

Overcoming Perceptual Features in Logical Reasoning: A Parametric Functional Magnetic Resonance Imaging Study

Jérôme Prado^{1,2} and Ira A. Noveck^{1,2}

Abstract

■ Participants experience difficulty detecting that an item depicting an *H-in-a-square* confirms the logical rule, “If there is not a T then there is not a circle.” Indeed, there is a perceptual conflict between the items mentioned in the rule (*T* and *circle*) and in the test item (*H* and *square*). Much evidence supports the claim that correct responding depends on detecting and resolving such conflicts. One aim of this study is to find more precise neurological evidence in support of this claim by using a parametric event-related functional magnetic resonance imaging (fMRI) paradigm. We scanned 20 participants while they were required to judge whether or not a conditional rule was verified (or falsified) by a corresponding target item. We found that the right middorsolateral prefrontal cortex (mid-DLPFC) was specifically engaged, together with the medial frontal (anterior cingulate and presupplementary motor area [pre-SMA]) and parietal cortices, when mismatch-

ing was present. Activity in these regions was also linearly correlated with the level of mismatch between the rule and the test item. Furthermore, a psychophysiological interaction analysis revealed that activation of the mid-DLPFC, which increases as mismatching does, was accompanied by a decrease in functional integration with the bilateral primary visual cortex and an increase in functional integration with the right parietal cortex. This indicates a need to break away from perceptual cues in order to select an appropriate logical response. These findings strongly indicate that the regions involved in inhibitory control (including the right mid-DLPFC and the medial frontal cortex) are engaged when participants have to overcome perceptual mismatches in order to provide a logical response. These findings are also consistent with neuroimaging studies investigating the belief bias, where prior beliefs similarly interfere with logical reasoning. ■

INTRODUCTION

One of the older, continuing debates in the reasoning literature opposes researchers who highlight the relative ease with which participants arrive at rational, normative conclusions against those who underline participants' tendency to be seduced by nonrational biases. This debate is often evinced in research that focuses on *deductive reasoning*, that is, inference making that allows one to reach new valid conclusions on the basis of given information. Whereas many highlight, experimentally and otherwise, the human ability to carry out fundamental logical inferences (Braine & O'Brien, 1998; Rips, 1994; Sperber & Wilson, 1986), research has also demonstrated that deductive reasoning can be subject to biases (Evans, 1983, 1998).

Recently, *dual-process* models have been proposed that, in effect, resolve this debate by providing theoretical tools to explain how both normative and nonnormative

reasoning can occur within a single person (Stanovich, 2004; Evans, 2003; Sloman, 1996). Such theories posit that two separate cognitive systems underlie human rationality. The first one, named System 1 or the *heuristic system*, which is described as evolutionarily old, fast operating, automatic, and parallel, is generally assumed to be shared by humans and animals and to be at the root of many nonrational behaviors, two of which are described in detail below. Competing with System 1 is an evolutionary more recent system, System 2 or the *analytic system*, which is considered to be relatively slow operating, rule based, and sequential in nature. Although System 2 is constrained by working memory capacity, it permits abstract logical reasoning and hypothetical thinking and it is related to measures of general intelligence (Stanovich & West, 2000). A defining feature of any dual-process model of reasoning is that System 2 (i.e., the analytic system) should be able to inhibit and override System 1 (i.e., the heuristic system) so that people can successfully carry out logical tasks (Evans, 2003; Houdé & Tzourio-Mazoyer, 2003). Much of the empirical support for dual-process models comes from findings that show how

¹CNRS-Université de Lyon, France, ²CERMEP-Imagerie du vivant, France

normative reasoning is challenged by specific types of materials that trigger biases among reasoners. In what follows, we review behavioral and neurological findings related to two sorts of biasing materials.

How Perceptual Mismatches Negatively Affect Normative Reasoning

The first bias concerns the way people anticipate perceptual matches between features mentioned in a rule and those in a test item. That is, normative performance can be negatively affected when a correct answer actually requires one to deal with a mismatch or two. For example, when participants are presented a conditional rule such as “If there is not an H then there is not a 6” in what is known as the *truth table task*, they have relative difficulty detecting that the pair P4 indeed confirms the rule. Indications are that this difficulty is due to the fact that the two elements mentioned in the rule (the H and the 6) do not correspond with those in the provided pair (the P and the 4 in this case).¹ If the rule had been “If there is a P then there is a 4,” rates of correct performance are optimal. According to dual-process theorists such as Evans, errors in mismatching cases arise because nonmatches are seen as irrelevant to the rule, the upshot being that participants do not immediately attend to the logical demands of the task. Detecting that P4 properly confirms the conditional rule “If there is not an H then there is not a 6” implies that (i) a reasoner has to overcome an initial System 1 tendency, which is to see only the elements H and 6 as relevant, and that (ii) a reasoner has to engage System 2 processes in order to redirect attention toward other logically relevant items (i.e., letters other than H and numbers other than 6) in order to complete the task successfully (Oaksford & Stenning, 1992). The mismatching effect can be seen as “a System 1 heuristic which competes with logical (System 2) processes in determining choice” (Evans, 2003, p. 456).

The neural bases of the mismatching effect have been investigated to some extent. Houdé et al. (2000) recorded neural activity as participants essentially performed a specific trial drawn from one of the conditions of the truth table paradigm, namely, one in which participants are required to “falsify” a conditional rule such as “If there is not a square on the left then there is a triangle on the right.” In order to provide the correct response to this problem, the participant had to falsify this rule, which entails choosing among (12) designs in such a way that the pair has a true antecedent (a nonsquare on the left) and a false consequent (a nontriangle on the right). One can see how a correct response compels a participant to create a double mismatch (e.g., by choosing the elements *circle–diamond*).² That this particular trial from the truth table paradigm is exceptionally difficult is confirmed by data from standard evaluation tasks in which a participant has to decide

whether a provided exemplar is true or false with respect to the rule; the double mismatch case in the falsification condition consistently provides the lowest rates of correct responses (Prado & Noveck, 2006; Evans, 1972). Houdé et al. (2000) measured neural activity (using positron emission tomography [PET] methodology) and they were specifically interested in the role of training as participants aimed to avoid errors while searching for a pair of shapes.

Houdé et al. highlighted two features of their data. First, they reported that only inhibition training (one that aids participants to get past the perceptual pitfalls of the task) was efficient at reducing errors (as opposed to providing a logical explanation or by simply repeating the task), suggesting that giving an incorrect response in this task lies in a lack of inhibitory control. Second, using PET to compare brain activation patterns in trials before and after this successful training, the authors found that a bilateral posterior network (involving parietal and occipitotemporal regions) was involved during pretest (i.e., when subjects were influenced by a perceptual bias) and that a bilateral prefrontal network, involving the right ventromedial prefrontal cortex (VMPFC), was engaged during post test (i.e., when subjects were more successful at overcoming the effect) (Houdé et al., 2000, 2001). These findings show that several reasoning strategies can compete in the human brain (a perceptual strategy linked to visuospatial regions and a more analytic one associated with the PFC), providing some support for dual-process theories (Houdé, 2000).

Belief Biases and Deductive Reasoning

Another way that bias can challenge normative reasoning arises with considerations of personal beliefs. The *belief-bias* effect refers to the way one’s personal beliefs can provide competing information that is inconsistent with judgments of validity (Goel & Dolan, 2003; Goel, Buchel, Frith, & Dolan, 2000; Evans, 1983). The effect was first illustrated in behavioral studies using Aristotelian syllogisms over two decades ago (Evans, 1983), and its neural bases have been investigated more recently with fMRI (Goel & Dolan, 2003; Goel et al., 2000). We describe the phenomenon here.

Consider the two isomorphic syllogisms below, each of which is logically valid. The first (1) concludes with relatively neutral content, whereas the second (2) ultimately contains a conclusion that goes counter to participants’ personal beliefs:

- (1) No millionaires are hard workers.
Some rich people are hard workers.
∴ Some rich people are not millionaires.
- (2) No rich people are hard workers.
Some millionaires are hard workers.
∴ Some millionaires are not rich people.

Behavioral studies show that the conclusion in (1) is endorsed by a large majority (over 90%) of participants, whereas the conclusion in (2), whose content makes the conclusion appear false on its face, halves the endorsement rates (for classic studies, see Evans, 1983).

In terms of neural activity, Goel et al. (2000) have shown that making a logical inference in cases such as (1), where there is no belief-validity conflict, engages the left inferior frontal lobe and the left middle temporal lobe when compared to controls that do not require one to draw any inference. That the left inferior frontal lobe plays an important role in deductive reasoning has been confirmed with studies using neutral conditional (if-then) syllogisms (Noveck, Goel, & Smith, 2004). On the other hand, the neuroimaging data of Goel et al. (2000) indicate that a region located at the level of the right middorsolateral prefrontal cortex (mid-DLPFC) ($x = 53, y = 28, z = 23$ in Talairach coordinates; Brodmann's area [BA] 46/45) exhibits greater activity when participants respond correctly on tasks that require one to overcome biasing materials (such as those in (2)) than when there is no belief-logic conflict (such as (1)). The right inferior PFC has also been found to be more active when subjects successfully inhibit the prepotent response linked to beliefs (i.e., giving a correct response in syllogisms such as (2)) than when they fail to inhibit such trials (i.e., giving an incorrect response in syllogisms such as (2)) (Goel, 2003). Conversely, when subjects' responses are biased by their beliefs (incorrect inhibitory trials > correct inhibitory trials), the VMPFC appears to be specifically engaged, consistent with the idea that this region is engaged in intuitive response tasks (Elliott, Rees, & Dolan, 1999). As a whole, these belief-bias findings provide supporting evidence that the right mid-DLPFC is involved in addressing a conflict between two response tendencies in logical reasoning (Goel, 2005; Goel & Dolan, 2003; Goel et al., 2000). More specifically, the role of the right mid-DLPFC could be to inhibit a prepotent System 1 response (associated with prior knowledge in the case of the belief bias) in order to give a correct System 2 logical response (Evans, 2003). This finding is also highly consistent with the role attributed to the right lateral PFC in inhibitory control (Buchsbaum, Greer, Chang, & Berman, 2005; Aron, Robbins, & Poldrack, 2004; Bunge, Dudukovic, Thomason, Vaidya, & Gabrieli, 2002) and supports the dual-process claim that one of the key features of System 2 is to detect and resolve conflict that is brought forward by System 1.

Overall, the neuroimaging data indicate that the mismatching effects and the belief bias effects linked to System 1 (those that are not overridden by System 2) rely on separate neural networks (visuospatial areas for those who cannot get past the mismatching effect and VMPFC for those who fall prey to the belief bias). Are there more than just two systems then? Dual-process theorists argue that System 1 is not a singular system,

but rather a set of subsystems that operate with some autonomy (System 1 is also called the autonomous set of systems [TASS]; Stanovich, 2004). Whereas each subsystem in System 1 can be triggered by specific stimuli, each of these subsystems potentially competes with System 2. Crucially, dual-process models propose that the neural network that is engaged in overcoming any System 1 response (whether it be perceptual or based on beliefs) is unique. Specifically, the main cortical region that is engaged when overcoming belief biases (the right mid-DLPFC) ought to be engaged when overriding other sorts of biasing materials, for example, those related to perceptual mismatching.

Prior Approaches to Mismatching and the Current One

At this point, it is difficult to know whether this specific claim about the right mid-DLPFC can be supported. Although Houdé et al. (2001) found evidence revealing right prefrontal activation when participants successfully produce a correct (mismatching) pair of items, the area they reported (i.e., right VMPFC, BA 10) does not correspond with the lateral prefrontal region that Goel et al. (2000, 2003) describe as activated in their studies on the belief bias (i.e., mid-DLPFC, BA 46/45). One possible explanation for this discrepancy is that the findings of Goel et al. (2000, 2003) were reported with event-related fMRI, whereas those of Houdé et al. (2000, 2001) were reported in a block-design PET study that does not permit one to separate reading of the stimuli from the reasoning task. The discrepancy could also be due to the design used. Whereas Goel and Dolan's study (2000) directly compared syllogisms whose conclusions contain a (belief vs. logic) conflict to syllogisms whose conclusions do not, the design of Houdé et al. (2000, 2001), being focused on training, did not directly compare trials containing perceptual conflict(s) to those that do not.

The present study has three empirical objectives as it tests claims from the dual-processing accounts and while employing a truth table task. The first is to test the dual-process prediction that the cortical area involved in overriding the mismatching effect is the one reported with respect to overriding the belief bias. Given that the right mid-DLPFC appears to be engaged in overriding a prepotent response linked with beliefs, dual-processing theorists ought to predict that this region will also be involved in a truth table task when participants have to deal with mismatching cases. If a perceptual mismatch is indeed sufficient for producing activity in the right mid-DLPFC then even one mismatch between the features mentioned in the rule and the features in the test item ought to lead to such activity.

The second aim is to take full advantage of the truth table task paradigm by introducing a parametric approach that is sensitive to degrees of mismatch (see Prado & Noveck, 2006). That is, the truth table task can provide

test items that have no mismatches (with respect to the items mentioned in the rule), one mismatch, or two. The paradigm allows one to investigate cerebral activity linked to correct performance as the number of mismatches increases. This leads to our third aim.

According to dual-process accounts, part of the effort involved in overcoming a perceptual bias is to suppress System 1 activity. In the case of the truth table task, the structure that requires suppressing is the posterior perceptual network. Houdé et al. (2000) showed that such a perceptual network is engaged when participants give a nonlogical response in their task and, if dual processing accounts are correct, then one ought to find a suppression of activity there when participants provide a correct response in the face of mismatching information. To test this hypothesis, we will conduct a psychophysiological interaction (PPI) analysis (Friston et al., 1997) to reveal changes in the functional integration of the right mid-DLPFC as mismatches increase. We hypothesize that the right mid-DLPFC should display (1) decreased functional integration with regions involved in perceptual processing (due to the inhibition of the mismatching effect) and (2) increased functional integration with regions involved in response selection (of a logical response) when participants have to overcome the bias (namely, the posterior parietal cortex [PPC]; see Bunge, Hazeltine, Scanlon, Rosen, & Gabrieli, 2002).

To make the paradigm (and its parametric aspect) concrete, consider the three conditional rules below in a task that requires verification:

- (3) If there is an H then there is a square.
- (4) If there is an H then there is not a square.
- (5) If there is not an H then there is not a square.

If the rule in (3) were followed by an item depicting an *H-in-a-square*, verification would not entail mismatches at all and serves as an ideal baseline (0-mismatch condition). Now consider the rule in (4) followed by the item *H-in-a-circle*; verification here would require dealing with one mismatch (1-mismatch condition). Finally, consider the rule in (5) followed by the item *P-in-a-circle*; a participant would be required to deal with two mismatches (2-mismatch condition).³ Thus, assuming that mismatching is predictive of cortical activity, we should be able to reveal which regions are affected and in a linearly modulated fashion.

The paradigm readily allows for a more exhaustive study of perceptual mismatches in reasoning and in two ways. First, one can look at cases that do not depend on the interpretive role of negation. For example, consider the case where one ought to say “no” when the rule is (3) and the item is an *H-in-a-circle*. Here the perceptual mismatch is the basis for *correctly rejecting* the item with respect to the conditional rule and there is no negation involved. Second, the paradigm allows for two sorts of judgments: verification or falsification. The fal-

sification condition (run as a separate block) requires one to determine whether the test item falsifies the rule (e.g., *H-in-a-square* falsifies the rule in (4)). In the falsification task, a “yes” means that the rule was falsified. This condition is important because it amounts to a second, orthogonal parameter that further varies task difficulty in a systematic way. As Prado and Noveck (2006) have shown, the falsification task in the truth table paradigm prompts lower accuracy rates and slower reaction times than the verification task; nevertheless, mismatching continues to play an interfering role.

That this paradigm should be predictive of neuroimaging outcomes comes from a previously conducted behavioral study (Prado & Noveck, 2006). We showed that when the items mentioned in the rule mismatch those in the exemplar, rates of correct responses are generally lower and reaction times are generally slower. For example, when participants are required to verify a rule such as “If there is an H then there is not a square,” a test item such as an *H-in-a-square* (which provides no mismatches) leads to reliably higher rates of Correct Rejections than does an item such as *H-in-a-circle* (one mismatch) provide Hits. Likewise, it takes much longer to verify that *P-in-a-circle* verifies the rule “If there is not an H then there is not a square” than it does the rule “If there is a P then there is circle.” Falsification tasks provide very similar results while requiring participants to deal with an extra layer of complexity. The data from that study supported Evans’s (1998, 2003) claims that rates of errors increase along with mismatches due to the perceptual conflicts that arise between the elements mentioned in the rule and those in the test item.⁴

For the current work, we hypothesized which critical cortical regions ought to be involved in overcoming the mismatching effect in the truth table paradigm just described (with a particular focus on the right mid-DLPFC). In order to assess our hypothesis, we apply three criteria to determine whether these regions are in fact active when responding to trials using the truth table paradigm: (i) These regions should exhibit greater activity for correct mismatching trials than for correct nonmismatching trials (even while controlling for negations in the rules); (ii) cortical activity should increase as the number of mismatches increases in rules (i.e., a parametric modulation); and (iii) mismatch related activity should be evident whether the block requires one to verify or falsify the rule.

METHODS

Participants

Twenty healthy native French-speaking volunteers (7 men and 13 women, aged 19–26 years, mean 21.4 years) with no history of neurological or psychiatric disorders participated in the study. All subjects were right-handed as measured by the Edinburg Handedness Inventory

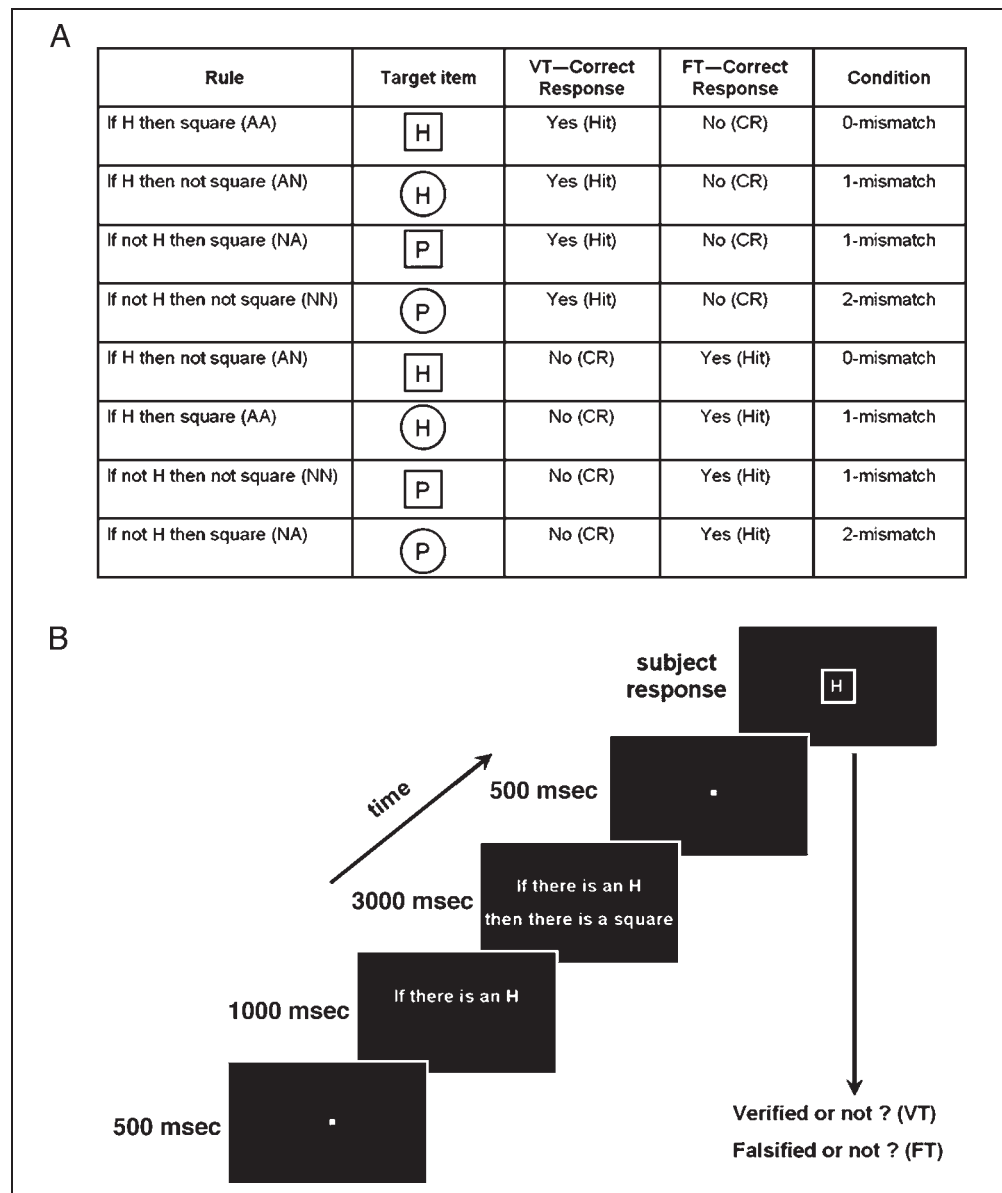
(Oldfield, 1971). They gave written informed consent and were paid for their participation. Procedures were approved by the local ethics committee (CCPPRB, Lyon, France).

Materials

A single trial was composed of a conditional if-then statement followed by a pictorial target item. The conditional rule described a letter and shape relation (e.g., “If there is an H then there is a square”) and the target item was based on a letter-in-shape combination, so that it consisted of a picture showing, for example, an H-in-a-square. This was designed to limit saccades. Trials were prepared based on the number of mismatching elements between the conditional rule and the target item (see Figure 1A). Hence, trials were separated into three

different conditions, 0-mismatch (i.e., no mismatch), 1-mismatch (i.e., moderate mismatch), and 2-mismatch (i.e., full mismatch). In the 0-mismatch condition, the pictorial item completely matched the letter and shape mentioned in the rule (e.g., to verify the rule “If there is an H then there is a square,” it would be followed by the target item H-in-a-square); in the 1-mismatch condition, only one element was present in the rule and the target item (e.g., to verify the rule “If there is not a J then there is a square,” it would be followed by a target item such as H-in-a-square); in the 2-mismatch condition, the pictorial item mismatched both the letter and shape mentioned in the rule (e.g., to verify the rule “If there is not a J then there is not a triangle,” it would be followed by a target item such as H-in-a-square). Four rules were used in the paradigm (based on the presence or absence of a negation in the antecedent of the rule and the

Figure 1. Stimuli and experimental procedure. (A) Overall experimental design. Participants were required to judge whether or not a conditional rule was verified (verification task, VT) or falsified (falsification task, FT) by a target item. Each trial was based on the presence or absence of mismatching elements between the rule and the target item, such that three main conditions were defined as: 0-mismatch, 1-mismatch, and 2-mismatch. Hit: Hit response; CR: Correct Rejection. (B) Timing of a sample trial. A trial started with the presentation of a visual central dot. The two parts of the conditional rule then appeared one line at a time and the entire rule remained on the screen until the target item appeared (after a briefly presented central dot). Subjects were then required to press one of the two response keys depending on whether the rule was verified or not (in the VT) or falsified or not (in the FT). The length of a trial varied randomly from 9250 msec to 13,350 msec.



presence or absence of a negation in the consequent of the rule). AA was affirmative throughout (e.g., “If there is an H then there is a circle”), AN presented a negation in the consequent of the rule (e.g., “If there is an H then there is not a square”), NA presented a negation in the antecedent of the rule (“If there is not a J then there is a circle”), and NN presented a negation in both the antecedent and consequent of the rule (e.g., “If there is not a J then there is not a square”).

We were concerned with those cases that yield unambiguous responses. Thus, 26 out of 30 trials per condition contained items that had a true antecedent. The remaining four stimuli per condition presented a letter that was irrelevant to the antecedent of the conditional rule; this leads to an nonobvious evaluation (consider the rule “If there is a J then there is a square” and the item H-in-a-square).⁵ These four were included to avoid predictability in the task and were considered as fillers. Half of the 26 relevant trials were confirming cases (where an item like H-in-a-square verifies the rule “If there is an H then there is a square”) and half disconfirming cases (where the item H-in-a-square ought to be rejected with respect to the rule “If there is an H then there is not a square”). The 36 target items were composed of one of six letters presented as a capital in bold (H, I, J, P, Q, and R) and one of six shapes (square, circle, star, diamond, rectangle, and triangle) in the central visual field.

Tasks and Procedure

Visual stimuli were generated with Presentation 9.20 software (Neurobehavioral Systems, www.neurobs.com) and projected onto a translucent screen with a Canon Xeed SX50 projector. The screen was viewed through a mirror. Each trial started with the presentation of a visual fixation mark (a central dot) in the center of the screen for 500 msec (see Figure 1B). The two parts of the conditional rule then appeared one line at a time, with the first part (e.g., “If there is an H”) appearing at 500 msec and the second part (“then there is a square”) at 1500 msec. The entire rule then remained on the screen for a further 3000 msec, at which point the rule disappeared and the central dot reappeared for 500 msec. This was immediately followed by the target item, which remained on the screen until subjects pressed one of two buttons on a keypad (yes/no response). Variable periods of visual fixations were added at the end of each trial to introduce jittering. That is, the duration of a trial varied randomly between 9250 and 13,350 msec (mean trial time = 11,300 msec).

Participants performed the experiment in four blocks (45 trial presentations in each) in order to complete both the verification task (VT) and the falsification task (FT). In two successive blocks, they had to perform the VT; that is, they had to judge whether or not the target item verified the rule presented. If the item confirmed

the rule, participants were required to press the “yes” key (i.e., “Yes, the rule is verified”); if the item did not confirm the rule, they were required to press the “no” key (i.e., “No, the rule is not verified”). In the two other successive blocks, participants performed the FT, which is different from the verification task in that participants had to (not verify but) falsify the rule. In other words, they had to determine whether or not the target item falsified the rule (by pressing the same yes/no response keys, as in the VT).

Task order was counterbalanced. Subjects were informed of the upcoming task only at the start of it. Trial order within each block was also randomized and the block order within each task was counterbalanced across participants. Participants were instructed to respond as quickly and as accurately as possible. Each task began with five training trials, which included four relevant (two confirming cases and two disconfirming cases) and one nonrelevant case. To summarize, the study used a within-subject 3×2 factorial design with the factors (i) Mismatch Level (0-mismatch, 1-mismatch, and 2-mismatch) and (ii) Task (VT, FT).

Imaging Procedures

Images were collected using the 1.5T MRI system (Siemens Sonata Maestro Class; Siemens, Erlangen, Germany) of the CERMEP Imagerie du vivant in Lyon. The fMRI blood oxygenation level dependent (BOLD) signal was measured using a T2*-weighted echo-planar sequence (repetition time [TR] = 2500 msec, flip angle = 90°, echo time [TE] = 60 msec). Twenty-six axial slices (4.40-mm thickness, field of view = 23 cm, 64×64 matrix) were acquired per volume. Following functional image acquisition, a high-resolution T1-weighted anatomical image (TR = 1880 msec, TE = 3.93 msec, FOV = 256 mm, flip angle = 15°, $176 \times 256 \times 256$ matrix, slice thickness = 1 mm) was collected for each subject.

fMRI Data Analysis

fMRI data were analyzed using SPM2 software (Wellcome Department of Cognitive Neurology, London, UK, www.fil.ion.ucl.ac.uk). Each block contained 216 functional volumes after rejecting the first four scans to eliminate nonequilibrium effects of magnetization. Functional images were corrected for slice acquisition delays and were spatially realigned to the first image of the first session on a voxel-by-voxel basis so as to correct for head movements. The realigned functional images and the anatomical scans for each subject were then normalized into a standard stereotaxic space by using the Montreal Neurological Institute (MNI) template. The functional images were spatially smoothed with an isotropic Gaussian filter (12-mm full width at half maximum). The event-related statistical analysis was performed according to the general linear model (Josephs,

Turner, & Friston, 1997) using the standard hemodynamic response function provided by SPM2. Events were time-locked to the appearance of the target item. The time series data were high-pass filtered (1/128 Hz) and serial correlations were corrected by an autoregressive AR(1) model.

Each activation event was categorized according to a 3 (mismatch level) \times 2 (task type: verification vs. falsification) factorial design. To this purpose we used a “parametric” approach to create functional maps of regions where the fMRI signal was positively or negatively modulated by the mismatch level (0-mismatch, 1-mismatch, 2-mismatch) as a function of the task (VT, FT). For each of the four blocks, the three conditions (three mismatch levels) were thus concatenated into one trial type and a parametric regressor was entered to test for a linear signal increase from 0-mismatch to 2-mismatch. Additionally, reaction times were added as covariates of no interest to remove effects of performance difficulty. Random effects analyses were applied to individual contrasts to account for between-subjects variance and to generalize to the population as a whole. The activations reported survived a voxel-level threshold of $p < .001$, uncorrected for multiple comparisons, and a cluster-level threshold of $p < .05$, corrected for multiple comparisons. The SPM2 coordinates were converted from MNI coordinate space into Talairach space (www.mrc-cbu.cam.ac.uk/Imaging/Common/mnispace.shtml) and localized using the Talairach atlas (Talairach & Tournoux, 1988).

Psychophysiological Interaction Analysis

PPI analysis is used to isolate brain areas (targets) showing an activity that can be explained in terms of an interaction between the influence of a distal area (source) and an experimental parameter (Friston et al., 1997). We apply this analysis to investigate the influence that our main a priori area, the right mid-DLPFC, could exert over other brain areas in relation to the mismatch level (i.e., a measure of effective connectivity). First, the deconvolved time course of activity was extracted in the right mid-DLPFC from the contrast showing linear parametric responses to increasing mismatch level in VT and in FT (an 8-mm-radius sphere centered at the peak of activity in the group analysis; $x = 45$, $y = 19$, $z = 27$). This activation time course constituted the first regressor in our PPI analysis. We then entered mismatch level (i.e., the psychological variable of interest) as the second regressor and calculated the product of time course with mismatch level to create the PPI term. The effect of the interaction term was investigated for each subject and each task and entered into a standard random effect group analysis at the second level (PPI maps were thresholded at an uncorrected voxel-level threshold of $p < .001$, and at a corrected cluster level threshold of $p < .05$).

RESULTS

The Results section is broken down into five parts. In the first part, we review the behavioral performance to verify that participants’ responses in the scanner are comparable to those found in standard experimental settings. In the second part, we take advantage of our 3 \times 2 parametric factorial design to highlight brain regions whose fMRI activity was linearly correlated with the increase of mismatches in rules in the VT and in the FT. In the third and fourth parts, we perform post hoc contrasts to verify which regions of the mismatch-related brain network highlighted above meet the three criteria defined in the Introduction. In the fifth part, we perform a PPI analysis aimed at investigating the functional interactions (or integrations) of the right mid-DLPFC when mismatch increases or decreases.

Overall Behavioral Performance

Overall behavioral performances (accuracy and mean reaction times) for each condition and task are shown in Table 1. Repeated measures analyses of variance (ANOVAs) with the factors Mismatch Level (0, 1, 2) and Task (VT, FT) were carried out on accuracy and response times data from correct trials. A main effect of Mismatch Level was found on both dependant variables: accuracy, $F(2,38) = 18.17$, $p < .001$; response times, $F(2,38) = 82.42$, $p < .001$. Responses were accurate and fast when the pictorial item completely matched with the letter and shape mentioned in the rule (0 mismatch; accuracy, 96%; mean response time, 1561 msec), they were less accurate and slower when there was one mismatch (1-mismatch; accuracy, 93%; mean response time, 1900 msec) and even less accurate and slower when there were two mismatches (2-mismatch; accuracy, 87%; mean response time, 2185 msec). That is, reaction times were linearly correlated with mismatch level in the VT as

Table 1. Overall Behavioral Performance

<i>Condition</i>	<i>Percentage Correct</i>	<i>Reaction Time (msec)</i>
<i>Verification task (VT)</i>		
0-Mismatch	97.12 \pm 0.73	1352.85 \pm 83.02
1-Mismatch	94.79 \pm 1.79	1671.60 \pm 97.24
2-Mismatch	93.65 \pm 2.11	1981.83 \pm 136.50
<i>Falsification task (FT)</i>		
0-Mismatch	94.98 \pm 1.05	1775.81 \pm 95.53
1-Mismatch	92.03 \pm 0.98	2174.02 \pm 101.90
2-Mismatch	80.66 \pm 3.20	2452.41 \pm 130.89

Values are mean accuracy and reaction times \pm SEM. Mean response times are based on correct responses only. SEM = standard error of the mean across participants ($n = 20$).

Table 2. Brain Areas Activated across the Different Contrasts

<i>Anatomical Location</i>	<i>~BA</i>	<i>No. of Voxels in Cluster</i>	<i>Voxel-level p Value (FDR Corrected)</i>	<i>Talairach Coordinates</i>			<i>Z Score</i>
				<i>x</i>	<i>y</i>	<i>z</i>	
<i>(2-Mismatch > 1-mismatch > 0-mismatch)_{VT}</i>							
R. middle/inferior frontal gyrus	46/45	2082	<.001	45	16	27	5.89
R. medial/superior frontal gyrus	6		<.001	9	11	49	4.88
R. inferior parietal lobule	39/40	726	<.001	30	-59	39	5.44
L. cuneus	17	1529	<.001	-9	-93	7	5.29
R. lingual gyrus	18		<.001	15	-85	-6	4.51
L. middle frontal gyrus	9/46	659	<.001	-50	28	35	4.94
L. inferior parietal lobule	40	384	<.001	-33	-50	41	4.73
R. middle/inferior frontal gyrus	46/45	2082	<.001	45	16	27	5.89
R. medial/superior frontal gyrus	6		<.001	9	11	49	4.88
R. inferior parietal lobule	39/40	726	<.001	30	-59	39	5.44
<i>(2-Mismatch > 1-mismatch > 0-mismatch)_{FT}</i>							
L. cuneus	17	1296	<.01	-9	-93	5	4.95
R. lingual gyrus	18		<.01	21	-85	-6	4.24
R. middle/inferior frontal gyrus	46/45	398	<.01	45	19	27	4.91
R. medial frontal gyrus	32	231	<.01	9	11	46	4.50
L. superior/middle frontal gyrus	6/7	158	<.001	-30	3	58	4.35
R. inferior parietal lobule	39/40	178	<.001	30	-56	39	4.19
<i>(1-Mismatch > 0-mismatch)_{AA}</i>							
L. middle frontal gyrus	46	573	<.01	-42	30	23	4.87
L. middle frontal gyrus	6		<.01	-39	8	52	3.98
L. middle frontal gyrus	6		<.01	-18	0	50	3.89
L. superior parietal lobule	7	589	<.01	-36	-62	50	4.68
L. inferior parietal lobule	40		<.01	-39	-33	46	4.44
L. inferior parietal lobule	39		<.01	-30	-59	36	3.96
R. middle frontal gyrus	9	492	<.01	53	19	32	4.64
R. precentral gyrus	6		<.01	48	2	36	4.13
R. middle frontal gyrus	46		<.01	48	19	21	3.97
R. medial frontal gyrus	32	349	<.01	12	14	44	4.61
L. medial frontal gyrus	8		<.01	0	20	46	4.12
R. medial frontal gyrus	8		<.05	9	31	40	3.72
R. superior parietal lobule	7	173	<.01	24	-56	39	4.22
R. superior parietal lobule	7		<.01	30	-59	44	4.07
R. middle frontal gyrus	9		<.05	30	-52	58	3.51
L. putamen		84	<.01	-21	23	-6	3.95

Table 2. (continued)

Anatomical Location	~BA	No. of Voxels in Cluster	Voxel-level <i>p</i> Value (FDR Corrected)	Talairach Coordinates			Z Score
				<i>x</i>	<i>y</i>	<i>z</i>	
<i>(2-Mismatch > 1-mismatch > 0-mismatch)_{VT, Hits only}</i>							
R. middle/inferior frontal gyrus	46/45	1728	<.001	42	16	21	5.40
R. medial frontal gyrus	6/32		<.01	6	17	46	4.62
R. middle frontal gyrus	10		<.01	39	52	-8	4.48
R. lingual gyrus	17	3222	<.001	21	-93	2	5.37
L. lingual gyrus	17		<.01	-24	-70	-3	4.71
L. cuneus	19		<.01	-3	-77	31	4.65
L. middle frontal gyrus	9	799	<.001	-42	28	26	5.10
L. middle frontal gyrus	8		<.001	-47	8	41	4.80
L. middle frontal gyrus	8		<.01	-42	22	40	4.41
L. inferior parietal lobule	40	777	<.01	-39	-53	52	4.65
L. inferior parietal lobule	40		<.01	-33	-54	45	4.41
L. precuneus	7		<.01	-3	-62	47	4.13
R. superior parietal lobule	7	387	<.01	33	-53	47	4.34
R. inferior parietal lobule	40		<.01	45	-41	42	3.85

L = left; R = right; ~BA = approximate Brodmann's area; FDR = false discovery rate; pre-SMA = presupplementary motor area; dACC = dorsal anterior cingulate cortex.

well as in the FT (VT: $r = .48$, $p < .001$; FT: $r = .50$, $p < .001$). A main effect of task was also found on accuracy and response times: accuracy, $F(1,19) = 11.48$, $p < .01$; response times, $F(1,19) = 20.40$, $p < .001$, showing that subjects were more accurate and responded more quickly for the VT than for the FT (accuracy, 95% vs. 89%; response times, 1662 vs. 2101 msec).

Overall Linear Effects of Increasing Mismatch in the VT and in the FT

fMRI data were entered in a random effects analysis using a 3×2 repeated measures ANOVA with the factors Mismatch Level (0-mismatch, 1-mismatch, 2-mismatch) and Task (VT, FT). That is, the four rules (AA, [AN, NA], and NN) and the correct two response types (Hits and Correct Rejections) were used to reveal brain regions that showed linear parametric BOLD responses to increasing mismatch in the VT and in the FT (Table 2 and Figure 2). In the VT, significant parametric activations to increasing mismatch ($[2\text{-mismatch} > 1\text{-mismatch} > 0\text{-mismatch}]_{\text{VT}}$) were found in the bilateral middle/inferior frontal gyrus (left: $x = -50$, $y = 28$, $z = 35$; $Z = 4.94$; right: $x = 45$, $y = 16$, $z = 27$; $Z = 5.89$), the dorsal anterior cingulate/presupplementary motor area (dACC/pre-SMA) ($x = 9$, $y = 11$, $z = 49$; $Z = 4.88$), the bilateral inferior parietal lobule (left: $x = -33$, $y = -50$, $z = 41$; $Z = 4.73$; right: $x = 30$, $y = -59$, $z = 39$; $Z = 5.44$), the

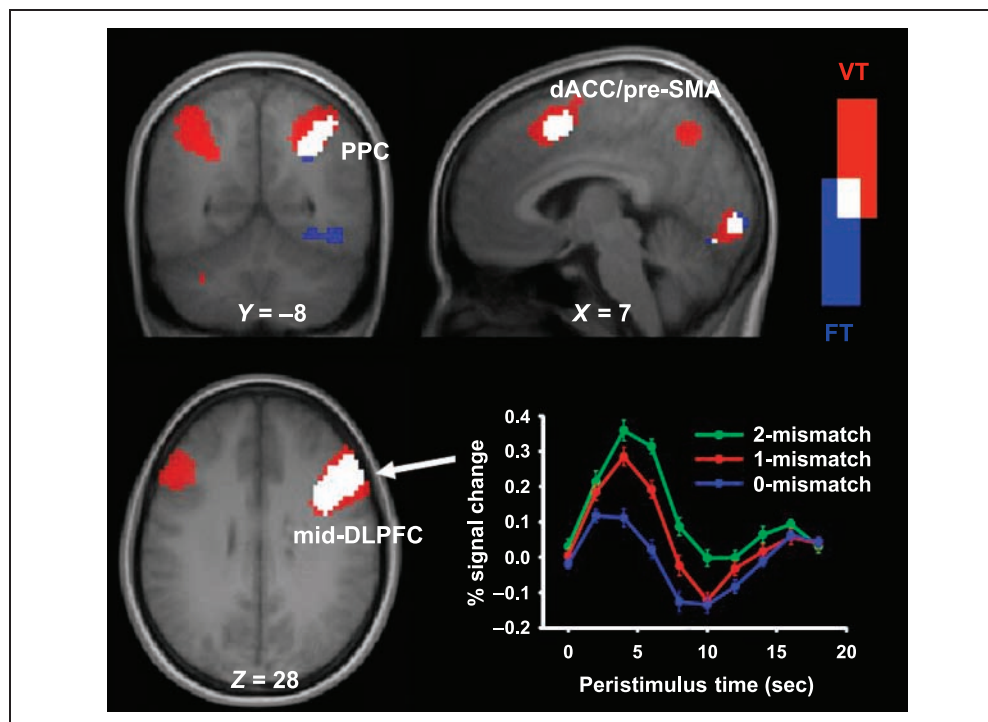
left cuneus ($x = -9$, $y = -93$, $z = 7$; $Z = 5.29$), and the right lingual gyrus ($x = 15$, $y = -85$, $z = -6$; $Z = 4.51$). In the FT, positive correlations ($[2\text{-mismatch} > 1\text{-mismatch} > 0\text{-mismatch}]_{\text{FT}}$) were observed in the right middle/inferior frontal gyrus ($x = 45$, $y = 19$, $z = 27$; $Z = 4.91$), the dACC/pre-SMA ($x = 9$, $y = 11$, $z = 46$; $Z = 4.50$), the right inferior parietal lobule ($x = 30$, $y = -56$, $z = 39$; $Z = 4.19$), the left superior/middle frontal gyrus ($x = -30$, $y = 3$, $z = 58$; $Z = 4.35$), the left cuneus ($x = -9$, $y = -93$, $z = 5$; $Z = 4.95$), and the right lingual gyrus ($x = 21$, $y = -85$, $z = -6$; $Z = 4.24$).

Consistent with the results above, a Boolean intersection of the linear main effects of mismatching in the VT and in the FT ($[2\text{-mismatch} > 1\text{-mismatch} > 0\text{-mismatch}]_{\text{VT}} \text{ AND } [2\text{-mismatch} > 1\text{-mismatch} > 0\text{-mismatch}]_{\text{FT}}$) revealed a right-lateralized cortical network. More specifically, this analysis revealed activations in the right mid-DLPFC ($x = 45$, $y = 19$, $z = 27$), in the dACC/pre-SMA ($x = 9$, $y = 11$, $z = 46$), and in the right PPC at the level of the inferior parietal lobule ($x = 30$, $y = -56$, $z = 39$).

Hits versus Correct Rejections: Effect of Mismatch Controlled for Effect of Negations

As can be seen in Figure 1A, an increase of mismatch in the previous sections could partially depend on a general increase of the amount of negations in the rules. That is, part of the effect (with respect to both the

Figure 2. Overall linear effects of increasing mismatch. (A) Brain regions showing linear parametric BOLD responses to increasing mismatch in the VT ([2-mismatch > 1-mismatch > 0-mismatch]_{VT}; red), in the FT ([2-mismatch > 1-mismatch > 0-mismatch]_{FT}; blue) and in both tasks ([2-mismatch > 1-mismatch > 0-mismatch]_{VT} AND [2-mismatch > 1-mismatch > 0-mismatch]_{FT}; white) when considering the whole paradigm (i.e., four rules and two response types). SPMs are superimposed on to axial, coronal and sagittal slices of the mean anatomical brain of the 20 participants (voxel level $p < .001$, cluster level $p < .05$ corrected). The graph shows the group-averaged event-related responses to the three mismatch levels in the right mid-DLPFC. mid-DLPFC: midsolateral prefrontal cortex; PPC: posterior parietal cortex; dACC/pre-SMA: dorsal anterior cingulate/presupplementary motor area.



behavioral and imaging data) could be due to the extra processing necessitated by negations. To control for this potential confound, we now turn to effects of mismatching when the mismatches are not linked to the inclusion of negations in the rules, that is, in comparing Hits versus Correct Rejections across the same, single (AA) rule (see Introduction).

We focus on the single rule where there are no negations (e.g., “If there is an H then there is a square” [AA]). When one considers the entire paradigm (VT, FT, Hits, and Correct Rejections), the mismatch effect on these rules (i.e., contrast [1-mismatch > 0-mismatch]_{AA}) depends on the type of correct response (i.e., Hit or Correct Rejection) as well as the task (VT vs. FT) but not on the number of negations (see Figure 3A). In other words, single mismatch cases are compared to no-mismatch cases. To illustrate with respect to the rule “If there is an H then there is a square,” the single mismatch in the item H-in-a-circle typically prompts a false response in the VT (a Correct Rejection) and a true response in the FT (a Hit), whereas the no-mismatch item H-in-a-square typically prompts a true response in the VT (a Hit) and a false response in the FT (a Correct Rejection). Any activated regions found in the contrast [1-mismatch > 0-mismatch]_{AA} cannot be explained in terms of extra processing of negations nor in terms of a search for contrasting items (other letters or other shapes). These led to two analyses.

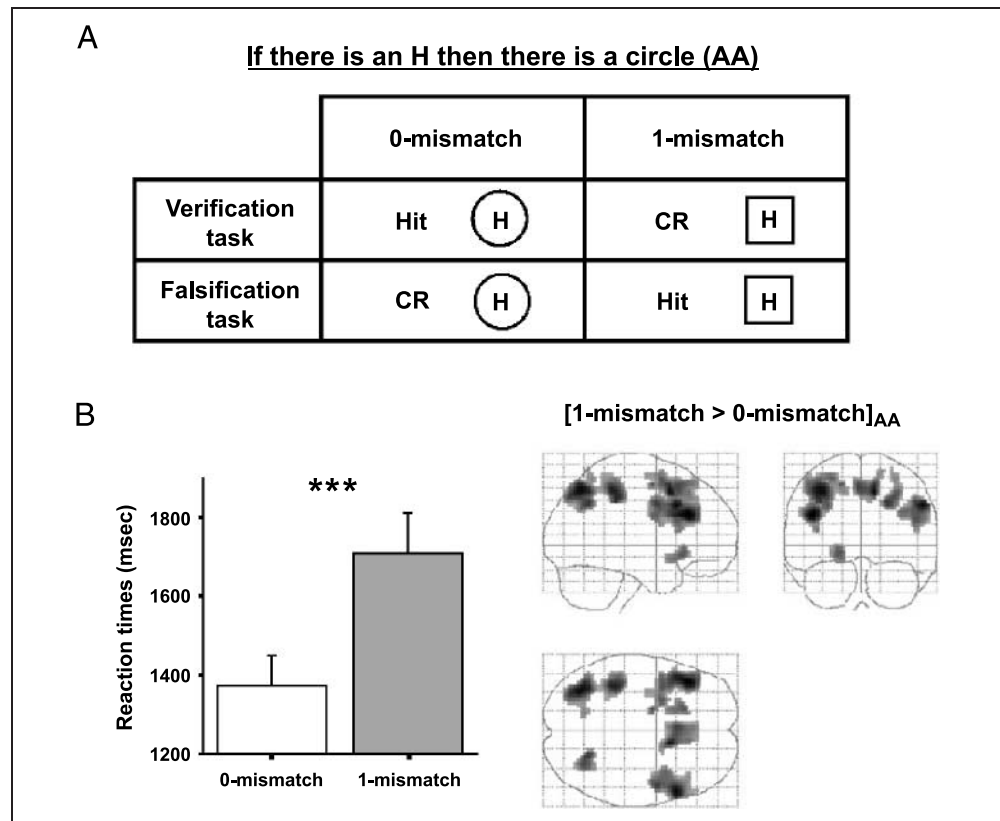
First, behavioral responses were significantly faster for 0-mismatch items than for 1-mismatch items, 1374 vs. 1711 msec; $t(19) = 6.22$ $p < .001$; two-tailed paired t test. Second, giving a correct response in these mismatching cases ([1-mismatch > 0-mismatch]_{AA}) activated a network similar to that found in the previous section (i.e., when mismatching was due to the role of negations) (see Table 2 and Figure 3B). That is, this contrast revealed activation of the bilateral middle/inferior frontal gyrus (left: $x = -42$, $y = 30$, $z = 23$; $Z = 4.87$; right: $x = 48$, $y = 19$, $z = 21$; $Z = 3.97$), the dACC/pre-SMA ($x = 12$, $y = 14$, $z = 44$; $Z = 4.61$), and the bilateral PPC (left: $x = -36$, $y = -62$, $z = 50$; $Z = 4.68$; right: $x = 24$, $y = -56$, $z = 39$; $Z = 4.22$).

Within these brain regions, we then searched for voxels also active in the overall main effects of mismatch in the VT *and* in the FT (see previous section). Only three regions survive this stringent analysis (i.e., [1-mismatch > 0-mismatch]_{AA} AND [2-mismatch > 1-mismatch > 0-mismatch]_{VT} AND [2-mismatch > 1-mismatch > 0-mismatch]_{FT}): the right mid-DLPFC, the dACC/pre-SMA, and the right PPC (see Figure 4A).

Linear Effects of Increasing Number of Negations (and Mismatches)

As seen above, the mismatching effect persists even when disentangled from potential effects of negation. However,

Figure 3. Hits vs. Correct Rejections: effect of mismatch controlled for effect of negations. (A) Sample trials used in the analysis of the contrast $[1\text{-mismatch} > 0\text{-mismatch}]_{AA}$. (B) Left: Response times data. Participants were faster to judge a 0-mismatch trial than a 1-mismatch trial (independently of negations): $t(19) = 7.02, p < .001$. Right: fMRI data. Activated brain network in the contrast $[1\text{-mismatch} > 0\text{-mismatch}]_{AA}$ shown on a glass brain (voxel level $p < .001$, cluster level $p < .05$ corrected).



it also pays to investigate how negation affects activation as it interacts with mismatching in the rules because correct responding depends on properly interpreting negations (e.g., participants need to recognize that *not-H* allows for a *P*, etc.). This is why we investigate the *Hit* responses in the Verification task (see Figure 1A) (i.e., contrast $[2\text{-mismatch} > 1\text{-mismatch} > 0\text{-mismatch}]_{VT, Hits}$). Results are reported in Table 2 and Figure 4B.

First, a repeated measures ANOVA with the factor Mismatch Level (0, 1, 2) on response times data from correct trials revealed a significant main effect of negations/mismatch number, $F(2,38) = 74.25, p < .001$. Reaction times were also linearly correlated with the number of negations/mismatches ($r = .60, p < .001$).

Second, significant parametric activations to increasing negations/mismatches in the $[2\text{-mismatch} > 1\text{-mismatch} > 0\text{-mismatch}]_{VT, Hits}$ contrast were found in a large bilateral network including the bilateral lateral PFC (with a peak of activity at the level of the mid-DLPFC: left: $x = -42, y = 28, z = 26; Z = 5.10$; right: $x = 42, y = 16, z = 21; Z = 5.40$), the right anterior PFC (aPFC) ($x = 39, y = 52, z = -8; Z = 4.48$), the dACC/pre-SMA ($x = 6, y = 17, z = 46; Z = 4.62$), the PPC (left: $x = -39, y = -53, z = 52; Z = 4.65$; right: $x = 33, y = -53, z = 47; Z = 4.34$), and the bilateral lingual gyrus (left: $x = -24, y = -70, z = -3; Z = 4.71$; right: $x = 21, y = -93, z = 2; Z = 5.37$). That is, we found basically the same neural network as above with the notable exception that, in addition, the right aPFC appeared to be specifically engaged here.

Psychophysiological Interaction Analysis

As shown in Figure 5, the PPI analysis revealed that increase of activity in the right mid-DLPFC (using as seed an 8-mm-radius sphere centered at the coordinates $x = 45, y = 19, z = 27$) when mismatch increased (across the whole paradigm, VT and FT included) was accompanied by decreased functional interaction with bilateral primary visual cortex (left cuneus, $x = -12, y = -84, z = 4; Z = 4.48$; right lingual gyrus, $x = 15, y = -82, z = 2; Z = 4.37$) and increased functional interaction with the right PPC (inferior parietal lobule, $x = 50, y = -27, z = 46; Z = 3.84$).

DISCUSSION

Performance with the truth table task investigated here requires participants to determine whether or not a conditional rule was verified or falsified by a pictorial target item while the level of mismatch between the target item and the elements contained in the rule were manipulated in a parametric way (via three levels). Our theoretical analysis was guided by the a priori prediction coming from dual-process accounts of reasoning that propose that cognitive inhibition mechanisms in general—whether they are linked with mismatching or prior beliefs—should emerge from the same neural sites. That is, overcoming a System 1 response based on perceptual processes should yield activity in the same cortical regions

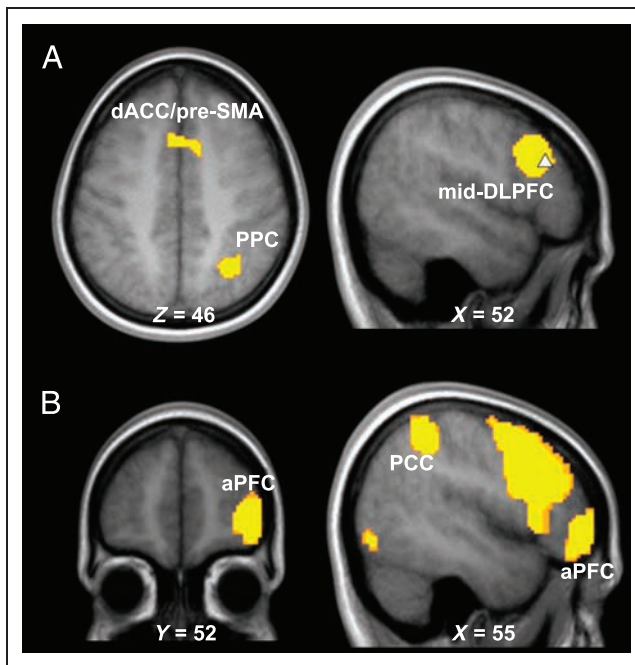


Figure 4. Brain networks involved in the inhibition of the mismatching effect and in the processing of negations. (A) Brain regions showing both greater activity for correct mismatching trials than for correct nonmismatching trials, and a linear correlation with the level of mismatch in the VT as well in the FT. Negations are not predictive of activity in this network (see Methods and Results). White triangle: approximate location of the main peak of activity reported in Goel et al.'s study on the inhibition of the belief bias for the contrast [correct inhibitory trials > correct neutral trials] (2000). (B) Linear effects of increasing number of negations (and mismatches) in the VT (Hits responses only). This analysis reveals specific activation of the right anterior cortex (aPFC).

as overcoming a System 1 response associated with prior knowledge. Although mismatching and belief bias effects are not likely to share the same subsystems in System 1, the conflicts they engender are likely to activate the same singular System 2. Given that Goel and colleagues found activation mainly in the right lateral PFC (i.e., right mid-DLPFC) during inhibition of the belief bias (Goel & Dolan,

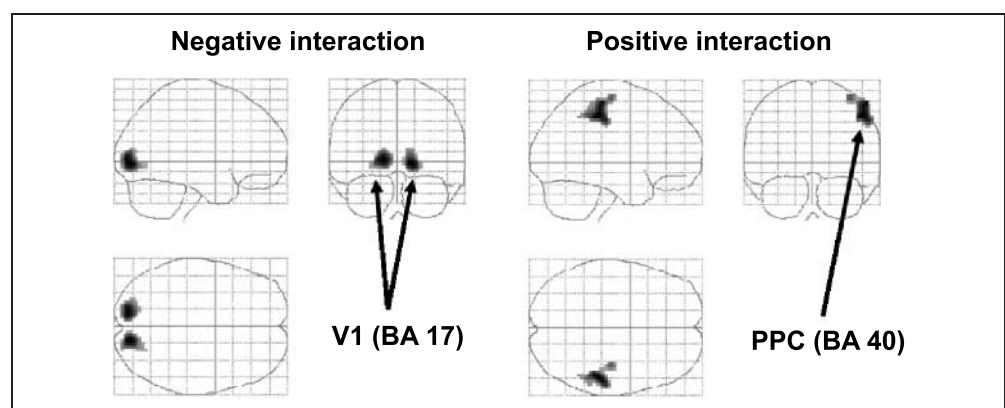
2003; Goel et al., 2000), dual-process accounts predict the involvement of the right mid-DLPFC for effects due to perceptual mismatching.

The behavioral data confirmed that mismatches interfered with correct performance, as much of the prior literature has demonstrated (Evans, 1998). Confirming Prado and Noveck (2006), participants made significantly more errors and their reaction times were significantly slower when elements in the test item increasingly mismatched with respect to the elements in the rule (see Table 1). This demonstrates that performance in the fMRI scanner is robust.

More central to our current interests, the fMRI study was designed to reveal the cortical network engaged in a deductive task when it contends with such perceptual mismatches. For the purposes of our study, we argued that any area that is proposed as being involved in the inhibition of the mismatching effect ought to fulfill three requirements. First, such a region should be most active when subjects have to deal with mismatches (and this activity should be independent of the presence of negations in rules). Second, the magnitude of the activity should be positively correlated with the number of mismatches. Third, this mismatch-related activity should be observed in the VT as well as in the FT (i.e., regardless of the specific task in which the item is tested).

To test our predictions, we first searched for brain regions exhibiting a parametric linear response to increasing mismatch (when using the whole paradigm that allows three levels of mismatching in the VT and in the FT: 0-mismatch, 1-mismatch and 2-mismatch). In the VT, we showed that a bilateral frontoparietal network, including the bilateral lateral PFC (mid-DLPFC), the bilateral PPC, and the dACC/pre-SMA is involved when participants reason with these mismatching cases. However, only the right part of this network was engaged when computing the main effect of mismatch in the FT. More precisely, the main cluster of activation in the two tasks was located at the level of the right mid-DLPFC ($x = 42, y = 16, z = 21$; BA 9/46), the main a priori region predicted by dual-process theories of reasoning (see Figure 2).

Figure 5. Brain regions showing negative (left) and positive (right) functional interaction with the right mid-DLPFC when mismatching increases (PPI analysis). Random effect analysis, voxel level $p < .001$, cluster level $p < .05$ corrected.



One concern is that mismatches often overlap with the number of negations in the rules (see Figure 1A). Our effort to control for this potential confound, in effect, further tests our initial hypothesis. We performed an additional analysis using the most basic rule, AA. Here we analyzed brain regions responsive to perceptual mismatch when the mismatching did not depend on the inclusion of negations; that is, we investigated Hit responses and Correct Rejections across the VT and in the FT. This involved contrasting cases containing a single mismatch only to those having no mismatches (regardless of the sort of response given). The effect of mismatch here ($[1\text{-mismatch} > 0\text{-mismatch}]_{AA}$) yields activations in a quite similar bilateral network involving the bilateral PPC, the bilateral mid-DLPFC, and the dACC/pre-SMA (see Figure 3).

Although activity in the right mid-DLPFC is shown to appear as a result of mismatching independently of the presence of negations, the word *not* does have a role to play in these tasks. When we analyzed the VT as negations are parametrically manipulated (see examples 3 through 5), we found activity, not only in the bilateral frontoparietal network reported in the above analyses but in the right aPFC as well, a region not typically shown to be engaged in studies involved with overriding a prepotent response (as the mismatching effect is supposed to yield) (see Figure 4B).⁶ Moreover, this region was not found to be involved when mismatching was independent of negations, even at a lenient threshold ($p < .05$ uncorrected, voxelwise). We thus hypothesize that activity in this region is explained specifically by the increase of the number of negations in rules. Although the aPFC has been found to be engaged in a variety of tasks involving episodic memory, working memory, or task switching, it has been proposed recently that this region is required when subjects need to integrate the outcomes of two or more cognitive processes (Ramnani & Owen, 2004). Specifically, Koechlin, Basso, Pietrini, Panzer, and Grafman, 1999 have demonstrated that the aPFC is involved in cognitive branching (i.e., the ability to keep in mind a main goal while performing concurrent subgoals). Given that the word *not* arguably compels one to consider alternatives from the relevant contrast class (e.g., not-H can prompt a search for other letters; Oaksford & Stenning, 1992), reasoning on a conditional rule that contains negations might involve a certain amount of cognitive branching. That a search among members in the contrast class should arise when reasoning from a negated element was described in Prado and Noveck (2006).

To summarize, we found only three mismatch-related regions that satisfied the three requirements we imposed in the Introduction: the right mid-DLPFC, the right PPC, and the dACC/pre-SMA (see Figure 4A). In this network, mismatch-related activity is (i) evident although not depending on the interpretative role of negations, (ii) parametrically modulated by the level of mismatch,

and (iii) present in the verification as well as in the falsification task.

This cortical network is highly similar to the network classically engaged in attentional control (i.e., the cognitive process that is involved in overriding a rare or highly salient interference). Using perceptual interference tasks like the Stroop paradigm (MacLeod & MacDonald, 2000), most neuroimaging studies have systematically demonstrated the engagement of a cerebral network composed of the medial frontal cortex (dACC and pre-SMA), the lateral PFC, and the PPC in attentional control (Kerns et al., 2004; Milham et al., 2001; Banich et al., 2000a). Although the respective roles of these regions are still under debate, it has been argued that the dACC, together with the pre-SMA, detect the occurrence of a conflict (Rushworth, Walton, Kennerley, & Bannerman, 2004; Botvinick, Braver, Barch, Carter, & Cohen, 2001) and in response recruit the DLPFC, which would be essential for resolving this conflict in a top-down manner in the PPC by imposing an attentional “set” (Bunge, Hazeltine, et al., 2002; Corbetta & Shulman, 2002; Banich et al., 2000b).

The present findings are also consistent with previous studies investigating the belief bias described earlier, which similarly interferes with logical reasoning but in ways that have little to do with perceptual features. To highlight the role of the right mid-DLPFC, consider again the two isomorphic syllogisms presented in the Introduction, one that presents a valid conclusion that is consistent with beliefs (1) and one that is inconsistent (2) with beliefs. While investigating the neural bases of this belief bias effect, Goel et al. (2000) have demonstrated that the right mid-DLPFC (along with the medial prefrontal cortex) is activated more in conflicting conditions than congruent ones. Moreover, this region also appears to be engaged when participants apparently inhibit the tendency to endorse a believable conclusion in order to provide a validity judgement (Goel & Dolan, 2003). Studying the belief bias effect and using a cognitive subtraction similar to that used in the present study (correct belief trials > correct neutral trials), Goel et al. found activity in a region located around Talairach coordinates $x = 53, y = 28, z = 23$ (BA 46/45, mid-DLPFC). In the present study, as shown in Figure 4A, the coordinates were nearly identical, $x = 45, y = 19, z = 27$ (BA 9/46), and yet the bias was introduced by interference that does not arguably require high-level concerns like beliefs.

It has been recently suggested that the right lateral PFC could play a crucial role in cognitive control and response inhibition (Aron et al., 2004). For example, it has been shown that a lesion at the level of the right inferior frontal cortex disrupts response inhibition in a stop-signal task (Aron, Fletcher, Bullmore, Sahakian, & Robbins, 2003), and this region was shown to be activated in many neuroimaging studies involving inhibitory control (Hampshire & Owen, 2006; Bunge, Dudukovic, et al., 2002; Garavan, Ross, & Stein, 1999; Konishi et al., 1999). Thus, the right

lateral PFC could resolve conflicts between concurrent response tendencies by modulating activity of posterior regions involved in response selection.

To assess the hypothesis that the right mid-DLPFC interacts with posterior regions when participants have to override interference from mismatching, a PPI analysis was performed to examine the functional integration of the right mid-DLPFC when mismatching increases (using as seed an 8-mm-radius sphere centered at the peak of activity in the present study: $x = 45$, $y = 19$, $z = 27$). We found two interactions, one positive and another negative (see Figure 5). The positive interaction that satisfied our criteria ($p < .001$, voxelwise; $p < .05$, corrected clusterwise) was in the right PPC ($x = 50$, $y = -27$, $z = 46$, $Z = 3.84$). Interestingly, it has been proposed that the PPC houses the representation of possible responses invoked by the environment (Bunge, Hazeltine, et al., 2002). As far as the present study is concerned, we speculate that the right mid-DLPFC is recruited when one needs to override interference caused by mismatching elements (in order to inhibit the tendency to see them as irrelevant to the rule). The right mid-DLPFC could then bias the PPC in amplifying cerebral responses to task-relevant information (i.e., seeing the relevance of the mismatching cases), a role that has been already attributed to the DLPFC in a recent neuroimaging study investigating attentional control in a Stroop task (Egner & Hirsch, 2005).

The other finding from the PPI analysis was that increases of activity in the right mid-DLPFC as mismatches increase were accompanied by decreased functional interaction with bilateral primary visual cortex ($x = 15$, $y = -82$, $z = 2$, $Z = 4.48$; $x = -12$, $y = -84$, $z = 4$, $Z = 4.37$). This result directly favors the view that the right mid-DLPFC plays a central role in overriding, in this case, a perceptual interference. Such a negative interaction between this region and the primary visual cortex when mismatching increases can be interpreted as a need to overcome the perceptual cues (i.e., the elements in the rule) in order to give a correct logical response. This finding is in line with the data reported by Houdé and Tzourio-Mazoyer (2003), who measured cerebral activity of participants while they were performing the most difficult falsification trial in the truth table task (a total mismatch in the Falsification Task) (Houdé et al., 2000). They found a striking shift from a posterior network to a prefrontal network after bias-inhibitory training on this task, a result highly consistent with the decreased functional interaction observed in the present study between the prefrontal and the occipital cortex as mismatching increases. Furthermore, they found that the mismatching effect elicited activity in perceptual and visuospatial regions when subjects gave a heuristic response. Here we showed that the right mid-DLPFC is active when participants have to override this heuristic response, whereas its effective connectivity with two perceptual regions decreased, strongly suggesting a role

of the right mid-DLPFC in overcoming a perceptual response yielded by a visuospatial brain network.

Overall, the findings here are consistent with dual-process accounts, which suggest that two cognitive systems compete for control of the response in reasoning tasks, a nonrational system, System 1, associated with, among other things, beliefs and perceptual processes, and an analytic system, System 2, capable of logical abstraction and hypothetical thinking (Stanovich, 2004; Evans, 2003; Sloman, 1996). The brain system composed of the right mid-DLPFC, the dACC/pre-SMA, and the right PPC observed here—when subjects overcome the perceptual bias—is basically the same as the one involved when participants have to inhibit the belief bias effect (Goel & Dolan, 2003; Goel et al., 2000) and is in line with the view that two behavioral response tendencies (a rational and a perceptual one) can compete with respect to the mismatching effect. That is, this brain network arguably reflects System 2's inhibition of the heuristic system (System 1). Our findings also suggest that, even if System 1 is likely to rely on a set of separate subsystems (with beliefs and mismatching underlying two distinct cerebral correlates; see Goel & Dolan, 2003; Houdé et al., 2000), each subsystem could be overridden nevertheless by a singular System 2, which principally depends on the inhibitory role of the right mid-DLPFC. This hypothesis is supported by an important line of research that emphasizes the role of the right lateral PFC in attentional control (see above).

More generally, the present work's contribution to the evolving field of neuroimaging of reasoning is threefold. First, to our knowledge, this is the first study to show that, when participants have to overcome a perceptual bias in a *logical reasoning* task, a right-lateralized system involving the mid-DLPFC, the PPC, and the dACC/pre-SMA is recruited. Second, the findings indicate that this particular truth table paradigm can be useful for studying this network because its parametric character allows for varying levels of perceptual interference. Third, these findings are consistent with those from previous studies showing that such a network is activated when participants have to overcome a prepotent response associated with *prior knowledge* in logical reasoning (Goel, 2005) as well as in other tasks involving response competition (Buchsbaum et al., 2005; Aron et al., 2004). Overall, these results support dual-mechanism accounts of reasoning that postulate that a single analytic reasoning system inhibits a more heuristic one. Indeed, a perceptual mismatch between features mentioned in a rule and those in a test item can yield a conflict that needs to be resolved in order to carry out logical reasoning.

Acknowledgments

We thank Jean-Baptiste Van der Henst, Mathilde Bonnefond, Coralie Chevallier, and the reviewers for helpful comments on the manuscript; Danielle Ibarrolla, Christine Vighi and

Fabienne Poncet for assistance with data collection; and Christian Scheiber for having been our project's principal medical researcher. This work was supported by Emergence, a Rhône-Alpes Developmental grant, awarded to both authors.

Reprint requests should be sent to Jerome Prado, Laboratoire sur le Langage, le Cerveau et la Cognition, UMR 5230, Institut des Sciences Cognitives, 67 Boulevard Pinel, 69675 Bron cedex, France, or via e-mail: jprado@isc.cnrs.fr.

Notes

1. Based on related findings on other reasoning tasks, errors are said to result from a *matching bias*. However, we do not use this nomenclature here because, as one will see, the “bias” can also be viewed as salutary, rendering this expression confusing (see Prado & Noveck, 2006). To keep the focus on the difficulties rendered by mismatching cases in the truth table task, we will refer to *mismatching effects* here.
2. Note how a falsification task is easier when the rule is “If there is a square on the left then there is not a triangle on the right.” In this case, a correct response entails no mismatches (square–triangle).
3. The examples in (3) through (5) are given for illustrative purposes; the experiment also includes a fourth type of rule that has a negated antecedent and an affirmative consequent. Although there are four types of rules, note that there are only three levels of mismatching.
4. In that study we compared Evans's account to one known as the *contrast-class* account, which assumes that error rates are due to the computational role of negations that could prime appropriate alternatives. Findings from our two experiments supported Evans's account (Prado & Noveck, 2006).
5. According to standard logic truth tables, conditionals containing a false antecedent are always true, but participants in reasoning experiments typically consider these ambiguous or confusing.
6. We thank an anonymous reviewer for pointing out that this ROI ought to be considered independently of those considered relevant to dual-process accounts.

REFERENCES

- Aron, A. R., Fletcher, P. C., Bullmore, E. T., Sahakian, B. J., & Robbins, T. W. (2003). Stop-signal inhibition disrupted by damage to right inferior frontal gyrus in humans. *Nature Neuroscience*, *6*, 115–116.
- Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences*, *8*, 170–177.
- Banich, M. T., Milham, M. P., Atchley, R., Cohen, N. J., Webb, A., Wszalek, T., et al. (2000a). fMRI studies of Stroop tasks reveal unique roles of anterior and posterior brain systems in attentional selection. *Journal of Cognitive Neuroscience*, *12*, 988–1000.
- Banich, M. T., Milham, M. P., Atchley, R. A., Cohen, N. J., Webb, A., Wszalek, T., et al. (2000b). Prefrontal regions play a predominant role in imposing an attentional “set”: Evidence from fMRI. *Brain Research, Cognitive Brain Research*, *10*, 1–9.
- Botvinick, M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*, 624–652.
- Braine, M. D. S., & O'Brien, D. P. (1998). *Mental logic*. Hillsdale, NJ: Erlbaum.
- Buchsbaum, B. R., Greer, S., Chang, W. L., & Berman, K. F. (2005). Meta-analysis of neuroimaging studies of the Wisconsin card-sorting task and component processes. *Human Brain Mapping*, *25*, 35–45.
- Bunge, S. A., Dudukovic, N. M., Thomason, M. E., Vaidya, C. J., & Gabrieli, J. D. (2002). Immature frontal lobe contributions to cognitive control in children: Evidence from fMRI. *Neuron*, *33*, 301–311.
- Bunge, S. A., Hazeltine, E., Scanlon, M. D., Rosen, A. C., & Gabrieli, J. D. (2002). Dissociable contributions of prefrontal and parietal cortices to response selection. *Neuroimage*, *17*, 1562–1571.
- Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*, 201–215.
- Egner, T., & Hirsch, J. (2005). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nature Neuroscience*, *8*, 1784–1790.
- Elliott, R., Rees, G., & Dolan, R. J. (1999). Ventromedial prefrontal cortex mediates guessing. *Neuropsychologia*, *37*, 403–411.
- Evans, J. S. B. T. (1972). Interpretation and matching bias in a reasoning task. *British Journal of Psychology*, *24*, 193–199.
- Evans, J. S. B. T. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory and Cognition*, *11*, 295–306.
- Evans, J. S. B. T. (1998). Matching bias in conditional reasoning: Do we understand it after 25 years? *Thinking and Reasoning*, *4*, 45–110.
- Evans, J. S. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*, 454–459.
- Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage*, *6*, 218–229.
- Garavan, H., Ross, T. J., & Stein, E. A. (1999). Right hemispheric dominance of inhibitory control: An event-related functional MRI study. *Proceedings of the National Academy of Sciences, U.S.A.*, *96*, 8301–8306.
- Goel, V. (2005). Cognitive neuroscience of deductive reasoning. In K. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 475–492). Cambridge: Cambridge University Press.
- Goel, V., Buchel, C., Frith, C., & Dolan, R. J. (2000). Dissociation of mechanisms underlying syllogistic reasoning. *Neuroimage*, *12*, 504–514.
- Goel, V., & Dolan, R. J. (2003). Explaining modulation of reasoning by belief. *Cognition*, *87*, B11–B22.
- Hampshire, A., & Owen, A. M. (2006). Fractionating attentional control using event-related fMRI. *Cerebral Cortex*, *16*, 1679–1689.
- Houdé, O. (2000). Inhibition and cognitive development: Object, number, categorization, and reasoning. *Cognitive Development*, *15*, 63–73.
- Houdé, O., & Tzourio-Mazoyer, N. (2003). Neural foundations of logical and mathematical cognition. *Nature Reviews Neuroscience*, *4*, 507–514.
- Houdé, O., Zago, L., Crivello, F., Moutier, S., Pineau, A., Mazoyer, B., et al. (2001). Access to deductive logic depends on a right ventromedial prefrontal area devoted to emotion and feeling: Evidence from a training paradigm. *Neuroimage*, *14*, 1486–1492.
- Houdé, O., Zago, L., Mellet, E., Moutier, S., Pineau, A., Mazoyer, B., et al. (2000). Shifting from the perceptual brain to the logical brain: The neural impact of cognitive inhibition training. *Journal of Cognitive Neuroscience*, *12*, 721–728.
- Josephs, O., Turner, R., & Friston, K. (1997). Event-related fMRI. *Human Brain Mapping*, *5*, 217–327.

- Kerns, J. G., Cohen, J. D., MacDonald, A. W., III, Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science*, *303*, 1023–1026.
- Koechlin, E., Basso, G., Pietrini, P., Panzer, S., & Grafman, J. (1999). The role of the anterior prefrontal cortex in human cognition. *Nature*, *399*, 148–151.
- Konishi, S., Nakajima, K., Uchida, I., Kikyo, H., Kameyama, M., & Miyashita, Y. (1999). Common inhibitory mechanism in human inferior prefrontal cortex revealed by event-related functional MRI. *Brain*, *122*, 981–991.
- MacLeod, C. M., & MacDonald, P. A. (2000). Interdimensional interference in the Stroop effect: Uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences*, *4*, 383–391.
- Milham, M. P., Banich, M. T., Webb, A., Barad, V., Cohen, N. J., Wszalek, T., et al. (2001). The relative involvement of anterior cingulate and prefrontal cortex in attentional control depends on nature of conflict. *Brain Research, Cognitive Brain Research*, *12*, 467–473.
- Noveck, I. A., Goel, V., & Smith, K. W. (2004). The neural basis of conditional reasoning with arbitrary content. *Cortex*, *40*, 613–622.
- Oaksford, M. R., & Stenning, K. (1992). Reasoning with conditionals containing negated constituents. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 835–851.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97–113.
- Prado, J., & Noveck, I. A. (2006). How reaction times elucidate the matching bias and the way negations are processed. *Thinking and Reasoning*, *12*, 309–328.
- Ramnani, N., & Owen, A. M. (2004). Anterior prefrontal cortex: Insights into function from anatomy and neuroimaging. *Nature Reviews Neuroscience*, *5*, 184–194.
- Rips, L. (1994). *The psychology of proof*. Cambridge: MIT Press.
- Rushworth, M. F., Walton, M. E., Kennerley, S. W., & Bannerman, D. M. (2004). Action sets and decisions in the medial frontal cortex. *Trends in Cognitive Sciences*, *8*, 410–417.
- Soloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22.
- Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Oxford: Blackwell.
- Stanovich, K. E. (2004). *The robot's rebellion: Finding meaning in the age of Darwin*. Chicago: Chicago University Press.
- Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*, 645–665.
- Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*. New York: Thieme.