perception of change, and locus of attention. This was followed by the Big Five Inventory (BFI) (John, O. P., & Srivastava 1999) a 44 item version which produces results that correlate highly with longer tests designed to measure so-called “Big 5” traits. We omitted 4 questions while constructing the Qualtrics version of this test, so that our version had 40 questions, but we do not believe this should affect the results in any important way. As a measure of English Proficiency, we included the 20 item University of Washington Online English Language Self-Placement Test (UWOELST), which we modified slightly because we found some questions to be unclear. After data collection we also removed three questions that native English speakers performed at a low rate on. This was followed by a series of demographic questions about age, sex, level of education, whether participants had heard of the Wason task before, what their first language was, and whether they had lived in another country before. The next test was the Cognitive Reflection Test 2 (CRT2) (Thomson & Oppenheimer, 2016). This four item test is an updated version of the older Cognitive Reflection Test (CRT) (Frederick, 2005) and was created because of concerns that the older CRT may have become too widely known about to maintain validity. Like the original CRT, the CRT2 is designed as a measure of the tendency to conserve cognitive resources (Thomson & Oppenheimer, 2016) and consists of questions which tend to mislead the inattentive or unreflective participant into giving incorrect answers. The final measure which participants were asked to complete was a short 12 item version of the Advanced Raven’s Progressive Matrices (ARPM) (Arthur Jr & Day, 1994; Raven, 1938). This test has been shown to correlate with the longer version of the ARPM, and with a variety of other cognitive abilities, especially abstract intelligence.

Design

The experiment itself was hosted on the Qualtrics platform. When participants at Mechanical Turk agreed to participate, they followed a link which took them to Qualtrics.

They then read brief instructions indicating that they should complete the task on a laptop or desktop, and not on a mobile phone. This was followed by a UCLA approved consent form. After the indicated understanding and consent, they read instructions which told them how long the test would take, how much time they had to complete them, and which warned them about the presence of catch trials and the importance of reading instructions carefully. Once they had read the instructions, they could continue to Part 1, which contained the Wason scenarios, and which they were told should take them approximately 20 minutes to complete.

In Part 1, each participant viewed ten Wason scenarios. Some properties of these differed between participants, and some elements differed for the same participant within the same sitting. The presentation order of all ten scenarios was randomized for each participant. There were four catch trials that were seen by each participant, and these were scenarios that were not used for the experimental trials. They looked like normal scenarios, except that at the end there was a sentence that said “Please select all four cards and proceed to the next trial.” If they failed to do this they received a warning that they should read each scenario carefully before answering. Each participant also saw one familiar and one unfamiliar abstract scenario, although whether these contained the word “must” or not in the rule was randomized across participants. Each participant also saw one familiar and one unfamiliar hazard management scenario, although whether each of these was the all cues, no hazard, or no precaution version was randomized across participants. Similarly, each participant saw one familiar and one unfamiliar social exchange scenario, although whether these included the version with all cues, no benefit, no intention, or the version where it was not possible to cheat was randomized across participants.

After completing the Wason scenarios, the participant was shown the instructions for Part 2. Like the first part, it was expected to take 20 minutes, and included the individual difference measures listed above. The order of the measures was not randomized, but the within each test the order of questions was randomized (with the exception of the short ARPM, which is intended to be taken in order from easiest to most difficult)(Arthur Jr. & Day, 1994). At the end of the experiment, the participant was given a code which they could enter into their Mechanical Turk interface so that they would be paid. They were also provided with the researcher’s email address so that they could contact the researcher if difficulties arose.

Analysis

After removing participants who failed to meet requirements (as described above) the first stage of analysis focused on the Wason scenarios. First, a performance score was calculated for each scenario. This was done by scoring each individual card in each scenario as either correct or incorrect, so that a participant who, for example, chose only the cards corresponding to P and $\neg$Q would receive a score of 1, while a participant who chose the cards corresponding to P, $\neg$P, and $\neg$Q would receive a score of .75. A series of ANOVAS were then conducted to see which categorical variables of interest had an effect on either raw performance or on facilitation (with facilitation being defined as the difference in accuracy between the hazard scenarios and the abstract scenarios, or between the social exchange scenarios and the abstract scenarios). This analysis focused primarily on the role of the “cues” that Cosmides has suggested are necessary to trigger that cognitive modules that lead to facilitation, and on the role of familiarity (Cheng & Holyoak, 1985; Cosmides, Barrett, & Tooby 2010).

Following, the next stage of the analysis focused on associations between raw performance on the Wason scenarios and the various measures of individual differences, and then on the association between facilitation scores (obtained by subtracting performance on abstract scenarios from performance on hazard or social exchange scenarios. In order to reduce the number of factors that needed to be tested, a principal components analysis (PCA) was conducted, with the eigenvalue threshold set at 1, and then variables that loaded at .7 or higher onto each of the resulting components were combined with those. Subsequent correlations were conducted between the scenario scores and the components that came out of the PCA. All analyses were conducted with SPSS version 21 on a desktop computer running Windows 10.

Results

Core analysis

We began by analyzing the raw accuracy scores (with accuracy being calculated for each trial on a per card basis, as described above). Across all participants and conditions, the mean accuracy was 73.2% (*SD* = 8.1%). A 2 X 3 repeated-measures ANOVA was conducted to test the effects of familiarity (familiar vs. unfamiliar) and scenario type (abstract vs. hazard vs. social exchange on accuracy, as reflected by the raw scores. ASQ score (low vs high) and PSQ score (low vs high) were treated as between subjects variables (note that for this analysis the ASQ and PSQ scores were divided in half and the lower 50% treated as "low" while the upper 50% was treated as "high"). From this analysis we found a main effect of both familiarity and scenario type (*F*(1, 224) = 8.5, *p* = .004, and *F*(2, 448) = 69.9, *p* < .001, respectively). Specifically, participants exhibited lower mean accuracy for unfamiliar scenarios (*M* = 71.6%, *SD* = 19.6%), as compared to familiar ones (*M* = 74.7%, SD = 16.6%). Follow up paired sample t-tests determined that the effect of familiarity was driven by the difference in raw score accuracy between familiar and unfamiliar abstract scenarios (*t*(227) = 2.86, *p* = .005), with accuracy being higher for familiar abstract trials than for unfamiliar. The difference between familiar and unfamiliar scenarios was not significant for either hazard or social exchange scenarios (*t*(227) = .720, *p* = .472 and *t*(227) = 1.02, *p* = .310), respectively (see Figure 1). Among the scenario types, participants exhibited the lowest accuracy on the abstract scenarios (*M* = 63.4%, *SD* = 22.6%), second lowest on the social exchange scenarios (*M* = 75.5%, *SD* = 22.6%), and highest on hazard management (*M* = 80.5%, *SD* = 21.1%). Follow up pairwise comparisons of the marginal means confirmed that all three scenario types were significantly different from each other at p < .001 (see Figure 1).

With regard to ASQ score, there was a main effect of low vs high *F*(1, 224) = 8.67, *p* = .004), with those who scored high on the ASQ being more accurate across scenario types (*M* = 76.3%, *SD* = 16%) as compared to those who scored low on the ASQ (*M* = 70.1%, *SD* = 16%). There was no interaction between ASQ score and any of the other factors in this analysis. Hypothesis-driven follow up t tests determined that individuals with high ASQ scores did experience facilitation on familiar hazard scenarios

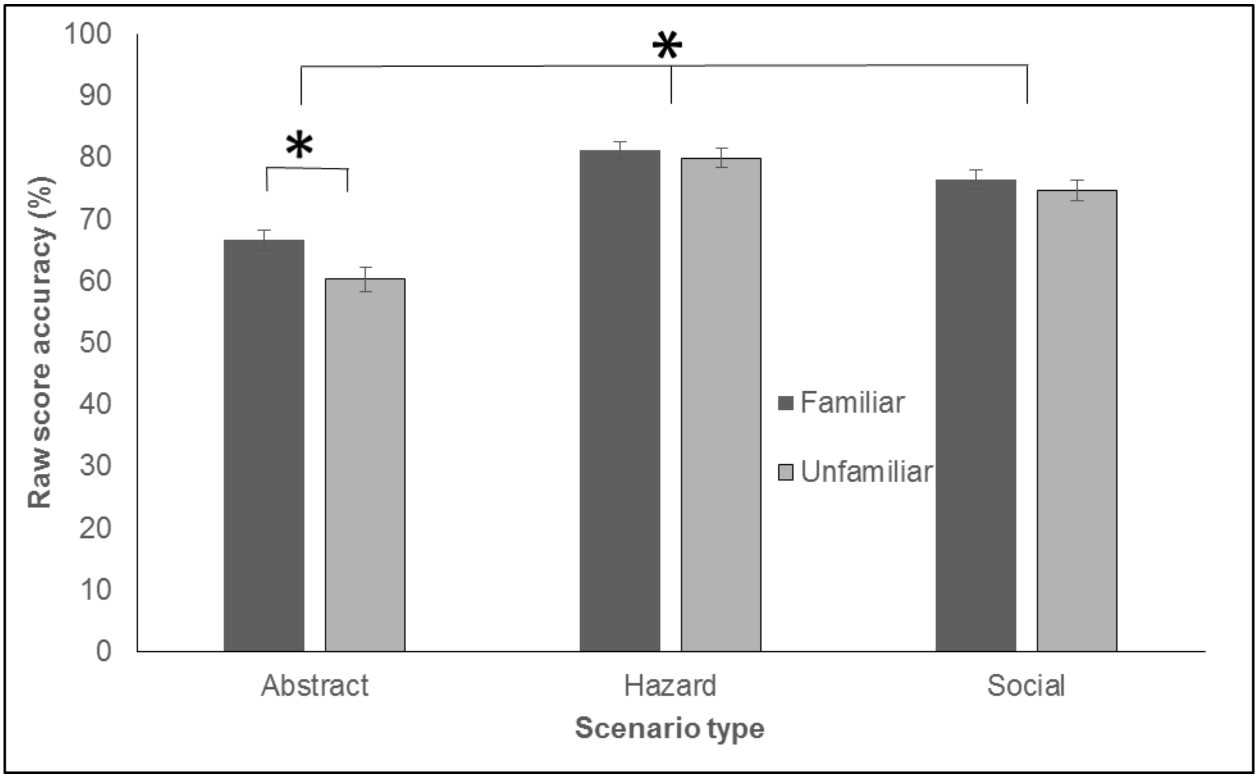


Figure 1. Accuracy by familiarity and scenario type. Raw scores. Familiarity was a significant factor, as was scenario type. All three scenario types differed significantly from each other. Within scenarios, only abstract had a significant difference between familiar and unfamiliar.

($t(114) = 3.63$, $p < .001$), and unfamiliar hazard scenarios ($t(114) = 6.58$, $p < .001$), as compared to familiar and unfamiliar abstract scenarios, respectively. Individuals with high ASQ scores also experienced facilitation on familiar social exchange scenarios ($t(114) = 3.50$, $p = .001$), and unfamiliar social exchange scenarios ($t(114) = 4.53$, $p < .001$), as compared to familiar and unfamiliar abstract scenarios, respectively (see Figure 2).

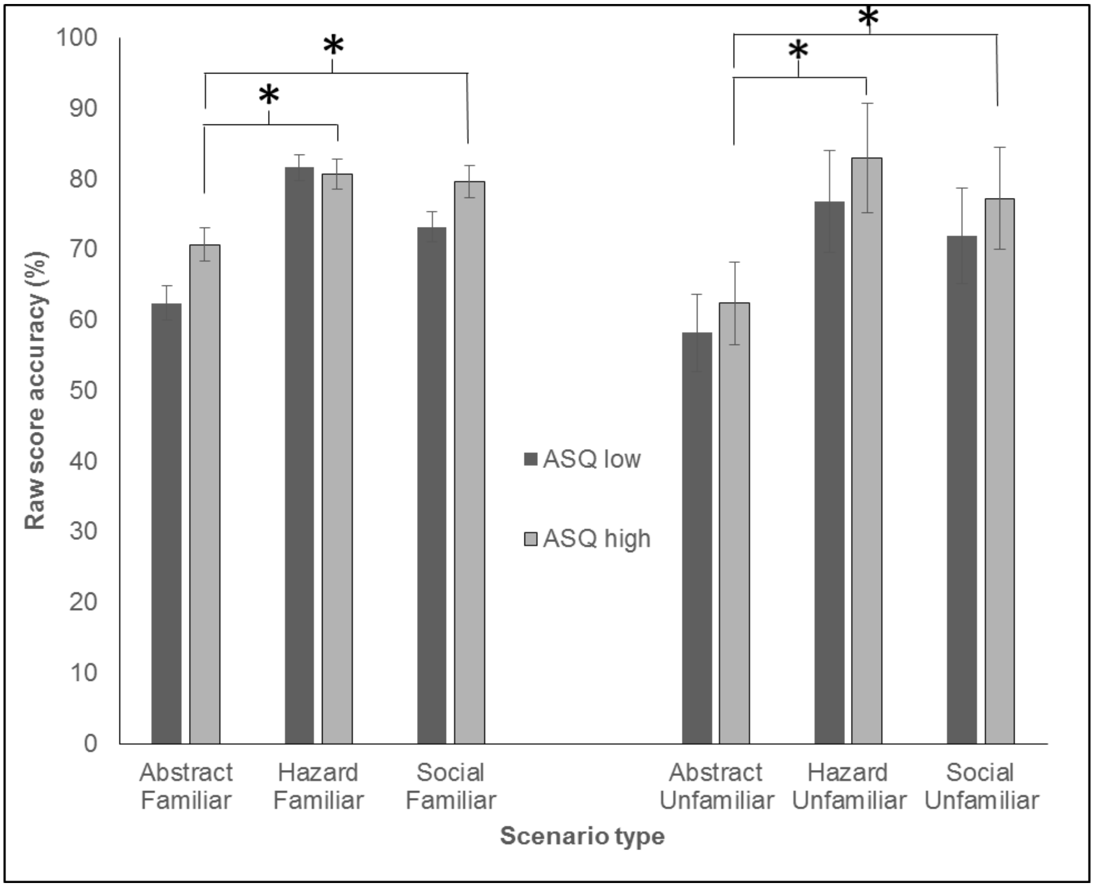


Figure 2. Raw score accuracy by ASQ score (low vs high), scenario type, and familiarity. Individuals with high ASQ scores performed better than individuals with low ASQ scores. Facilitation was exhibited for those with high ASQ scores on hazard scenarios and social exchange scenarios, regardless of familiarity.

The diagnostic threshold for the ASQ is 32 (Baron-Cohen et al., 2001) with individuals above that threshold considered likely to have an autism spectrum disorder. In our sample we only had 14 individuals who met this criteria, but because of the importance for our hypothesis we decided to test this sample anyway. For these individuals, there was facilitation on unfamiliar hazard scenarios (*t*(13) = 3.98, *p* = .002) as compared to unfamiliar abstract scenarios, and there was facilitation on unfamiliar social exchange scenarios (*t*(13) = 3.65, *p* = .003) as compared to unfamiliar abstract

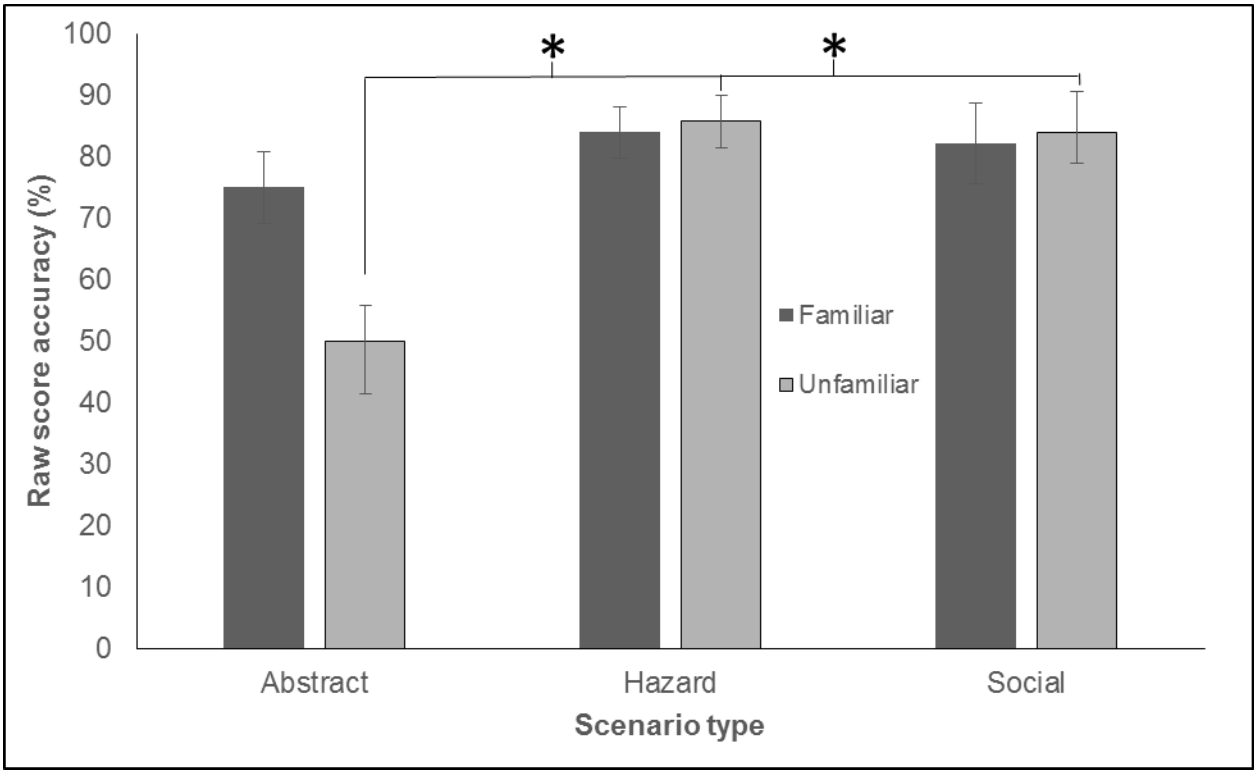


Figure 3. Facilitation for persons with ASD. These individuals had ASQ scores above 32 meaning they are likely to have autism spectrum disorder (ASD). They exhibited facilitation on the unfamiliar hazard and social scenarios, but not on the familiar ones.

scenarios, but there was no significant facilitation between familiar hazard and familiar abstract scenarios, or between familiar social exchange scenarios and familiar abstract scenarios (see Figure 3).

With regard to PSQ score, there was a main effect of low vs high *F*(1, 224) = 11.6, *p* = .001), with those who scored high on the PSQ being less accurate across scenario types (*M* = 69.6%, *SD* = 16.0%) as compared to those who scored low on the PSQ (*M* = 76.8%, *SD* = 16.0%). There was no interaction between PSQ score and any of the other factors in this analysis. Rather, higher PSQ scores were associated with lower scores on all scenarios, regardless of familiarity or scenario type (see Figure 4). This was the opposite of the pattern seen with ASQ scores. There was also no

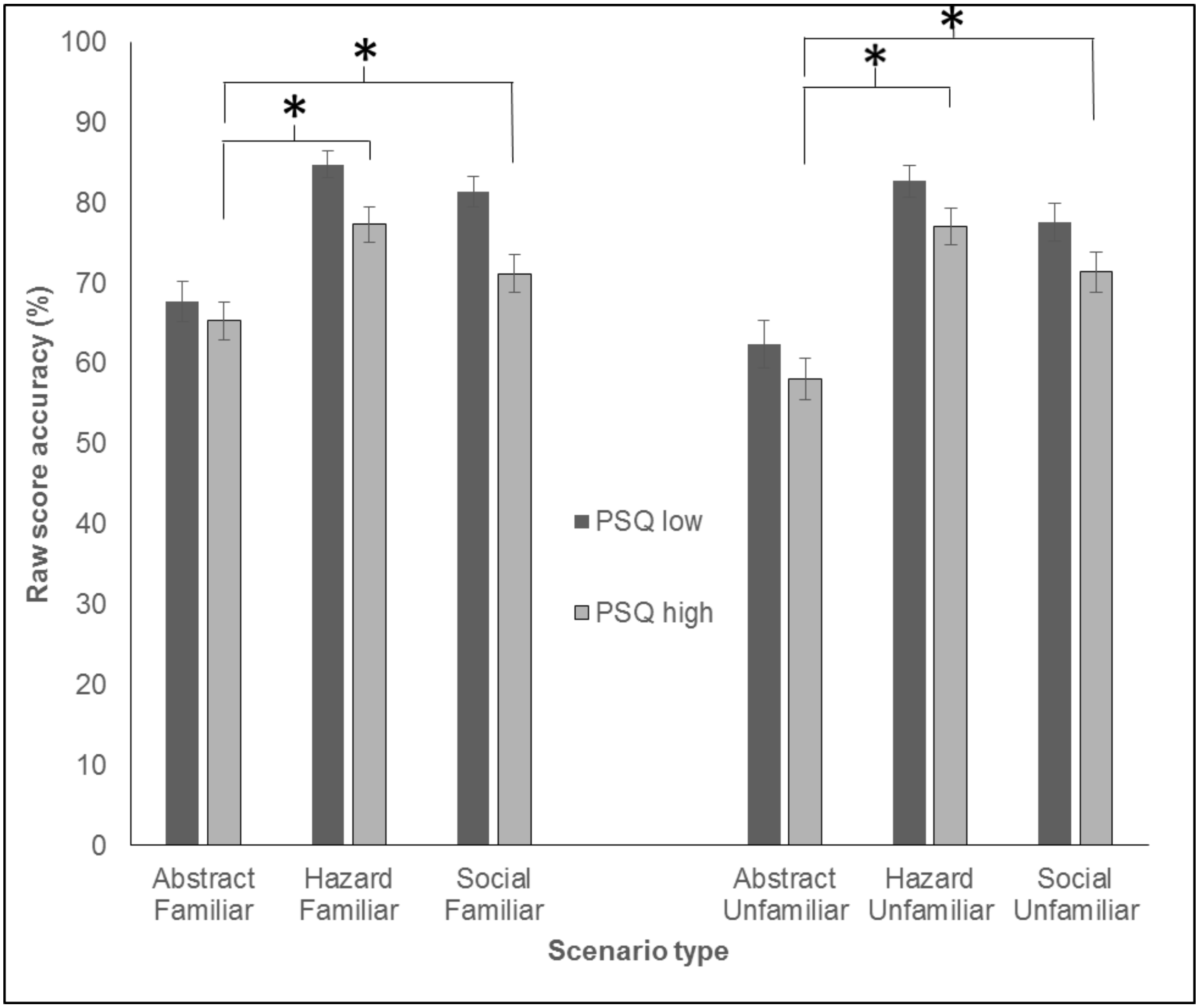


Figure 4. Facilitation for participants with high PSQ scores. High PSQ scores were associated with lower scores across scenarios, as compared participants with low PSQ scores. High PSQ scores were still associated with facilitation on hazard vs abstract and on social vs abstract scenarios.

interaction between ASQ and PSQ scores. Hypothesis driven follow up paired sample t tests showed that for persons high on PSQ there was facilitation on familiar hazard scenarios ($t(110) = 4.60$, $p < .001$) and on familiar social exchange scenarios ($t(110) = 6.00$, $p < .001$) as compared to familiar abstract scenarios, and there was facilitation on unfamiliar hazard scenarios ($t(110) = 2.17$, $p = .032$) and on unfamiliar social exchange scenarios ($t(110) = 3.98$, $p < .001$) as compared to unfamiliar abstract scenarios (see Figure 4).

Next, a test was conducted to determine whether the country of residence affected scores on either the abstract, hazard, or social scenarios. Because the number of participants from the US and from India were so different following the removal of people who scored below the English Proficiency threshold, a non-parametric Mann-Whitney U test was used. There was no difference based on country among the abstract familiar ($p = .154$) or abstract unfamiliar scenarios ($p = .310$), but there was a significant difference among the hazard (both familiar ($p < .001$) and unfamiliar ($p = .001$)) and social exchange (both familiar ($p = .002$) and unfamiliar ($p = .002$)) scenarios, with US accuracy scores being higher than Indian scores for all of these

Next, a series of four one way ANOVAs were used to test for cue-based differences among the raw performance scores. For hazard – familiar, there was no significant difference among the three subtypes (all cues, no hazard, no precaution) ($F(2, 225) = .028$, $p = .972$). For hazard management – unfamiliar, there was also no significant difference among the three subtypes ($F(2, 225) = .638$, $p = .529$. For social exchange – familiar, there was no significant difference among the four subtypes (all cues, no benefit, no intention, not possible) ($F(3, 224) = 1.17$, $p = .323$). For social – exchange, unfamiliar, there was also no significant difference among the four subtypes ($F(3, 224) = 2.26$, $p = .083$). In summary, there was no detectable effect of the cue subtypes on the raw performance scores in either the hazard management or the social exchange scenarios, regardless of whether they were of the familiar or the unfamiliar type.

Next, an independent samples t-test was conducted on the abstract familiar scenarios to determine whether the use of “must” vs. “not must” in the rule made a difference in the accuracy. It did not ($t(226) = .193$, $p = .847$). Another independent samples t-test was conducted to determine whether the use of “must” vs. “not must” made a difference in the accuracy among the abstract-unfamiliar scenarios. It did not ($t(226) = .059$, $p = .953$). Because abstract scenarios differed with regard to familiarity but not with regard to must vs. not must, when facilitation scores were calculated, this was accomplished by subtracting abstract-familiar from hazard-familiar and from social-familiar, and by subtracting abstract-unfamiliar from hazard-unfamiliar, and from social-unfamiliar, without regard to the use of must vs not must.

Once facilitation scores were calculated (as described above) another set of four one way ANOVAs was used for each of the four resulting sets of facilitation scores. For hazard familiar, there was no significant difference among the three subtypes (all cues, no hazard, no precaution) ($F(2, 225) = .078$, $p = .925$). For hazard unfamiliar, there was also no significant difference among the three subtypes ($F(2, 225) = .376$, $p = .687$). For social exchange familiar, there was no significant difference among the four subtypes (all cues, no benefit, no intention, not possible) ($F(3, 224) = 1.94$, $p = .125$). For social exchange, unfamiliar, there was also no significant difference among the four subtypes ($F(3, 224) = .709$, $p = .547$). There was no detectable effect of the cue subtypes on the facilitation scores in either the hazard management or the social exchange scenarios, regardless of whether they were of the familiar or the unfamiliar type.

Next, a 2 X 2 repeated measures ANOVA was conducted on the facilitation scores, with the first factor being familiarity and the second factor being scenario type (hazard vs. social exchange). ASQ score (low high) and PQ score (low high) were included as between subjects variables. The effect of familiarity in this case was not significant (*F*(1, 227) = 3.45, *p* = .065). The effect of scenario type was significant, with hazard showing more facilitation than social exchange (*F*(1, 227) = 17.5, *p* < .001). The interaction was not significant. ASQ and PQ also failed to be significant in this test.

Other individual differences

We began by doing a PCA analysis of all the continuous variable individual difference measures of interest to us. This included scores on the ASQ, LSRP (and the two subfactors), Analytic/Holistic scale (and the four subfactors), English Proficiency, Age, CRT2, and the 12 item ARPM (although we had already conducted analyses with ASQ scores and with PSQ scores, we decided it was justifiable to still include these as part of the PCA, since determining the effect to which these measures were associated with other individual difference measures may have explanatory value). We used a varimax rotation, and required Eigenvalues of >1. The PCA yielded four components, and our measures were considered to load onto a particular component if they had correlation estimates of .7 or higher. Tests that load positively onto component 1 include the total LSRP score, both of its subfactors, and (surprisingly) subfactor 3 from the Analytic/Holistic scale, which was described as “Perception of Change” (Choi, Koo, & Choi, 2007). Component 1 can be thought of as the “psychopathy factor.” Tests that load positively onto component 2 include the personality trait of Extraversion, while those that load negatively include ASQ score and the personality trait of Neuroticism.

Component 2 can be thought of as an “anti-autism” or “extraversion” factor. Tests that load positively onto component 3 include the overall score from the Analytic/Holistic Scale, as well as subfactor 2 from that test, described as “Attitude Toward Contradictions.” Component 3 can be thought of as a “holism” factor. Tests that load positively onto Component 4 include the CRT2 score and the ARPM. Component 4 can be thought of as a “cognitive ability” factor. See Table 1 for details.

Table 1. Rotated Component Matrix for PCA

	Component			
	1	2	3	4
ASQ score	.116	**-.806**	-.005	.010
LSRP score	**.930**	-.045	-.146	-.224
LSRP subfactor 1	**.845**	.151	-.157	-.241
LSRP subfactor 2	**.751**	-.388	-.074	-.111
Analytic Holistic overall score	-.156	.104	**.939**	.019
Analytic Holistic subfactor 1	-.198	-.139	.626	-.047
Analytic Holistic subfactor 2	-.183	.053	**.751**	-.017
Analytic Holistic subfactor 3	.870	-.092	-.181	-.137
Analytic Holistic subfactor 4	.270	.336	.567	.050
Big Five: Openness	-.107	.402	.213	.296
Big Five: Conscientiousness	-.539	.539	.174	-.025
Big Five: Extraversion	.172	**.771**	-.027	-.224
Big Five: Agreeableness	-.482	.499	.324	-.079
Big Five: Neuroticism	.160	**-.730**	.009	.047
Age	-.438	.060	-.056	-.280
CRT2 score	-.124	-.001	-.081	**.749**
ARM score	-.096	-.137	.006	**.743**

Loadings >= .7 are in ***bold****.*

Next, a bivariate correlational analysis was run between the four components derived from the PCA and each of the six scenario score measures (abstract, hazard, and social exchange, each in the familiar and unfamiliar version). With regard to the four PCA components, social exchange familiar was significantly correlated with component 1 (psychopathy) although this did not survive a Bonferroni correction. Component 2 (anti-autism) exhibited no relationship with any of the raw scores. Component 3 (holistic) also exhibited no relationship with any of the raw scores. Component 4 (cognition) had a significant correlation with raw performance on all scenarios, and all of these also survived a Bonferroni correction (with the exception of social exchange familiar) (see Table 2).

Table 2. Pearson correlations between PCA components and raw scenario scores

	Abstract Familiar	Abstract Unfamiliar	Hazard Familiar	Hazard Unfamiliar	Social Exch. Familiar	Social Exch. Unfamiliar
Comp. 1	-.062	-.037	-.131*	-.068	-.179*	-.116
Comp. 2	-.175*	-.096	-.032	-.142*	-.192*	-.141*
Comp. 3	.006	-.032	-.005	-.088	1	.044
Comp. 4	**.220***	**.278***	**.313***	**.276***	.156*	**.290***

* p < .05. Correlations that survive Bonferroni correction (p<.002) are in **bold**.

Next, a set of correlations was conducted between the four facilitation scores (hazard management and social exchange, both the familiar and the unfamiliar versions) and the four PCA components. There was one significant correlation, between hazard management familiar and component 4 (cognition) but it did not survive a

Bonferroni correction (see Table 3). Given the absence of meaningful correlations in the this facilitation score analysis, we did not attempt to obtain intra-scenario correlations for the different versions of each scenario (i.e. hazard familiar all cues, hazard familiar no hazard, etc.).

Table 3. Pearson correlations between PCA components and scenario facilitation scores

	Hazard Familiar	Hazard Unfamiliar	Social Exch. Familiar	Social Exch. Unfamiliar
Comp. 1	-.039	-.014	-.086	-.057
Comp. 2	.136*	-.013	.000	-.025
Comp. 3	-.010	-.032	.073	.061
Comp. 4	.030	-.056	-.068	-.019

* p < .05. No correlations survived the Bonferroni correction (p<.003).

Discussion

This is one of the largest, and certainly the most diverse participant samples that social exchange theory and hazard management theory have been tested on. That being the case, it is worth noting which traits of the scenarios had an effect on performance. First, familiarity had a significant effect on the raw scores, with performance on familiar scenarios being greater than performance on unfamiliar scenarios. The effect of familiarity did not extend, however, to the facilitation scores. The type of scenario also had an effect, at both the level of raw scores and the level of facilitation scores. Amongst raw scores, both hazard and social exchange scenarios engendered more accuracy than abstract scenarios. Also, amongst facilitation scores, hazard scenarios engendered more accuracy than social exchange scenarios. However, the “cues” that Cosmides (2010) has suggested are necessary to trigger the purported modules failed to have a detectable effect at every level of the analysis. This raises a question: if it is not the cues of hazard and precaution that make a hazard scenario what it is, and it is not the cues of benefit, intentionality, and possibility of cheating that make a social exchange scenario what it is, then from where were the facilitatory effects of these scenarios derived? It is unclear, but one possibility is that participants did not actually read the scenarios closely enough to notice which cues were present and which were not, but instead read them just closely enough to get the “gist” of the situation, and then filled in the blanks once they felt they knew what was going on. Such a strategy is fully compatible with the atmosphere heuristic, something that comes up frequently in the study of deductive reasoning (Reverberi et al., 2009).

With regard to autism, we failed to find any support for Scott and Baron-Cohen's (1996) hypothesis that autism would be associated with reduced facilitation in response to social exchange scenarios. Instead, high scores on the ASQ were associated with facilitation on both hazard and social exchange scenarios, whether they were familiar or unfamiliar, and participants with scores above the diagnostic threshold still displayed facilitation on the unfamiliar versions of the hazard and social exchange scenarios. However, it is still of interest that high ASQ scores were associated with higher scores across scenarios, and, that based on the PCA, higher ASQ scores did not appear to be associated with higher general cognitive ability (component 4).

With regard to psychopathy, we failed to find any support for our hypothesis that individuals with high PQ scores would not display facilitation on the hazard scenario.

Indeed, these participants displayed facilitation on hazard and social scenarios, whether they were familiar or unfamiliar. It is of interest, however, that individuals high in psychopathy displayed lower accuracy across scenarios, and that this reduced performance was not associated with general cognitive ability (component 4).

The clearest associational result in our study was with Component 4, cognitive ability. The CRT2 and the ARPM both loaded onto this component, and it had strong positive correlations with the raw scores for 5 out of 6 of the scenario types (although not with facilitation scores). This is perhaps not surprising, given that these test scores likely tap into exactly the sort of abilities that would be required to do well on the Wason task, either in its abstract or non-abstract forms.

This study had several limitations. The first is that it is difficult to show participants enough trials of this sort of task to make a repeated measures analysis possible. As the number of scenarios increases, participants will most likely become less likely to read each one carefully. As such, most studies of this sort have relatively few trials for the core test, and most of the analysis is done between participants (Cheng & Holyoak, 1985; Cosmides, 1985). Additionally, we did not expect that so many of our Indian participants would turn out to score so low on English proficiency. The reasons for this are unclear, but given that many other research studies make use of Indian participants on Mechanical Turk, they may want to consider including tests of English proficiency.

Future directions for this sort of research include further neuroimaging work on the Wason task, such as the work done by Stone et al. (2002). It could also be useful to use neuromodulation to try and selectively reduce or enhance facilitation on the task, perhaps in a manner similar to that done by Tsujii et al. (2010) for belief bias reasoning. Finally, given the artificial nature of the Wason task, it could be fruitful to translate the mechanics of the task into more naturalistic real world situations.

In conclusion, this study adds to the literature regarding contextual facilitation of deductive reasoning in several ways. First, it provides strong evidence that, even in a participant pool with diverse ages, educational achievements, and nationalities, such facilitation still does occur. Second, despite the large participant pool, it fails to find evidence that the cues proposed by Cosmides (2010) play any significant role. Third, although autism and psychopathy do not appear to play a role in the presence or absence of facilitation, they do appear to play a surprising role in overall performance on tasks such as these, and this is worthy of further investigation, since it may shed light on the cognitive abilities or obstacles faced by these populations.

References

Arthur Jr, Winfred and David V Day. 1994. “Development of a Short Form for the Raven Advanced Progressive Matrices Test.” *Educational and Psychological Measurement* 54(2):394–403.

Baron-Cohen, Simon, Sally Wheelwright, Richard Skinner, Joanne Martin, and Emma Clubley. 2001. “The Autism Spectrum Quotient : Evidence from Asperger Syndrome/High Functioning Autism, Males and Females, Scientists and Mathematicians.” *Journal of Autism and Developmental Disorders* 31(1):5–17.

Cheng, Patricia W. and Keith J. Holyoak. 1985. “Pragmatic Reasoning Schemas.” *Cognitive Psychology* 17(4):391–416. Retrieved October 27, 2013 (http://www.sciencedirect.com/science/article/pii/0010028585900143).

Cheng, Patricia W. and Keith J. Holyoak. 1989. “On the Natural Selection of Reasoning Theories.” *Cognition* 33(3):285–313. Retrieved August 29, 2015 (http://www.sciencedirect.com/science/article/pii/0010027789900310).

Choi, Incheol, Minkyung Koo, and Jong An Choi. 2007. “Individual Differences in Analytic versus Holistic Thinking.” *Personality and Social Psychology Bulletin* 33(5):691–705.

Choi, Incheol, Minkyung Koo, and Jong An Jong An Choi. 2007. “Individual Differences in Analytic Versus Holistic Thinking.” *Personality and Social Psychology Bulletin* 33(5):691–705. Retrieved July 14, 2017 (http://journals.sagepub.com/doi/10.1177/0146167206298568).

Cosmides, Leda. 1985. “Deduction or Darwinian Algorithms? An Explanation of the ‘Elusive’ Content Effect on the Wason Selection Task.” Retrieved November 7, 2015 (http://www.researchgate.net/publication/33882408_Deduction_or_Darwinian_Algorithms_An_Explanation_of_the_Elusive_Content_Effect_on_the_Wason_Selection_Task).

Cosmides, Leda, H. Clark Barrett, and John Tooby. 2010. “Colloquium Paper: Adaptive Specializations, Social Exchange, and the Evolution of Human Intelligence.” *Proceedings of the National Academy of Sciences of the United States of America* 107 Suppl(Supplement_2):9007–14. Retrieved September 23, 2015 (http://www.pnas.org/content/107/Supplement_2/9007.full).

Cosmides, Leda and John Tooby. 1987. “From Evolution to Behavior: Evolutionary Psychology as the Missing Link.”

Cosmides, Leda and John Tooby. 1992. “Cognitive Adaptations for Social Exchange.” Pp. 163–228 in.

Cosmides, Leda and John Tooby. 2005. “Neurocognitive Adaptations Designed for Social Exchange.” Pp. 584–627 in.

Evans, J. St. B. T., D. E. Over, and K. I. Manktelow. 1993. “Reasoning, Decision Making and Rationality.” *Cognition* 49(1–2):165–87. Retrieved October 27, 2013 (http://www.sciencedirect.com/science/article/pii/001002779390039X).

Fiddick, Laurence, Maria Vittoria Spampinato, and Jordan Grafman. 2005. “Social Contracts and Precautions Activate Different Neurological Systems: An FMRI

Investigation of Deontic Reasoning.” *NeuroImage* 28(4):778–86. Retrieved October 23, 2013 (http://www.sciencedirect.com/science/article/pii/S1053811905003873).

Frederick, Shane. 2005. “Cognitive Reflection and Decision Making.” *Journal of Economic Perspectives* 19(4):25–42. Retrieved (http://pubs.aeaweb.org/doi/10.1257/089533005775196732).

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Beyond WEIRD: Towards a broad-based behavioral science. Behavioral and Brain Sciences, 33(2-3), 111-135.

John, O. P., & Srivastava, S. 1999. “Big Five Inventory (Bfi).” *Handbook of Personality: Theory and Research* 2:102–138. Retrieved (http://www.fetzer.org/sites/default/files/images/stories/pdf/selfmeasures/Personality.pdf).

Lansingh, Van C. et al. 2015. “Affordability of Cataract Surgery Using the Big Mac Prices.” *Revista Mexicana de Oftalmologia* 89(1):21–30. Retrieved (http://dx.doi.org/10.1016/j.mexoft.2014.07.001).

Levenson, Michael R., Kent A. Kiehl, and Cory M. Fitzpatrick. 1995. “Assessing Psychopathic Attributes in a Noninstitutionalized Population.” *Journal of Personality and Social Psychology* 68(1):151–58. Retrieved (http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.68.1.151).

Miller, Joshua D., Eric T. Gaughan, and Lauren R. Pryor. 2008. “The Levenson Self-Report Psychopathy Scale: An Examination of the Personality Traits and Disorders Associated with the LSRP Factors.” *Assessment* 15(4):450–63.

Raven, John C. and others. 1938. *Raven's Progressive Matrices*. Western Psychological Services.

Reverberi, C., Rusconi, P., Paulesu, E., & Cherubini, P. (2009). Response demands and the recruitment of heuristic strategies in syllogistic reasoning. *Quarterly Journal of Experimental Psychology*, 62(3), 513-530.

Scott, Fiona J. and Simon Baron-Cohen. 1996. "Logical, Analogical, and Psychological Reasoning in Autism: A Test of the Cosmides Theory." *Development and Psychopathology* 8(01):235–45. Retrieved October 31, 2015 (files/17399/displayAbstract.html).

Smith, Adam. 2006. "Cognitive Empathy and Emotional Empathy in Human Behavior and Evolution." *The Psychological Record* 56(1):3–21.

Sperber, Dan and Vittorio Girotto. 2003. "Does the Selection Task Detect Cheater-Detection?" Pp. 197–226 in.

Stone, Valerie E., Leda Cosmides, John Tooby, Neal Kroll, and Robert T. Knight. 2002. "Selective Impairment of Reasoning about Social Exchange in a Patient with Bilateral Limbic System Damage." *Proceedings of the National Academy of Sciences* 99(17):11531–36. Retrieved October 27, 2013 (http://www.pnas.org/content/99/17/11531).

Thomson, Keela S. and Daniel M. Oppenheimer. 2016. "Investigating an Alternate Form of the Cognitive Reflection Test." *Judgment and Decision Making* 11(1):99.

Tsujii, Takeo, Sayako Masuda, Takekazu Akiyama, and Shigeru Watanabe. 2010. "The

Role of Inferior Frontal Cortex in Belief-Bias Reasoning: An RTMS Study." *Neuropsychologia* 48(7):2005–8. Retrieved (https://www.ncbi.nlm.nih.gov/pubmed/20362600).

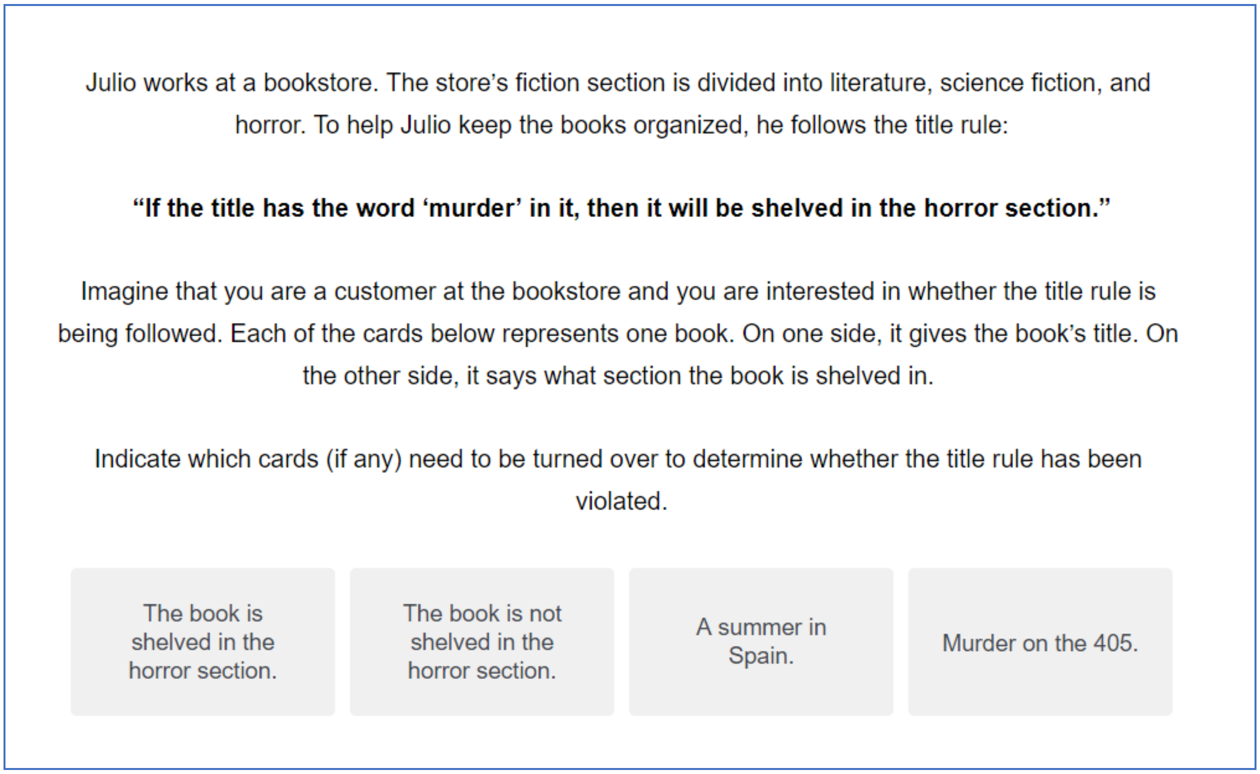
Walters, Glenn D., Chad A. Brinkley, Philip R. Magaletta, and Pamela M. Diamond. 2008. "Taxometric Analysis of the Levenson Self-Report Psychopathy Scale." *Journal of Personality Assessment* 90(5):491–98.

Wason, P. C. 1968. "Reasoning about a Rule." *Quarterly Journal of Experimental Psychology* 20(3):273–81. Retrieved October 27, 2013 (http://www.tandfonline.com/doi/abs/10.1080/14640746808400161).

Wason, Peter C. 1966. "Reasoning." Pp. 135–151 in, vol. 1.

Wilkins, M. C. 1929. "The Effect of Changed Material on Ability to Do Formal Syllogistic Reasoning." *Archives of Psychology* 102:83. Retrieved (http://psycnet.apa.org/psycinfo/1929-04403-001).

APPENDIX: SUPPLEMENTS



Julio works at a bookstore. The store's fiction section is divided into literature, science fiction, and horror. To help Julio keep the books organized, he follows the title rule:

"If the title has the word 'murder' in it, then it will be shelved in the horror section."

Imagine that you are a customer at the bookstore and you are interested in whether the title rule is being followed. Each of the cards below represents one book. On one side, it gives the book's title. On the other side, it says what section the book is shelved in.

Indicate which cards (if any) need to be turned over to determine whether the title rule has been violated.

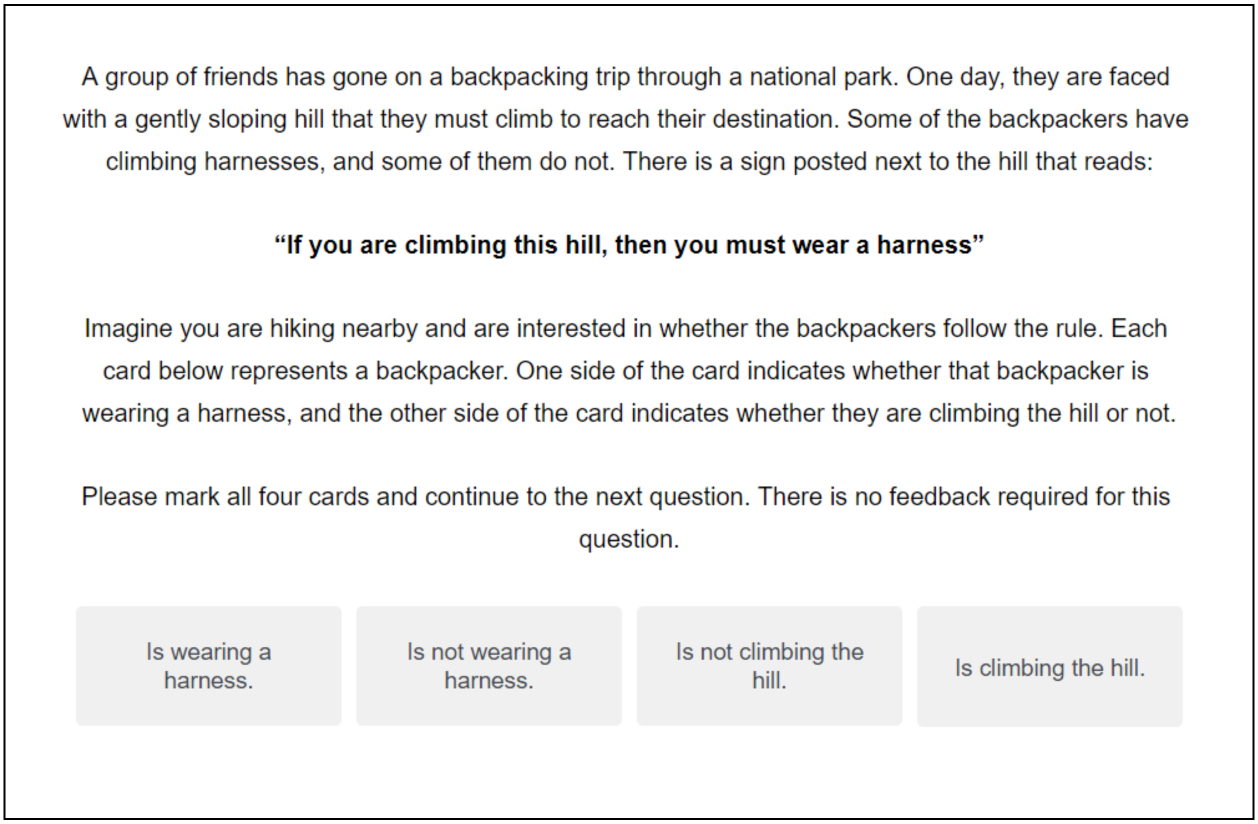
The book is shelved in the horror section.

The book is not shelved in the horror section.

A summer in Spain.

Murder on the 405.

Supplement 1. An example of the Abstract – Familiar – No Must condition.



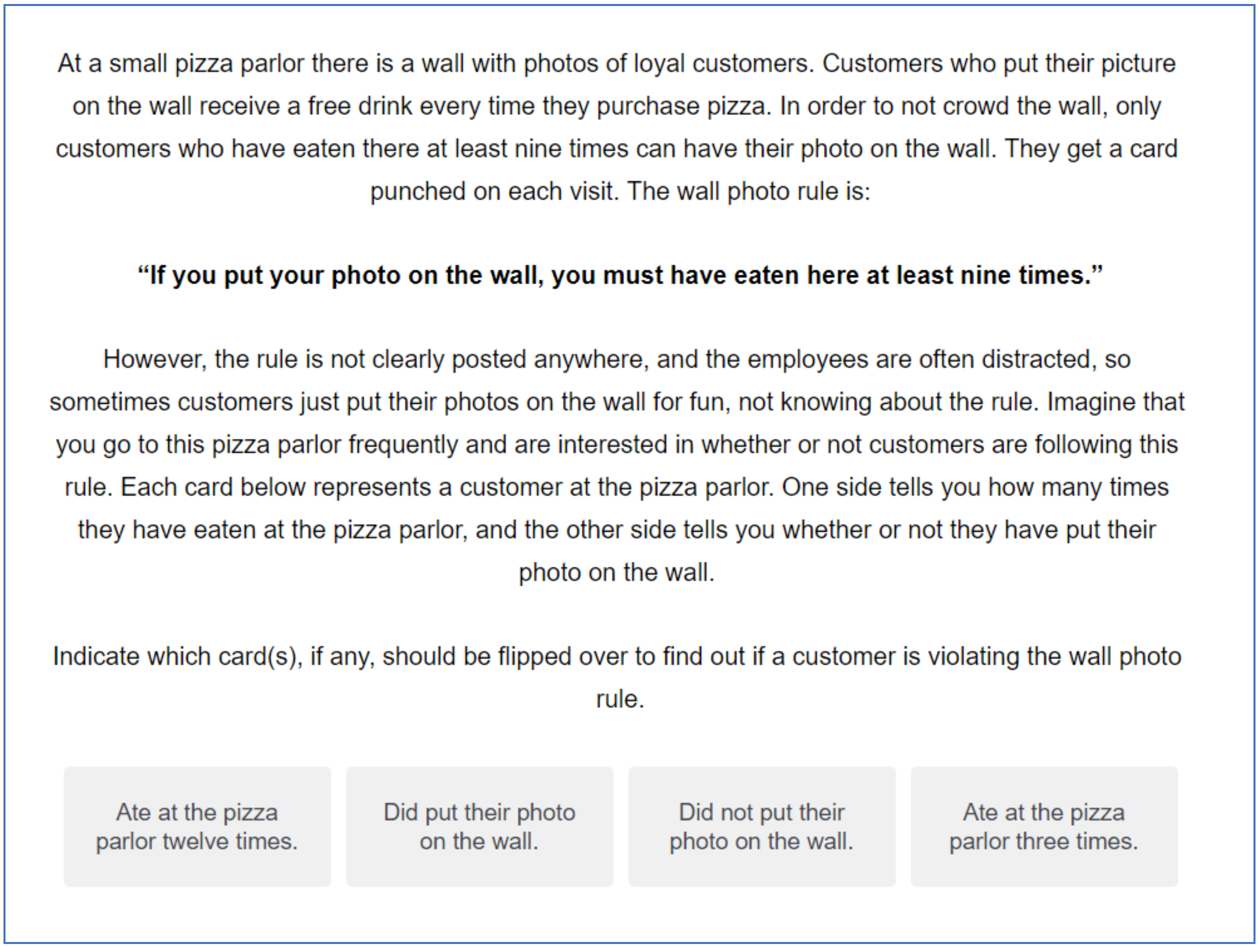
A group of friends has gone on a backpacking trip through a national park. One day, they are faced with a gently sloping hill that they must climb to reach their destination. Some of the backpackers have climbing harnesses, and some of them do not. There is a sign posted next to the hill that reads:

"If you are climbing this hill, then you must wear a harness"

Imagine you are hiking nearby and are interested in whether the backpackers follow the rule. Each card below represents a backpacker. One side of the card indicates whether that backpacker is wearing a harness, and the other side of the card indicates whether they are climbing the hill or not.

Please mark all four cards and continue to the next question. There is no feedback required for this question.

Supplement 2. An example of a catch trial.



At a small pizza parlor there is a wall with photos of loyal customers. Customers who put their picture on the wall receive a free drink every time they purchase pizza. In order to not crowd the wall, only customers who have eaten there at least nine times can have their photo on the wall. They get a card punched on each visit. The wall photo rule is:

"If you put your photo on the wall, you must have eaten here at least nine times."

However, the rule is not clearly posted anywhere, and the employees are often distracted, so sometimes customers just put their photos on the wall for fun, not knowing about the rule. Imagine that you go to this pizza parlor frequently and are interested in whether or not customers are following this rule. Each card below represents a customer at the pizza parlor. One side tells you how many times they have eaten at the pizza parlor, and the other side tells you whether or not they have put their photo on the wall.

Indicate which card(s), if any, should be flipped over to find out if a customer is violating the wall photo rule.

Supplement 3. An example of the Social Exchange – Familiar – No Intention condition.

Conclusions

This dissertation is a report on three studies investigating the nature of deductive reasoning, one of our most distinctively human traits, and what its relationship is to other psychological processes. Study 1 was an investigation of the relationship in the brain between deductive reasoning and both language and working memory. The fMRI study reported there both replicated prior work (Monti, Osherson, Martinez, & Parsons, 2007; Monti, Parsons, & Osherson, 2009; Reverberi, Shallice, D'Agostini, Skrap, & Bonatti, 2009; Rodriguez-Moreno & Hirsch, 2009), providing further evidence for a model of deductive reasoning centered on core regions in frontomedial cortex (Monti et al., 2007) and extended that work by demonstrating that deductive reasoning is dissociable from not only from language but also from working memory.

However, because of the fractionated nature of the field, it was necessary to find more decisive evidence for the view that language and deduction are separable. In Study 2, cTBS was successfully applied in order to generate causal evidence that language is dissociable in the brain from deductive reasoning. The evidence here for the role of mesial BA8 in deduction was less decisive, and that is an open avenue for future research. Frontopolar BA10 is a promising target for future neuromodulation of reasoning research, based on the findings in Study 1 and on previous work by Monti et al., (2007). The area near the border between left BA10 and left BA47 is especially promising (it was also an initial target of Study 2, but was abandoned because of the associate physical discomfort, future attempts to target this area should either be between participants (A. Costa et al., 2011; Alberto Costa et al., 2013), or should apply methods that are easier for participants to tolerate, such as focused ultrasound (Bystritsky et al., 2011; Monti, Schnakers, Korb, Bystritsky, & Vespa, 2016).

In Study 3, a large and diverse online participant pool was used to test Social Exchange Theory, and to investigate whether autism, psychopathy, personality, and other individual differences are associated with performance on the Wason task, as well as with facilitation by the contents of that task. The study provided support for the role of familiarity, and for the facilitatory effects of the different scenario categories, but not for the specific cues outlined by Cosmides (Cosmides, Barrett, & Tooby, 2010). With regard to individual differences, there was strong support for performance on the Wason task being associated with a factor that correlates with scores on the CRT2 and with scores on the 12 item APRM, but evidence was more mixed for the other combinations of individual differences that we tested. In this area future studies should apply neuromodulation techniques (Tsujii, Masuda, Akiyama, & Watanabe, 2010; Tsujii, Sakatani, Masuda, Akiyama, & Watanabe, 2011) or should test theories about deductive facilitation in more naturalistic ways.

Hopefully the research presented here has moved our understanding of the relationship between language and thought forward, and can serve as a stepping stone for future investigators.

References

Bystritsky, A., Korb, A. S., Douglas, P. K., Cohen, M. S., Melega, W. P., Mulgaonkar, A. P., … Yoo, S. S. (2011). A review of low-intensity focused ultrasound pulsation. *Brain Stimulation*, *4*(3), 125–136. https://doi.org/10.1016/j.brs.2011.03.007

Cosmides, L., Barrett, H. C., & Tooby, J. (2010). Colloquium paper: adaptive specializations, social exchange, and the evolution of human intelligence. *Proceedings of the National Academy of Sciences of the United States of America*, *107 Suppl*(Supplement_2), 9007–14. https://doi.org/10.1073/pnas.0914623107

Costa, A., Oliveri, M., Barban, F., Bonnì, S., Koch, G., Caltagirone, C., & Carlesimo, G. A. (2013). The Right Frontopolar Cortex Is Involved in Visual-Spatial Prospective Memory. *PLoS ONE*, *8*(2), 1–7. https://doi.org/10.1371/journal.pone.0056039

Costa, A., Oliveri, M., Barban, F., Torriero, S., Salerno, S., Lo Gerfo, E., … Carlesimo, G. A. (2011). Keeping Memory for Intentions: A cTBS Investigation of the Frontopolar Cortex. *Cerebral Cortex*, *21*(12), 2696–2703. https://doi.org/10.1093/cercor/bhr052

Monti, M. M., Osherson, D. N., Martinez, M. J., & Parsons, L. M. (2007). Functional neuroanatomy of deductive inference: A language-independent distributed network. *NeuroImage*, *37*(3), 1005–1016. https://doi.org/10.1016/j.neuroimage.2007.04.069

Monti, M. M., Parsons, L. M., & Osherson, D. N. (2009). The boundaries of language and thought in deductive inference. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(30), 12554–9. https://doi.org/10.1073/pnas.0902422106

Monti, M. M., Schnakers, C., Korb, A. S., Bystritsky, A., & Vespa, P. M. (2016). Non-

Invasive Ultrasonic Thalamic Stimulation in Disorders of Consciousness after Severe Brain Injury: A First-in-Man Report. *Brain Stimulation*. https://doi.org/10.1016/j.brs.2016.07.008

Reverberi, C., Shallice, T., D'Agostini, S., Skrap, M., & Bonatti, L. L. (2009). Cortical bases of elementary deductive reasoning: Inference, memory, and metadeduction. *Neuropsychologia*, *47*(4), 1107–1116. https://doi.org/10.1016/j.neuropsychologia.2009.01.004

Rodriguez-Moreno, D., & Hirsch, J. (2009). The dynamics of deductive reasoning: An fMRI investigation. *Neuropsychologia*, *47*(4), 949–961. https://doi.org/10.1016/j.neuropsychologia.2008.08.030

Tsujii, T., Masuda, S., Akiyama, T., & Watanabe, S. (2010). The role of inferior frontal cortex in belief-bias reasoning: An rTMS study. *Neuropsychologia*, *48*(7), 2005–2008. https://doi.org/10.1016/j.neuropsychologia.2010.03.021

Tsujii, T., Sakatani, K., Masuda, S., Akiyama, T., & Watanabe, S. (2011). Evaluating the roles of the inferior frontal gyrus and superior parietal lobule in deductive reasoning: An rTMS study. *NeuroImage*, *58*(2), 640–646. https://doi.org/10.1016/j.neuroimage.2011.06.076