BRAIN RESEARCH 1428 (2012) 24–32

Available online at www.sciencedirect.com

SciVerse ScienceDirect

www.elsevier.com/locate/brainres

BRAIN RESEARCH

Research Report

Levels of conflict in reasoning modulate right lateral prefrontal cortex

Melanie Stollstorff[a,b,*], *Oshin Vartanian*[a,c], *Vinod Goel*[a,d]

[a]*Department of Psychology, York University, Toronto, Ontario, Canada*
[b]*Department of Psychology and Neuroscience, University of Colorado, Boulder, CO, USA*
[c]*Department of Psychology, University of Toronto, Toronto, Ontario, Canada*
[d]*Department of Psychology, The University of Hull, Hull, UK*

ARTICLE INFO

Article history:
Accepted 18 May 2011
Available online 25 May 2011

Keywords:
Belief bias
Conflict
fMRI
Logic
Prefrontal cortex
Transitive inference

ABSTRACT

Right lateral prefrontal cortex (rlPFC) has previously been implicated in logical reasoning under conditions of conflict. A functional magnetic resonance imaging (fMRI) study was conducted to explore its role in conflict more precisely. Specifically, we distinguished between belief–logic conflict and belief–content conflict, and examined the role of rlPFC under each condition. The results demonstrated that a specific region of rlPFC is consistently activated under both types of conflict. Moreover, the results of a parametric analysis demonstrated that the same region was modulated by the level of conflict contained in reasoning arguments. This supports the idea that this specific region is engaged to resolve conflict, including during deductive reasoning.

This article is part of a Special Issue entitled "The Cognitive Neuroscience of Thought".

1. Introduction

Deductive reasoning can be tested using abstract, content-free arguments such as, "X is to the left of Y. Y is to the left of Z. Therefore, X is to the left of Z." Judging the validity of this abstract argument probes pure logical reasoning ability. However, it is not often that we encounter arguments in this pure, abstract form in everyday life. By adding meaningful content to the argument, we emulate real-life reasoning conditions more closely.

The content of an argument can affect one's ability to reason logically. Content in accordance with one's beliefs can help, whereas conflicting content can hinder our ability to reason – this is known as the belief–bias effect (Evans, 1989; Evans et al., 1983; Johnson-Laird and Bara, 1984; Oakhill and Johnson-Laird, 1985). An example of a congruent argument (one that is easier to solve because the belief-based response and the logic response are the same) is the following: "Canada is North of USA. USA is North of Mexico. Therefore, Canada is North of Mexico." We believe that Canada is in fact North of Mexico, and are more likely to accept this conclusion to be logically valid. An example of an incongruent argument (one that is more difficult to solve because the belief-based and logical responses diverge) is: "USA is North of Mexico. Mexico is North of Canada. Therefore, USA is North of Canada." This conclusion is not true according to our beliefs, but it is logically valid, and thus creates a conflict between logic and beliefs. Incongruent problems contain belief–logic conflict and are more difficult to solve than belief–logic congruent problems (Gilinsky and Judd, 1994; Goel et al., 2000; Goel and Dolan, 2003).

* *Corresponding author at:* Department of Psychology and Neuroscience, University of Colorado, Muenzinger D244, 345 UCB, Boulder, CO 80309–0345, USA.
E-mail address: melanie.stollstorff@colorado.edu (M. Stollstorff).

0006-8993/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.brainres.2011.05.045

The present study is concerned primarily with the cognitive and neural bases of reasoning under conditions of conflict. Several previous studies have found a brain region to be activated during reasoning when beliefs and logic conflict. In two functional magnetic resonance imaging (fMRI) studies, right lateral prefrontal cortex (rlPFC) was activated during reasoning trials containing belief–logic conflict compared to belief–logic congruent trials (at nearby MNI voxel coordinates): 54, 28, 26 and 51, 21, 12 (Goel et al., 2000; Goel and Dolan, 2003). Subsequent research using event-related potentials (ERP; Luo et al., 2008), functional near-infrared spectroscopy (fNIRS; Tsujii et al., 2010b; Tsujii and Watanabe, 2009, 2010), and regional transcranial magnetic stimulation (rTMS;Tsujii et al., 2010a) have also found rlPFC to be involved in reasoning trials containing belief–logic conflict in reasoning. Although this region appears to be activated by belief–logic conflict, further research is necessary to determine (1) whether this result extends to other types of deductive reasoning, and (2) whether its involvement is specific to the reasoning process or whether the presence of unbelievable sentences itself is sufficient for rlPFC activation.

To address the first issue, we conducted an fMRI experiment of belief–logic conflict using a different type of reasoning. Previous studies have used categorical syllogisms such as, "All apples are red fruit. All red fruit are poisonous. Therefore, all apples are poisonous." The present study utilizes three-term spatial relational arguments to determine whether or not the rlPFC finding will extend to transitive inference. We also introduce a new condition – neutral content – into the belief–logic conflict analysis. Previous studies have emphasized real-world knowledge that is either believable or unbelievable. By adding neutral content we can compare trials with belief–logic conflict to neutral trials, as opposed to comparing them only to congruent trials. If it is the conflict that is activating rlPFC, this comparison to neutral trials should also reveal rlPFC activation.

To address the second issue, we investigate a previously uninvestigated source of interference in reasoning which we termed "belief–content conflict." Previous fMRI studies only measured belief–logic conflict, which occurs only at the conclusion (when the conclusion's validity conflicts with what one believes to be true). Belief–content conflict, however, refers to the conflict between one's beliefs and the content of a statement (when a statement in the logic problem conflicts with what one believes to be true). Thus, belief–content conflict is a source of interference that does not arise from an interaction between beliefs and logical validity (as does belief–logic conflict). Rather, it arises from an interaction between beliefs and the proposition being read (premises or conclusion). Specifically, an unbelievable sentence in a reasoning problem creates a conflict between what we know to be true, and what we must temporarily accept as true for the purpose of constructing a mental representation of the problem. Let us consider an example: "Canada is north of Mexico. Mexico is north of USA. Therefore, USA is north of Canada." This is a logically invalid argument, and the conclusion is not true according to our beliefs, so this example contains no belief–logic conflict and is considered congruent (i.e., our beliefs say "false" and the logic says "invalid"). However, two sentences (the second premise and the conclusion) are false according to our beliefs ("unbelievable"), thereby creating belief–content conflict (our beliefs say one thing, but the sentence says another). An argument can contain no belief–logic conflict, but still contain belief–content conflict (arising from the unbelievable sentences, such as those in the example given above). Thus, belief–content conflict is characterized by a mismatch between the content of a reasoning problem and one's beliefs (resulting from unbelievable sentences). Neural activation associated with this type of conflict has not yet been investigated.

Belief–logic conflict has only two levels: (1) present – trials containing belief–logic conflict (incongruent), or (2) absent – trials with no belief–logic conflict (congruent or belief-neutral). Belief–content conflict, however, varies parametrically in the amount of conflict, that is, the number of unbelievable sentences, contained in the argument. Because belief–content conflict has multiple levels of interference, a parametric analysis can be conducted to measure neural activation under varying levels of conflict. Thus, it can be determined whether rlPFC, implicated in previous studies of belief–logic conflict, is also modulated by the amount of belief–content conflict contained within the reasoning problem. In addition, the present study investigates whether the mere presence of false propositions (unbelievable sentences) is sufficient for rlPFC activation independent of belief–logic conflict.

2. Results

2.1. *Behavioral*

Separate analysis of variance (ANOVA) tests were conducted for both reaction time and accuracy to test for behavioral differences in belief–logic conflict (congruent, belief-neutral, incongruent) and belief–content conflict (5 levels of interference).

2.1.1. *Belief–logic conflict*

Reaction time and accuracy were assessed to determine whether the presence of belief–logic conflict (incongruent trials) affects reasoning relative to trials without belief–logic conflict (congruent or belief-neutral trials). This would establish whether or not the classic belief–bias effect was demonstrated. Incongruent, congruent and belief-neutral trials were included in a within-subjects ANOVA. For reaction time, as predicted, there was a main effect of congruency [$F(2, 30)=4.74$, $p<.05$]. Post-hoc analysis revealed that participants responded significantly faster on congruent (3472 ms) relative to incongruent (3796 ms) trials [$F(1, 30)=3.48$, $p<.01$]. Reaction time for neutral trials (3542 ms) fell in between the two means, but did not differ significantly from either congruent or incongruent trials ($ps>.05$). The main effect of congruency on accuracy was marginally significant, $F(2, 30)=3.12$, $p<.06$. Post-hoc comparisons revealed that participants were more accurate for congruent (79.5%) than incongruent (73.5%) ($F(1, 30)=2.65$, $p<.05$) and belief-neutral (74.0%) ($F(1, 30)=2.46$, $p<.05$) trials. Therefore, results indicated that incongruent trials, characterized by a conflict between logic and beliefs, were associated with increased reaction times and decreased accuracy relative to congruent trials, thereby replicating the belief–bias effect (Gilinsky and Judd, 1994; Goel et al., 2000; Goel and Dolan, 2003).

Table 1 – Experimental design—belief–logic conflict (incongruent, congruent and neutral trials).

	Familiar content (72 trials) [comprised of Canadian Content (36 trials) and Learned Fictional Environment (36 trials)]		Unfamiliar content (belief-Neutral) (36 trials)
Reasoning (108 trials)	*Congruent (36 trials)*	*Incongruent (36 trials)*	
	Toronto is East of Halifax. Edmonton is West of Halifax. Toronto is East of Edmonton. *(valid, conclusion believable)*	Toronto is East of Halifax. Edmonton is East of Toronto. Halifax is West of Edmonton. *(valid, conclusion unbelievable)* ***Belief–Logic Conflict***	Fells River is west of Rostall. Jaredon is east of Rostall. Jaredon is east of Fells River.
	Toronto is East of Halifax. Edmonton is West of Halifax. Toronto is West of Edmonton. *(invalid, conclusion unbelievable)*	Toronto is East of Halifax. Edmonton is East of Toronto. Halifax is East of Edmonton. *(invalid, conclusion believable)* ***Belief–Logic Conflict***	
Baseline (40 trials)	Toronto is East of Halifax. Edmonton is East of Toronto. Vancouver is North of Montreal. *(conclusion unrelated to premises)*		

2.1.2. *Belief–content conflict*

Performance for the 5 levels of belief–content conflict was analyzed (level 1: 3 believable sentences; level 2: belief-neutral sentences, unfamiliar map condition; level 3: 1 unbelievable sentence; level 4: 2 unbelievable sentences; level 5: 3 unbelievable sentences); see Table 2. As predicted, the results showed that as the level of belief–content conflict increased, participants had slower reaction times, $F(4, 56)=3.70$, $p<0.01$. Post-hoc comparisons revealed that reaction times were significantly faster in level 1 (all believable sentences, 3281 ms) relative to levels 3 (3679 ms), 4 (3758 ms) and 5 (3641 ms; $p<0.05$), but was not significantly different from level 2 (neutral sentences, 3542 ms). No other post-hoc comparisons were significant. Accuracy did not differ by level of belief–content conflict ($p>0.05$; level 1: 81.1%, level 2: 74.0%, level 3: 74.5%, level 4: 72.8%, level 5: 73.3%).

2.2. *fMRI*

Data was analyzed using SPM2 (Friston et al., 1995). The first five volumes of the session were discarded, leaving 280 volumes per participant in total. Each volume was realigned to the first image of the session. The images were smoothed with an isotropic Gaussian kernel with FWHM of 12mm to allow for between-subject comparisons (Worsley and Friston, 1995). A random effects design was utilized. Condition effects at each voxel were estimated using the general linear model and compared using linear contrasts. Each contrast produced a statistical parametric map of the *t*-statistic at each voxel, which was transformed to a normal Z-distribution (Friston, et al., 1995).

At the group level, a categorical analysis was conducted to determine neural activation associated with belief–logic conflict: (a) incongruent minus congruent trials; (b)

Table 2 – Experimental design—levels of belief–content conflict: example stimuli and behavioral results.

Level of interference	Content	Examples	Accuracy	Reaction time
Level 1	All believable sentences (18 trials)	Toronto is west of Halifax. ✓ Edmonton is west of Toronto. ✓ Halifax is east of Edmonton. ✓	81.1%	3281 ms
Level 2	Belief-neutral, unfamiliar (36 trials)	Rostall is west of Triarch. ? Triarch is west of Rheemore. ? Rheemore is west of Rostall. ?	74.0%	3542 ms
Level 3	1 unbelievable sentences (18 trials)	Toronto is west of Halifax. ✓ Edmonton is east of Toronto. ✗ Halifax is east of Edmonton. ✓	74.5%	3679 ms
Level 4	2 unbelievable sentences (18 trials)	Toronto is west of Halifax. ✓ Edmonton is east of Toronto. ✗ Halifax is west of Edmonton. ✗	72.8%	3758 ms
Level 5	3 unbelievable sentences (18 trials)	Toronto is east of Halifax. ✗ Edmonton is east of Toronto. ✗ Halifax is west of Edmonton. ✗	73.3%	3641 ms

incongruent minus belief-neutral trials; (c) incongruent minus congruent trials, Canadian geography; (d) incongruent minus congruent trials, familiar fictional geography. A parametric analysis was conducted to determine regions modulated by level of belief–content conflict. Parameter estimates were extracted from the activated cluster in rlPFC using MARSBAR (Brett et al., 2002). Finally, a categorical analysis was conducted to determine activation associated with unbelievable conclusions (unbelievable>believable) in both congruent and incongruent trials separately. Unless otherwise indicated, both types of familiar content trials (fictional and Canada) were collapsed for analysis. All results reported survived $p<0.001$ uncorrected for multiple comparisons, unless otherwise indicated. Results from all imaging analyses are reported in Table 3.

2.2.1. *Belief–logic conflict*

As predicted, incongruent trials, characterized by a conflict between logic and belief, were associated with rlPFC activation in the same area that was activated in two previous studies (Goel et al., 2000; Goel and Dolan, 2003). Regions that were more active during incongruent than congruent reasoning trials (incongruent>congruent) included rlPFC (see Fig. 2a), as well as left fusiform gyrus. Regions more active during incongruent than belief-neutral trials (incongruent>neutral) included rlPFC, as well as left inferior parietal cortex, left superior parietal cortex, left putamen and left precuneus (see Fig. 2b). Both comparisons (incongruent>congruent and incongruent>neutral) revealed activation in the same region, rlPFC, found in previous studies.

Belief–logic conflict was also examined within the two conditions of content familiarity. Two separate analyses were conducted to identify regions more active for incongruent relative to congruent reasoning trials for each of the two familiar map conditions (Canadian geography and familiar fictional content). This comparison (incongruent>congruent) revealed significant activation exclusively in rlPFC, both for the Canadian geography condition (see Fig. 2c) and for the learned fictional environment condition (see Fig. 2d). Thus, belief–logic conflict activated the same area of the brain (a region within rlPFC) for both types of familiar content (learned fictional and real geography).

Table 3 – Regions engaged during various forms of conflict in reasoning.

Location	MNI coordinates X	Y	Z	Z score
Incongruent>Congruent				
Right lateral prefrontal	58	14	28	3.31
Left fusiform gyrus	−28	−86	−12	3.98
Incongruent>Neutral				
Right lateral prefrontal	60	20	24	4.02
Right lateral prefrontal	58	22	28	3.66
Left inferior parietal	−46	−34	26	3.85
Left superior parietal	−36	−72	46	3.65
Left putamen	−20	18	0	3.45
Left precuneus	−2	−64	36	3.38
Incongruent>Congruent (Canadian Geography)				
Right lateral prefrontal	60	20	24	3.12
Incongruent>Congruent (Learned Fictional Environment)				
Right lateral prefrontal	60	12	32	2.96
Belief–Content Conflict (parametric analysis, 5 levels)				
Right lateral prefrontal	60	20	24	4.04
Left superior parietal	−38	−74	48	3.80
Left supramarginal gyrus	−48	−56	56	3.26
Unbelievable>Believable Conclusions (Congruent trials only)				
Right lateral prefrontal	50	36	30	2.12
Unbelievable>Believable Conclusions (Incongruent trials only)				
Right lateral prefrontal	58	12	20	1.82

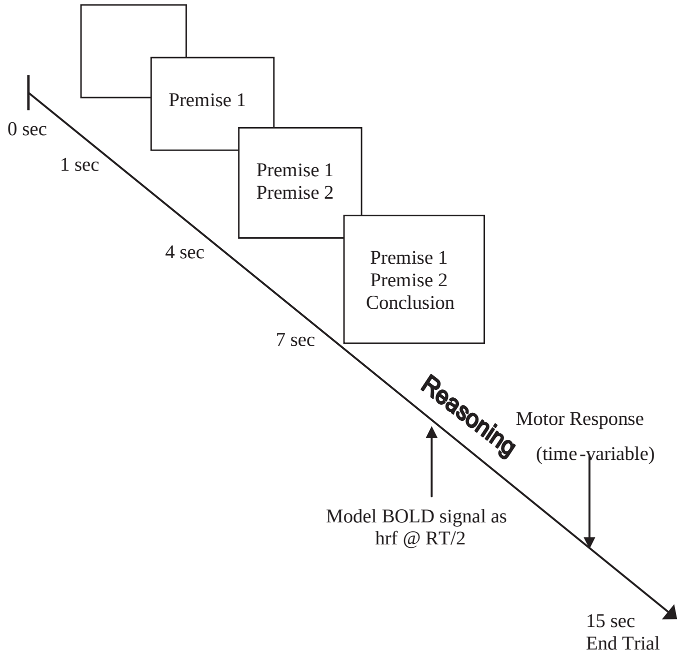
2.2.2. *Belief-content conflict*

The parametric analysis of belief–content conflict (5 levels) revealed significant activation in rlPFC (see Fig. 3), as well as left superior and inferior parietal cortex. Thus, the predicted area of rlPFC was modulated by the level of belief–content conflict present in the logic problem.

One problem with the above analysis is that the belief–logic conflict varied with belief–content conflict to some degree. That is, the problems with the highest number of unbelievable sentences (and thus high belief–content conflict) tended to be incongruent, whereas at lower levels of belief–content conflict, there were fewer incongruent problems. Therefore, the above 'belief–content conflict' analysis could have been primarily driven by belief–content conflict, belief–logic conflict, or both. To address this concern, a third type of analysis was conducted to test whether the presence of unbelievable conclusions was sufficient for rlPFC activation, controlling for congruency (and thus eliminating the effects of belief–logic conflict). We excluded the effect of belief–logic conflict (by holding this factor constant) and examined regions more active for unbelievable than believable conclusions. No regions survived the $p<.001$ threshold. We therefore reduced the threshold to $p<.05$ and report activation within rlPFC. Within congruent problems only, rlPFC was more active for unbelievable than believable conclusions ($p<.05$). Within incongruent problems, the same contrast (unbelievable>believable conclusions) revealed activation in the same region within rlPFC ($p<.05$). This demonstrates that the rlPFC activation increases with the presence of unbelievable sentences independent of belief–logic conflict, establishing belief–content conflict as a contributing factor in itself.

3. Discussion

The main focus of the present study was to investigate the role of the right lateral prefrontal cortex in reasoning during various conditions of conflict. Neural activation in a specific area within rlPFC was found in 'conflict' conditions of reasoning in the four analyses of belief–logic conflict. No other regions were consistently active in these comparisons. This is the first study to investigate the neural basis of belief–logic conflict in relational reasoning (transitive inference) problems and also the first to compare incongruent (belief–logic conflict) trials to belief-neutral trials. Not only was this region within rlPFC activated during belief–logic conflict, replicating and extending results of previous



Note. Event-related stimulus presentation for a single trial. Trial begins at 0 sec with a blank screen, followed by Premise 1 a second later, then Premise 2 at 4 seconds, then the conclusion at 7 seconds. Participants then have 8 seconds to make a response. The BOLD signal is measured at the halfway point between presentation of conclusion and the motor response, when reasoning occurs.

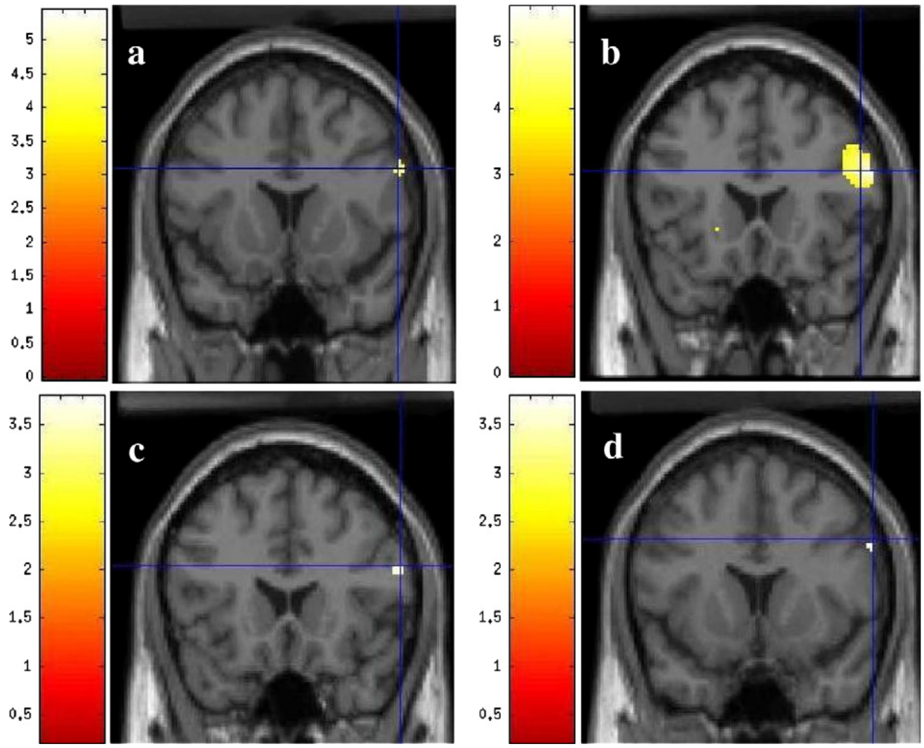
Fig. 1 – Trial structure.

studies that used categorical reasoning problems (Goel et al., 2000; Goel and Dolan, 2003; Tsujii et al., 2010a, 2010b; Tsujii and Watanabe, 2009, 2010), but the same area was also modulated by the amount of belief–content conflict in the argument; that is, the mere presence of counterfactual statements was sufficient to engage the rlPFC region. Additionally, rlPFC was activated independently of belief–logic conflict, establishing belief–content conflict as an important factor in itself. When there is conflict (either belief–logic or belief–content) in reasoning problems, the same area of the brain is consistently activated.

It is possible that the right inferior frontal cortex is engaged during reasoning to resolve conflict by inhibiting competing information (in this case, interfering beliefs). It is important to note that studies of inhibitory control and conflict resolution both report activation in rlPFC (Aron et al., 2004; Maril et al., 2001; Milham et al., 2001). Thus, rlPFC activation during conditions of conflict may be due to the exertion of inhibitory control (i.e., to suppress interfering beliefs) (see DeNeys and Franssens, 2009; Goel, et al., 2000; Handley et al., 2004; Houdé, 1997, 2000; Markovits and Doyon, 2004; Markovits and Potvin, 2001; Moutier and Houdé, 2003; Moutier et al., 2006). This is because when reasoners are asked to “suppose the premises are true,” they must inhibit real-world knowledge and suspend their beliefs. Likewise, when judging the validity of a conclusion, reasoners must inhibit their prior knowledge to focus on the logic. Thus, both false statements (containing belief–content conflict) and belief–logic conflict require decontextualization – a separation between previous knowledge and the information held in working memory – and therefore inhibition, to complete the task. Thus, as the level of interference is increased, so would the need to exert inhibitory control. However, participants appear unable to completely overcome the conflicting information of their beliefs, as evidenced by slower reaction times for trials containing belief–logic and belief–content conflict. Further research is necessary to determine the extent to which inhibition is exclusively responsible for the activation observed in rlPFC under conditions of conflict.

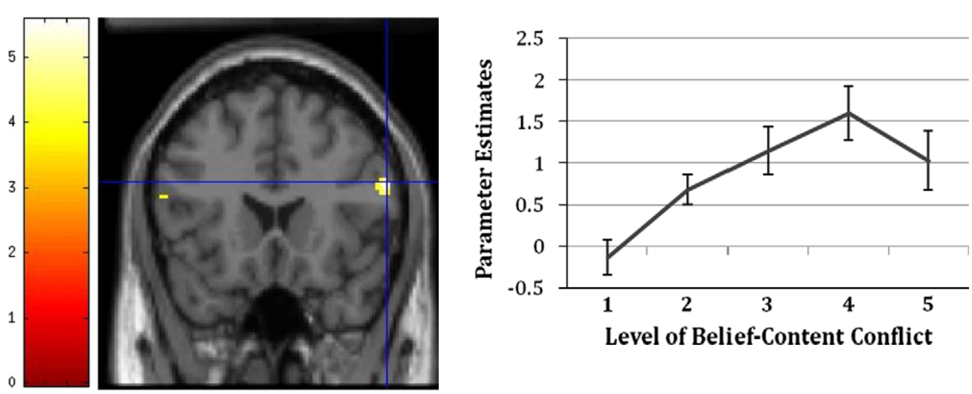
It is interesting to note the similarity of neural activation for belief–logic conflict between Canadian content and familiar-fictional content trials (see Fig. 2c and d). Although accuracy of knowledge about these two environments did not differ according to a post-scan questionnaire analysis, the familiar fictional content was only learned during a 1-month period before the scan, so the memory trace would not have had years of consolidation. Even though the beliefs about the familiar fictional environment would not be so deeply engrained as the Canadian content, it still activated the same area of rlPFC during belief–logic conflict, and almost to the same degree ($Z=2.96$ for familiar fictional versus $Z=3.12$ for Canadian content). Thus, even conflict in relation to recently acquired beliefs is sufficient to engage rlPFC during deductive reasoning.

In conclusion, results from the present study indicate that rlPFC is consistently engaged to resolve conflict in deductive reasoning. We demonstrated that multiple forms of conflict (belief–logic or belief–content conflict arising from statements about true-familiar or learned-fictional environments) contribute to the belief–bias effect. Although the present study focused on conditions of conflict stemming from content effects (beliefs) in deductive reasoning, this is not the only process that consistently engages rlPFC. Specifically, given this region’s involvement in



Note. a. Belief-logic conflict: incongruent > congruent (58, 14, 28, z = 3.31). b. Belief-logic conflict: incongruent > neutral (60, 20, 24, z = 4.02). c. Belief-logic conflict: Canadian geography only, incongruent > congruent (60, 20, 24, z = 3.12). d. Belief-logic conflict: familiar fictional environment only, incongruent > congruent (60, 12, 32, z = 2.96). SPMs rendered into standard stereotactic space and superimposed on to coronal MRI in standard space. Bars represent magnitude of z-score.

Fig. 2 – Belief–logic conflict activates same region within rlPFC.



Note. Belief-content conflict: Parametric modulation of rlPFC by level of interference (60, 20, 24, z = 4.04). SPM rendered into standard stereotactic space and superimposed on to coronal MRI in standard space (bar represents magnitude of z-score). Graph depicts parameter estimates extracted from the activated cluster in rlPFC (Level 1 = 3 believable sentences; Level 2 = belief-neutral sentences; Level 3 = 1 unbelievable sentence; Level 4 = 2 unbelievable sentences; Level 5 = 3 unbelievable sentences).

Fig. 3 – Parametric modulation of rlPFC by level of belief–content conflict.

uncertainty and indeterminacy (Goel et al., 2007; Goel et al., 2009) as well as simple motor or perceptual conflict, its involvement in conflict may be general rather than specific to deductive reasoning (Goel, 2009). As such, future research is needed to determine the specific features of conflict that activate this region consistently across different processes.

4. Experimental procedures

4.1. *Participants*

Sixteen participants (5 male; 11 female) with a mean age of 19.2 years (*SD*=2.3) and mean education of 13.8 years (*SD*=1.5) were recruited. All participants were normal healthy subjects with no history of neurological or psychiatric disorder. They were given course credit, monetary reimbursement and a picture of their brain for participation. All participants gave informed consent in accordance with the guidelines of the Human Participants Review Sub-Committee of York University.

4.2. *Experimental procedures*

Participants were instructed to learn one of two fictitious maps created by the first author, creating two conditions: a familiar fictional geography about which participants would have prior beliefs and an unfamiliar fictional geography about which participants would have no prior belief. On subsequent visits to the lab (each visit was 1 week apart), subjects were tested on the map both in written form (fill-in-the-blanks) and verbally (experimenter would ask questions such as, "Is Jaradon north of Veeron?"). If a participant demonstrated less than 100% accuracy, he or she was asked to return to the lab for a third test. Participants were also tested verbally on their knowledge of the geography of Canada until 100% accuracy on relevant facts was obtained. All participants met this criterion by the third test. On each visit, participants viewed and read aloud to the experimenter the words/places from the unfamiliar map, then repeated them silently to themselves several times. Therefore, although the words themselves from the unfamiliar map were not novel, geographical knowledge about their locations remained unfamiliar. On the second and third visits to the lab, participants were trained in logical reasoning and were given basic instructions on how to solve different types of three-term spatial relational problems. They were explicitly reminded that the task was to determine logical validity (valid or invalid) and to ignore content believability. Practice problems were given to participants as part of their training.

For data collection in the scanner, the 108 reasoning trials and 60 baseline trials were divided randomly into three runs of 56 trials, each run lasting 14 min. The following is an example of a reasoning trial:

Vancouver is west of Toronto.
Toronto is west of Halifax.
Vancouver is west of Halifax.

The task was to determine whether the third sentence, the conclusion, followed logically (i.e., necessarily) from the premises. Participants were specifically instructed that believability of content should be disregarded and only logical validity of the argument should be evaluated. Responses were made by pressing one of two buttons on a button box, corresponding with either 'valid' or 'invalid.' Baseline trials were created by pairing the first two sentences of one argument with the conclusion of a different argument in the study. Here is an example of a baseline trial:

Vancouver is west of Toronto.
Toronto is west of Halifax.
Montreal is north of St. Lawrence River.

For baseline trials, participants could disengage from the reasoning process upon the presentation of the unrelated conclusion and were instructed to press the button corresponding with "invalid." These baseline trials would control for neural activation associated with cognitive and motor components of no interest (e.g., reading and pressing a button).

Stimuli varied on two key dimensions: belief–logic conflict and belief–content conflict. First, reasoning problems are classified according to the presence or absence of belief–logic conflict. Trials containing no belief–logic conflict were either "congruent" or "neutral" trials. Congruent trials are those in which the conclusion believability and validity are in accordance (either valid-believable or invalid-unbelievable conclusions). For example, consider this valid-believable congruent problem: Vancouver is west of Toronto; Toronto is west of Montreal; Therefore, Vancouver is west of Montreal. This conclusion is logically valid (because the conclusion follows from the premises) and is also believable (because we know that Vancouver is actually west of Montreal), thus there is no conflict between conclusion validity nor believability ("belief–logic congruent"). Beliefs are actually facilitatory because they are in accordance with logic. Neutral trials (from the unfamiliar map condition) are belief-neutral and therefore contain no belief–logic conflict; that is, there can be no belief–logic conflict since by definition there are no beliefs that can conflict with logic. For example, consider the following logically valid, belief-neutral problem: Finn is north of Rheemore; Rostall is south of Rheemore; Therefore, Finn is north of Rostall. Participants who had not studied the map for this fictitious country would have no beliefs about whether Finn is north or south of Rostall, and therefore would have no beliefs as to its truth or falsity. The conclusion is valid, as it follows logically from the premises. Thus, there is no belief–logic conflict because there are no beliefs. [Half (8) of the participants would have studied this fictitious map and would therefore have beliefs about the sentences. The above logic problem would therefore be classified as either congruent or incongruent for these participants, depending on the believability and logical validity of the conclusion.] Both congruent and neutral conclusions contain no conflict between logic and beliefs. In contrast, incongruent trials are those in which the conclusion believability and validity are not in accordance (either valid-unbelievable or invalid-believable), thereby creating belief–logic conflict. For example, consider this valid-unbelievable incongruent problem: Canada is North of USA; Mexico is North of Canada. Therefore, Mexico is North of USA. This argument is logically valid (as the conclusion follows logically from the premises); however, the conclusion conflicts with our knowledge (as we do not believe Mexico is north of the United States), thereby creating a conflict between logic and beliefs—"belief–logic incongruent." Accuracy tends to decrease and reaction time tends to increase from congruent to neutral to incongruent trials (Gilinsky and Judd, 1994; Goel, et al., 2000; Goel and Dolan, 2003). There were an equal number of congruent, incongruent and neutral trials in the task (36 of each; see Table 1).

The second key variable on which problems varied was level of belief–content conflict (5 levels) (based on number of believable, unbelievable, or neutral sentences). Believable sentences from the Canadian geography or familiar-fictional (learned) map conditions (e.g., "Toronto is east of Vancouver") are in accordance with beliefs (no belief–content conflict), whereas unbelievable sentences (e.g., "Mexico is north of Canada") conflict with beliefs (thereby creating belief–content conflict). Neutral sentences (from the unfamiliar map condition) are belief-free (neither believable or unbelievable) and therefore do not cause belief–content conflict. If all three sentences in the argument (i.e., both premises and the conclusion) are believable or neutral, then no belief–content conflict would be present. The highest level of conflict would occur for arguments in which all three sentences were unbelievable, with problems containing one or two unbelievable sentences in between. Categorization of the levels of belief–content conflict was as follows: the lowest level (level 1) is an argument containing all believable sentences (i.e., 2 believable premises and a believable conclusion). Trials from the unfamiliar, belief-neutral condition contain no belief–content conflict (all sentences are neither believable nor unbelievable), and are thus classified as low conflict (level 2). Trials from levels 3, 4, and 5 contain one, two or three unbelievable sentences, respectively. Thus, trials varied in the level of interference caused by unbelievable sentences (with 3 unbelievable sentences as the highest level of interference). The number of trials in each belief–content conflict level was as follows: 18 "all believable" trials (level 1), 36 "neutral trials" (level 2) and 18 of each 1, 2, or 3 unbelievable sentences (levels 3, 4, and 5, respectively) (see Table 2). Thus, problems can vary in level of belief–content conflict, from all believable to neutral to 1, 2 or 3 unbelievable sentences.

Trials also varied in the familiarity of geographical content. The three conditions were: (1) familiar content (Canadian geography, 36 trials), (2) familiar fictional content (learned map, 36 trials), and (3) unfamiliar fictional content (unlearned map, 36 trials). The two fictional maps were counterbalanced so that 8 participants were trained on one map, and 8 on the other. Therefore, each participant learned one map, while the other remained unfamiliar (belief-neutral). Conditions 1 and 2, which were both from familiar geographical locations (Canada or learned fictitious country), were collapsed unless otherwise stated.

4.3. *fMRI data acquisition*

A 4-Tesla Oxford Magnet Technologies magnet with a Siemens Sonata gradient coil was used to acquire the data. Twenty-two $T2^*$-weighted interleaved multi-shot contiguous echo-planar images, 5 mm thick ($3\times3\times6$-mm voxels), were acquired axially, positioned to cover the whole brain. A total of 855 volume images were acquired over 3 runs (285 per run) with a repetition time (TR) of 3 s/volume.

Two subjects completed only one or two experimental runs due to time constraints. The other 14 participants completed all three runs in pseudo-random order. A 5-min anatomical scan was also collected from each participant. All participants were right-handed; however, half the participants were instructed to use their left hand as the response hand, while the other half used the right, to control for neural activation associated with motor control. Stimuli were presented randomly in an event-related design (see Fig. 1). Each trial began with a blank screen at 0 s, followed by the first sentence at 1 s, the second sentence at 4 s, and the conclusion at 7 s. Participants then had 8 s to respond by pressing a button box (half with the right hand, half with the left hand). The trials ended at 15 s. Data was sampled at the halfway point between the presentation of conclusion and the motor response (calculated for each individual trial for each participant).

Acknowledgments

This work was supported by training grants from the Natural Science and Engineering Research Council of Canada (NSERC) and Canadian Institutes for Health Research (CIHR) to M.S., and a McDonnell-Pew Program in Cognitive-Neuroscience Award, Premier's Research Excellence Award and an NSERC research grant to V.G.

REFERENCES

Aron, A.R., Robbins, T.W., Poldrack, R.A., 2004. Inhibition and the right inferior frontal cortex. Trends in Cognitive Science 8, 170–177.

Brett, M., Anton, J.-L., Valabregue, R., & J.-B., P. (2002). *Region of interest analysis using an SPM toolbox.*Presented at the 8th International Conference on Functional Mapping of the Human Brain.

DeNeys, W., Franssens, S., 2009. Belief inhibition during thinking: not always winning but at least taking part. Cognition 113, 45–61.

Evans, J.St.B.T., 1989. Bias in human reasoning. Lawrence Erlbaum, Hove, UK.

Evans, J.St.B.T., Barston, J.L., Pollard, P., 1983. On the conflict between logic and belief in syllogistic reasoning. Memory & Cognition 11, 295–306.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.-B., Frith, C.D., Frackowiak, R.S.J., 1995. Statistical parametric maps in functional imaging: a general approach. Human Brain Mapping 2, 189–210.

Gilinsky, A.S., Judd, B.B., 1994. Working memory and bias in reasoning across the life span. Psychol. Aging 9, 356–371.

Goel, V., 2009. Cognitive Neuroscience of Thinking. In: Berntson, G., Cacioppo, J.T. (Eds.), Handbook of Neuroscience for the Behavioral Sciences. Wiley, New York.

Goel, V., Stollstorff, M., Nakic, M., Knutson, K., Grafman, J., 2009. A role for right ventrolateral prefrontal cortex in reasoning about indeterminate relations. Neuropsychologia 47, 2790–2797.

Goel, V., Tierney, M., Sheesley, L., Bartolo, A., Vartanian, O., Grafman, J., 2007. Hemispheric specialization in human prefontal cortex for resolving certain and uncertain inferences. Cereb. Cortex 17, 2245–2250.

Goel, V., Buchel, C., Frith, C., Dolan, R., 2000. Dissociation of mechanisms underlying syllogistic reasoning. Neuroimage 12, 504–514.

Goel, V., Dolan, R., 2003. Explaining modulation of reasoning by belief. Cognition 87, 11–22.

Handley, S.J., Capon, A., Beveridge, M., Dennis, I., Evans, J.St.B.T., 2004. Working memory, inhibitory control and the development of children's reasoning. Thinking & Reasoning 10, 175–195.

Houdé, O., 1997. The problem of deductive competence and the inhibitory control of cognition. Current Psychology of Cognition 16, 108–113.

Houdé, O., 2000. Inhibition and cognitive development: object, number, categorization, and reasoning. Cognitive Development 15, 63–73.

Johnson-Laird, P.N., Bara, B.G., 1984. Syllogistic inference. Cognition 16, 1–61.

Luo, J., Yuan, J., Qiu, J., Zhang, Q., Zhong, J., Huai, Z., 2008. Neural correlates of the belief–bias effect in syllogistic reasoning: an event-related potential study. Neuroreport 19, 1073–1078.

Maril, A., Wagner, A.D., Schacter, D.L., 2001. On the tip of the tongue: an event-related fMRI study of semantic retrieval failure and cognitive conflict. Neuron 31, 653–660.

Markovits, H., Doyon, C., 2004. Information processing and reasoning with premises that are empirically false: interference, working memory, and processing speed. Memory, Cognition 32, 592–601.

Markovits, H., Potvin, F., 2001. Suppression of valid inferences and knowledge structures: the curious effect of producing alternative antecedents on reasoning with causal conditionals. Memory, Cognition 29, 736–744.

Milham, M.P., Banich, M.T., Webb, A., Barad, V., Cohen, N.J., Wszalek, T., Kramer, A.F., 2001. The relative involvement of anterior cingluate and prefrontal cortex in attentional control depends on nature of conflict. Brain Res. Cogn. Brain Res. 12, 467–473.

Moutier, S., Plagne-Cayeux, S., Melot, A.M., Houdé, O., 2006. Syllogistic reasoning and belief–bias inhibition in school children: evidence from a negative priming paradigm. Developmental Science 9, 166–172.

Moutier, S., Houdé, O., 2003. Judgement under uncertainty and conjunction fallacy inhibition training. Thinking and Reasoning 9, 185–201.

Oakhill, J., Johnson-Laird, P.N., 1985. The effects of belief on the spontaneous production of syllogistic conclusions. The Quarterly Journal of Experimental Psychology 37A, 553–569.

Tsujii, T., Masuda, S., Akiyama, T., Watanabe, S., 2010a. The role of inferior frontal cortex in belief–bias reasoning: an rTMS study. Neuropsychologia 48, 2005–2008.

Tsujii, T., Okada, M., Watanabe, S., 2010b. Effects of aging on hemispheric asymmetry in inferior frontal cortex activity during belief–bias syllogistic reasoning: a near-infrared spectroscopy study. Behav. Brain Res. 210, 178–183.

Tsujii, T., Watanabe, S., 2009. Neural correlates of dual-task effect on belief–bias syllogistic reasoning: a near-infrared spectroscopy study. Brain Res. 1287, 118–125.

Tsujii, T., Watanabe, S., 2010. Neural correlates of belief–bias reasoning under time pressure: a near-infrared spectroscopy study. Neuroimage 50, 1320–1326.

Worsley, K., Friston, K.J., 1995. Analysis of fMRI time-series revisited—again. Neuroimage 2, 173–181.