Overcoming perceptual features in logical reasoning: An event-related potentials study

Jérôme Prado*, Maria Kaliuzhna, Anne Cheylus, Ira A. Noveck

Laboratoire sur le Langage, le Cerveau et la Cognition (L2C2), Université de Lyon, CNRS UMR 5230, 67 Boulevard Pinel, 69675 Bron, France

ABSTRACT

It is more difficult for reasoners to detect that the letter–number pair H7 verifies the conditional rule *If there is not a T then there is not a 4* than to detect that it verifies the rule *If there is an H then there is a 7*. In prior work [Prado, J., & Noveck, I. A. (2007). Overcoming perceptual features in logical reasoning: a parametric functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience 19*(4), 642–657], we argued that this difficulty was due to *mismatching* effects, i.e. perceptual mismatches that arise when the items mentioned in the rule (e.g. T and 4) mismatch those presented in the test-pair (H and 7). The present study aimed to test this claim directly by recording ERPs while participants evaluated conditional rules in the presence or absence of mismatches. We found that mismatches, not only trigger a frontocentral N2 (an ERP known to be related to perceptual mismatch) but that they, parametrically modulate its amplitude (e.g. two mismatches prompt a greater N2 than one). Our results indicate that the main role of negations in conditional rules is to focus attention on the negated constituent but also suggest that there is some inter-individual differences in the way participants apprehend such negations, as indicated by a correlation between N2 amplitude and participants' reaction times. Overall, these findings emphasize how overcoming perceptual features plays a role in the mismatching effect and extend the mismatch-related effects of the N2 into a reasoning task.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

Deductive reasoning is an inference-making process that allows one to reach valid conclusions on the basis of prior information. While deductive inference-making has often been demonstrated to be fundamental to human cognition (Braine & O'Brien, 1998; Johnson-Laird & Byrne, 1991; Rips, 1994), it has also been shown that people fall prey to biases that are often inconsistent with normative theories of logic (Evans, 1989; Stanovich, 2004). For example, a large body of evidence in the reasoning literature shows that participants have difficulty disconfirming hypotheses (Nickerson, 1998), and are highly influenced by the semantic (Evans, 1983) content of logical arguments. How to interpret this discrepancy between the observed behavior of reasoners and the normative standards of logic has been the subject of a long-standing debate, some researchers arguing that the gap was reflecting irrationalities of thinking (Evans, 1984, 2003; Stanovich, 2004), while others emphasizing the fact that it was the manifestation of performance errors and computational limitations (Oaksford & Chater, 1995, 1996).

In conditional reasoning, it has been demonstrated that normative performances can be negatively influenced by perceptual mismatches between features mentioned in a rule and a test-item (Evans, 1972). In order to illustrate this bias, consider the two conditional rules below:

(1) If there is a P then there is a 4.
(2) If there is not a T then there is not a 7.

Both of the above rules are verified by the letter–number pair P4. However, it has been shown that participants experience more difficulty detecting that P4 confirms the rule in (2) than it confirms the rule in (1) (Evans, 1998). This effect is believed to occur because, for the rule in (2), the lexical content of the rule (the letter T and the number 7) does not match the letter–number pair verifying it (the letter P and the number 4). This effect has been termed *matching bias* (Evans, 1998), which underlines how matching cases encourage quicker responses, or *mismatching effects* (Prado & Noveck, 2007), which underlines how mismatches lead to errors and slowdowns.

* Corresponding author at: Department of Psychology, University of Michigan, 530 Church Street, Ann Arbor, MI 48109, United States. Tel.: +1 734 763 3321.
E-mail address: jprado@umich.edu (J. Prado).

As it is the case for other reasoning biases, there is a debate about the rational or irrational nature of the mismatching effects. One school of thought posits that the effect is due to a low-level *heuristic* that compels reasoners to consider relevant only features mentioned in the rule (i.e. T7 in the rule in (2)); the result is that they do not immediately attend to the logical demands of the task (i.e. detecting that P4 confirms the rule in (2) (Evans, 1998). Detecting that P4 confirms the rule in (1) does not involve such mismatches and thus prompts optimal performance.

Given the important role of negations in this type of task it is important to point out that this approach defends the idea that the word *not*, at its core, simply denies. We refer to this as the *narrow-scope* view. More concretely, when being informed that *There is not an T*, this says only what the letter in the item is *not*; the sentence is not necessarily suggesting a search among potential alternatives (this view represents a competing account that we describe immediately below). With a narrow-scope view, a second piece of information indicating, e.g., that the item has an *H* is viewed as a source of mismatch (between the two highlighted letters) rather than being the source of finding a logical compatibility with *not a T*. Given that this theory considers these mismatching effects as resulting from a heuristic process, we will refer to this account as the *heuristic* account throughout the present paper.

The alternative account defends a rational explanation of this phenomenon in arguing that participants' occasional failures with such (complex) sentences are due to computational errors linked to the processing of negations (Oaksford & Stenning, 1992). In this case, the word *not* compels one to consider alternatives from the relevant contrast-class (e.g. other letters; other numbers) and to then confirm that the elements in the test-item are among the alternatives. In other words, this account (which we will refer to as the *contrast-class* account) argues that a negation leads a listener to consider the many things that the denied proposition can be. When considering the proposition *There is not a T*, a reasoner has to imagine another possible letter (i.e. the phrase not T prompts a search for non-T letters). According to this account, the arrival of a second piece of information containing an acceptable alternative (e.g. where the letter is an H) should appear logically compatible and prompt ready confirmation. If a reasoner has to contend with two negations, as in *If there is not a T then there is not a square*, the number of contrast classes increases and so does the amount of needed calculation (and potential for error).

To date, behavioral studies investigating mismatching effects in conditional reasoning have provided mixed evidence regarding the heuristic or rational nature of the bias. On the one hand, Oaksford and Stenning (1992) have found that the mismatching effects receded when the construction of the contrast classes was facilitated, providing evidence for a rational explanation of the bias. For example, they showed that participants experience less difficulty dealing with mismatches when they reasoned with realistic conditional rules (e.g. If the boss doesn't want to see me, then I'll be home in time for dinner), or when the size of the contrast-class is relatively small (Oaksford & Stenning, 1992). More recently, Oaksford and Moussakowski (2004) directly contrasted predictions from the heuristic and the contrast-class accounts within the same experiment. The authors manipulated independently both mismatches and set sizes in conditional statements and found that reasoners' behavior was more driven by the size of the contrast set than by the presence of mismatches, providing additional evidence for the contrast-class account.

On the other hand, examination of participants' reaction times during a classic conditional reasoning task gives support for the heuristic account of the mismatching effects. Indeed, using a variation of the truth-table task, we found the mismatching effects to be systematic on conditional rules with respect to both reaction times and rates of correct responses, as subjects were required to either verify or falsify conditional rules (Prado & Noveck, 2006). That is, rates of correct responses were generally lower and reaction times were generally slower when the features mentioned in the rule mismatch those in the test-item. The effect is also present even in the absence of negations in rules. For example, it takes a significantly longer time to determine that the item *H-in-a-circle* disconfirms the rule *If there is an H then there is a square* (1-mismatch) than it does to verify the rule *If there is an H then there is circle* (0-mismatch). Our study also provided evidence in favor of a narrow-scope view of negations. Indeed, we showed that 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* (0-mismatch) leads to reliably higher rates of Correct Rejections than does an item such as *H-in-a-circle* (1-mismatch) provides Hits. Such a result argues against the idea that participants construct contrast classes when faced with negations in conditional rule. Instead, they more likely focus their attention on the negated constituent (*square* in the example above) and have to overcome perceptual mismatches when they are confronted to another piece of information (*circle* in the example above). These results are in-line with the heuristic account of the mismatching effects.

Recently, we investigated the neural bases of the mismatching effects using fMRI (Prado & Noveck, 2007). We demonstrated that a network composed of the medial prefrontal cortex (dorsal anterior cingulate [dorsal ACC] and pre-supplementary motor area [pre-SMA]), the mid dorso-lateral prefrontal cortex (mid-DLPFC) and the posterior parietal cortex (PPC) was more engaged when participants reasoned in the context of a mismatching rule than when the rule did not entail any mismatch. Interestingly, several studies have suggested that this frontoparietal network is associated with attentional orienting (Corbetta & Shulman, 2002) and is involved in selecting between task-relevant and task-irrelevant information, thus allowing goal-directed behavior. For example, these regions are found to be activated when participants are asked to ignore task-irrelevant information, like the meaning of the word in the Stroop paradigm or an infrequent stimulus in an oddball task (Milham, Banich, & Barad, 2003). In our fMRI study, we found that (i) activity in this network was parametrically modulated by the amount of perceptual mismatch between items in the rule and in the target item and (ii) was independent of the amount of negations in the rule. That is, reasoning on the rule *If there is an H then there is not a square* yielded more activation in this cerebral system when participants had to provide a Hit (1-mismatch; *H-in-a-circle*) than a CR (0-mismatch; *H-in-a-square*). Such a result is consistent with the idea that participants consider mismatching features in conditional rules as logically irrelevant, and need to overcome this tendency to provide a logical (task-relevant) response in this task (Evans, 2003).

Although both the behavioral and the fMRI findings described above suggest that negations in conditional rules are processed narrowly, arguments in favor of such a claim would be more convincing if one could provide additional, independent evidence. This is why we turn to investigations that concern an ERP component that has been reliably associated with mismatch: the frontocentral N2. As recently reviewed by (Folstein & Van Petten, 2008), there is long-standing literature suggesting that the frontocentral N2 (an ERP component typically elicited 200–400 ms following the presentation of a specific stimulus) reflects the mismatch between two stimuli. Such an effect has been demonstrated by Gehring, Gratton, Coles, and Donchin (1992), using a modified version of the Eriksen flanker task (Eriksen & Eriksen, 1974). In the typical flanker task, a central target stimulus is presented simultaneously with two surrounding flankers, and the subjects are asked to respond based on the target and to ignore the flankers. The flankers can indicate the same response as the target (e.g. SSSSS) or a different response

Table 1
Truth values and Mismatch levels for the rules AA and AN

Rule	Hit	Correct Rejection
If H then square (AA)	$H\text{-}in\text{-}square_0$	$H\text{-}in\text{-}circle_1$
If H then not circle (AN)	$H\text{-}in\text{-}square_1$	$H\text{-}in\text{-}circle_0$

Notes. Subscript refers to level of mismatching.

(e.g. SSHSS). Taking advantage of a cued version of this flanker task, Gehring et al. found that the amplitude of the N2 time-locked to the target array was parametrically modulated by the amount of mismatch between the cue and the target. For example, the N2 was greater when the target SSSSS was preceded by the cue H than by the cue S, or when the target HHSHH was preceded by the cue S than by the cue H.

More recently, Wang et al. (2003) presented participants with pairs of sequential geometrical figures differing or not in term of perceptual features. They found a greater N2 when the second stimulus mismatched the first one, as compared to when there was a perfect match. A subsequent study from the same team (Wang, Cui, Wang, Tian, & Zhang, 2004) showed that the amplitude of this mismatch-related N2 was also dependent on the number of mismatching features (with greater amplitude for maximal mismatch). Comparable results have been found in several other studies investigating perceptual changes in physical attributes as diverse as shape (Wang, Wang, Kong, Cui, & Tian, 2001), color (Wang et al., 2001), or spatial position (Yang & Wang, 2002).

The frontocentral N2 is also a component that is reliably elicited in paradigms involving a certain degree of stimulus novelty, like in oddball tasks. In an oddball task, subjects are presented with a mixture of frequent and typically unique target or non-target stimuli. Although participants' task is to detect only targets, a greater N2 is elicited when a novel (i.e. typically unique) stimulus appears (for a review, see Folstein & Van Petten, 2008). The amplitude of this 'novelty' N2 does not depend on the fact that the stimulus is a target or not (Courchesne, Hillyard, & Galambos, 1975), but is significantly influenced by its physical complexity (Courchesne et al., 1975; Daffner et al., 2000). This suggests that it may reflect the deviance from a mental template stored in short-term or long-term memory (Daffner et al., 2000). As noted by Folstein and Van Petten (2008), such an interpretation is consistent with results from sequential matching tasks discussed above and raises the hypothesis that this component is involved in more general mismatch detection mechanisms.

The mismatch sensitivity of the frontocentral N2 makes it a perfect candidate to shed light on the mismatching effects in conditional reasoning, as well as on the way negations are processed in such statements. As in the prior work, we employ a paradigm known as the *truth-table verification task* (Evans, 1972) in which subjects are presented with conditional rules similar to the one presented with respect to (1) and (2) in order to judge whether a given target item is verified or not by the rule. For example, see Table 1 and consider the two conditional rules below:

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

Both of the above rules are verified by the item *H-in-a-square* (i.e. Hit response) and falsified by the item *H-in-a-circle* (i.e. Correct Rejection). However, providing a Hit response in each rule involves dealing with different amounts of mismatch: a Hit response would not entail mismatch at all for the rule in (3) (0-mismatch), but it would require dealing with one mismatch for the rule in (4) (1-mismatch). On the contrary, correctly rejecting the rule in (3) would imply dealing with more mismatching information (1-mismatch)

than correctly rejecting the rule in (4) (0-mismatch). Based on our previous studies suggesting that the perceptual mismatches are on the basis of the mismatching effects, we predict a greater mismatch-related N2 for CRs than Hits for the rule in (3) (i.e. greater mismatch for 1-mismatch that for 0-mismatch), but a greater N2 for Hits than CRs for the rule in (4) (i.e. greater mismatch for 1-mismatch that for 0-mismatch). That is, one should show a crossover interaction between the Rule and the Type of response.

The same reasoning ought to apply to the rules (5) and (6) below, which differ from (3) and (4) in that they have negations in the antecedent as well. These, too, are verified by the item *H-in-a-square* (Hit) and falsified by the item *H-in-a-circle* (CR):

(5) If there is not a J then there is a square.
(6) If there is not a J then there is not a circle.

In the rule in (5), one would predict greater mismatch (thus greater N2) for CRs (2-mismatch) than Hits (1-mismatch). The reverse pattern should be observed for the rule in (6) (greater N2 for CRs than Hits). In other words, a crossover interaction between the Rule and the Type of Response should also be apparent on these rules (5) and (6).

Finally, we consider individual differences. In prior work (Prado & Noveck, 2006), we proposed that the two (narrow and contrast-class) accounts are compatible. We hypothesized that the initial reading of a negation will be narrow and in some scenarios this might be enough. However, the interlocutor (a participant) can further process the negation, leading to an analysis of alternatives; thus it could pay off to anticipate (that *not a square* can ready one for a circle). This led us to propose a potential two step process, where the search for alternatives is part of second process. Here, we aim to test this hypothesis by focusing on Hit items in AN rules (those having an affirmative in the antecedent and a negation in the consequent as in *If there is an H then there is not a square*), where only the consequent requires the interpretation of a negation.

When one considers these Hit items (e.g. *H-in-a-circle* with respect to the rule above), a participant with only a narrow-reading of negations would have to focus her attention on the inconsistency and would have slower reaction times and a greater N2, as our prior findings have shown. However, this (potentially dominant) strategy might be concealing a strategy from participants who, not only narrowly interprets the negation but, go a step further by considering alternatives. Arguably, such participants (anticipate and) compute a contrast-class that would simplify a search for an *H in a non-square*. Such a participant aims to anticipate potential contrasts and would ultimately have a relatively fast reaction time (from the moment that the test-item appears) *and* a small N2. If our hypothesis is supported, this would amount to a negative[1] (between-subject) correlation between reaction times and N2 amplitude in this condition.

2. Methods

2.1. Participants

Twenty healthy native French-speaking volunteers with no history of neurological or psychiatric disorders participated in the study. Five participants were excluded from the analyses due to excessive eye movement artifacts in the EEG signal, and one because her overall mean RT was above 2.5 standard deviations from the mean. The remaining 14 participants (8 males) were aged between 22 and 40 (mean age: 28 years). All subjects were right-handed as measured by the Edinburg Handedness Inventory (Oldfield, 1971).

---

[1] The correlation should be negative because the greater the N2, the more negative is its voltage.

2.2. Materials

A single trial was composed of a conditional *if-then* statement (in French) 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, e.g., 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 Fig. 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. in order 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. in order 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. in order 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 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, 90 out of 100 trials per condition contained items that had a true antecedent. The remaining 10 stimuli per condition presented a letter that was irrelevant to the antecedent of the conditional rule; this leads to a non-obvious evaluation (consider the rule *If there is a J then there is a square* and the item *H-in-a-square*).[2] These 10 were included in order to avoid predictability in the task and were considered fillers. Half of the 90 relevant trials were confirming cases (e.g., an item like *H-in-a-square* verifies the rule *If there is an H then there is a square*) and half disconfirming cases (e.g., 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.

2.3. Tasks and procedure

Visual stimuli were generated with Presentation 10.2 software (Neurobehavioral Systems, http://www.neurobs.com/) and displayed on a computer screen. Each trial started with the presentation of a visual fixation mark (a central dot) in the center of the screen for 500 ms (see Fig. 1). 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 ms and the second part ("then there is a square") at 1500 ms. The entire rule then remained on the screen for a further 3000 ms, at which point the rule disappeared and the central dot reappeared for 500 ms. This was immediately followed by the target item, which remained on the screen until subjects pressed one of two buttons with their right hand (yes/no response). Variable periods of visual fixations were added at the end of each trial. That is, the duration of a trial varied randomly between 9250 and 13350 ms (mean trial time = 11,300 ms).

Participants performed the experiment in five blocks (60 trial presentations in each). 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").

Trial order within each block was randomized and the block order within each task was counterbalanced across participants. Participants were instructed to respond as quickly and as accurately as possible. The task began with 10 training trials.

2.4. Electroencephalogram (EEG) recording

EEG was recorded with a 64 channel NetAmps 200 system (Electrical Geodesics Inc.). Amplified analogue voltages (0.1–100 Hz bandpass) were sampled at 250 Hz. Electrode impedance was kept below 60 kΩ. All channels were referenced to Cz during recording, and off-line re-referenced to the left mastoid. ERP analyses were conducted using ELAN-Pack software developed at INSERM U821 (Lyon, France). They consisted in averaging the EEG segments in synchronization with the onset of the target item in each trial over a 700 ms period including a 100 ms pre-stimulus interval. The signals were low-pass filtered (30 Hz) and a baseline correction was calculated from the 100 ms pre-stimulus interval. Trials contaminated by eye blinks

---

or eye movements (threshold: ±100 μV) were not included in the analyses. Also, analyses were restricted to trials on which subjects made the correct (i.e. logical) response. The N2 component was defined as the greatest negative peak occurring within a latency window between 200 and 400 ms at 12 central, frontocentral and frontal sites: Cz, C1, C1′, C2, C2′, Fcz, Fc1, Fc2, Fz, Fz′, F1, F2 (Fig. 1B). In order to simplify the analyses of ERP amplitudes, data from C1 and C1′ were collapsed into one C1 site, data from C2 and C2′ were collapsed into one C2 site, and data from Fz and Fz′ were collapsed into one Fz site. We thus analyzed nine sites, falling into a frontal/frontocentral/central × left/midline/right array (see Fig. 1B).

3. Results

3.1. Behavioral analysis

Behavioral performances (accuracy and mean reaction times) are shown in Table 2. Analyses are broken down into three parts. First, we focus exclusively on the rules containing no negation in the antecedent (i.e. AA and AN). Second, we analyze participants' performance using data from the two remaining rules NA and NN (with a negation in the antecedent). Third, we consider the overall main effect of mismatch collapsing all the four rules AA, AN, NA, and NN.

Using only behavioral data from the rules AA and AN, we carried-out a 2 × 2 repeated measures analysis of variance (ANOVA) with the factors Rule (AA vs. AN) and Response-Type (Hit vs. CR) on correct RTs. This ANOVA revealed a main effect of Response-type ($F(1,13) = 8.07$, $p < 0.05$; Hit: 1260 ms, CR: 1360 ms), a main effect of Rule ($F(1,13) = 32.74$, $p < 0.001$; AA: 1163 ms; AN: 1497 ms), and a Response-Type × Rule interaction ($F(1,13) = 18.33$, $p < 0.001$). That is, Hit responses were faster than CR responses for the rule AA (Hit: 942 ms; CR: 1304 ms), whereas the reverse pattern was observed for the rule AN (Hit: 1578 ms; CR: 1416 ms).

A second within-subject ANOVA with the factors Rule (NA vs. NN) and Response-type (Hit vs. CR) was subsequently performed on correct RTs from the rules NA and NN. This ANOVA showed a main effect of Response-Type ($F(1,13) = 17.29$, $p < 0.01$; Hit: 1552 ms, CR: 1890 ms) and a marginally significant interaction between Rule and Response-Type ($F(1,13) = 4.09$, $p < 0.07$). That is, it was faster to provide a Hit than a CR for the rule NA (Hit: 1418 ms; CR: 1881 ms), but this difference was smaller for the rule NN (Hit: 1686 ms; CR: 1900 ms). The main effect of Rule did not reach significance ($F(1,13) = 2.42$, $p = $ n.s.; NA: 1650 ms, NN: 1793 ms).

In a final analysis, we collapsed all the rules into three different conditions of mismatch (0-mismatch, 1-mismatch and 2-mismatch; see Table 2). Overall, responses were fast when the pictorial item completely matched with the letter and shape mentioned in the rule (0-mismatch; 1204 ms), they were slower when there was one mismatch (1-mismatch; 1542 ms) and even slower when there were two mismatches (2-mismatch; 1775 ms). A repeated measures ANOVA confirmed the results above, showing a significant main effect of Mismatch-Level on reaction times ($F(2,26) = 36.50$, $p < 0.001$).

3.2. ERP analysis

As for the behavioral analyses, we first focus on the N2 amplitude measured for the rules AA and AN across the nine sites studied. Fig. 2 displays the ERP data we obtained for these rules on two representative electrodes. We carried-out a within-subject ANOVA with the factors Rule (AA, AN), Response-Type (Hit, CR), Anterior-to-Posterior (frontal, frontocentral, central) and Left-to-Right scalp locations (left, midline, right) on N2 amplitudes. Although there was no main effect of Rule ($F(1,13) = 0.08$, $p = $ n.s.) or Response-Type ($F(1,13) = 0.31$, $p = $ n.s.), we found a significant interaction between these two factors ($F(1,13) = 5.95$, $p < 0.05$). As anticipated, the N2 amplitude was greater when the subjects had to correctly reject

---

[2] 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.
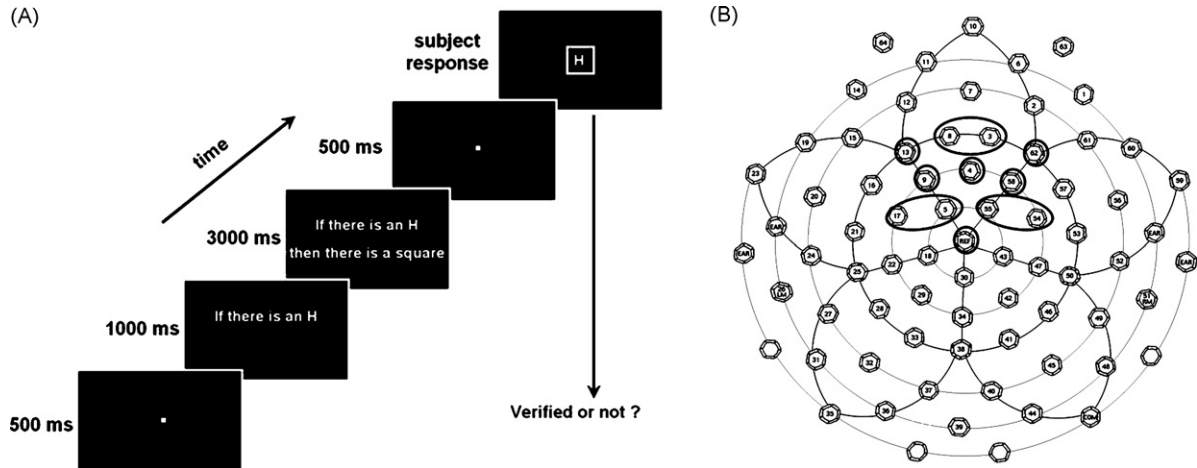
(A)

(B)

Fig. 1. Experimental procedure and layout of the electrode array. (A) Timing of a sample trial. (B) Electrode layout of the 64 channel geodesic sensor net. Sites of interest are circled in black and were grouped for statistical analysis. Frontal electrodes are shown at the top of the figure.

Table 2
Overall behavioral performance

Rule	Item	Correct Response	Mismatching level	Percentage Correct	RT[a]
If H then square (AA)	*H-in-square*	Yes (Hit)	0	98	942
If H then square (AA)	*H-in-circle*	No (CR)	1	97	1304
If H then not square (AN)	*H-in-circle*	Yes (Hit)	1	98	1578
If H then not square (AN)	*H-in-square*	No (CR)	0	97	1416
If not H then square (NA)	*P-in-square*	Yes (Hit)	1	96	1418
If not H then square (NA)	*P-in-circle*	No (CR)	2	91	1881
If not H then not square (NN)	*P-in-circle*	Yes (Hit)	2	96	1686
If not H then not square (NN)	*P-in-square*	No (CR)	1	87	1900

Mean accuracy and reaction times. *Notes.* This table is based on a model case *If there is an H then there is a square*, but letters and shapes were varied.

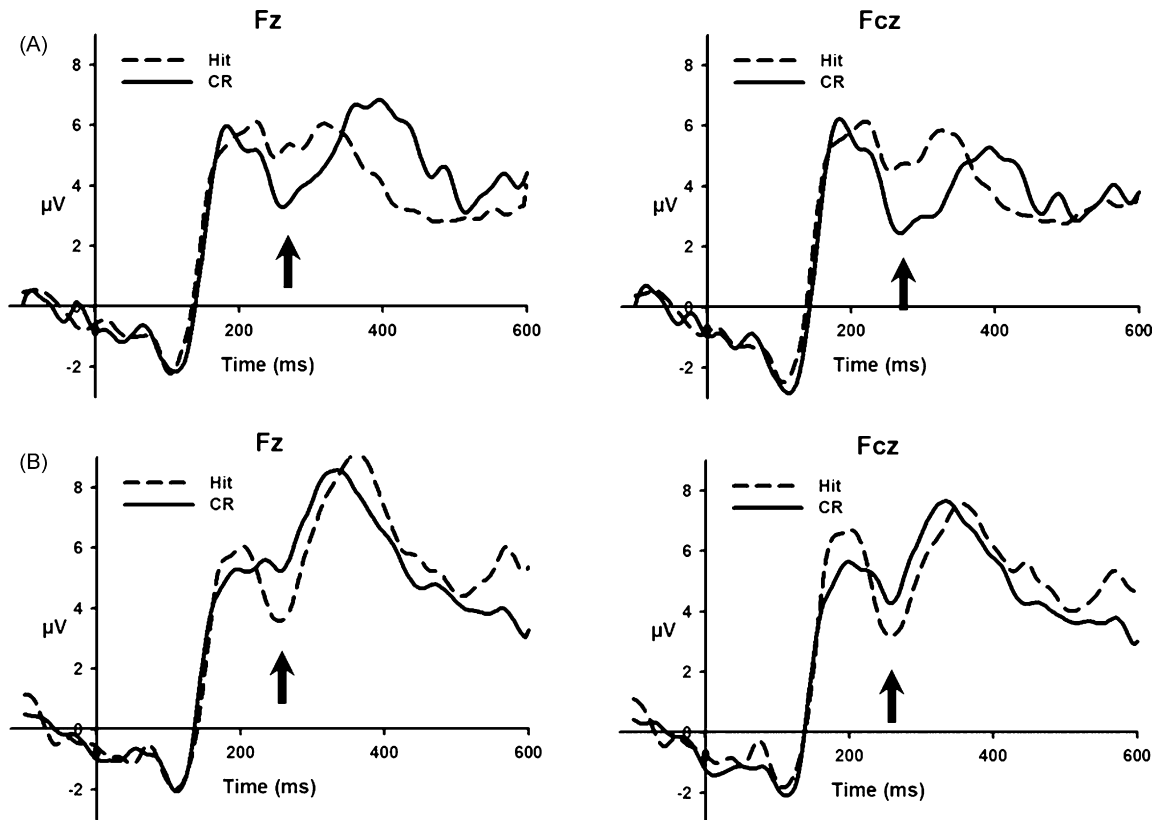 [a] These are reaction times to correct responses only.

Fig. 2. Stimulus locked grand-average waveforms evoked by the appearance of the target item in the rule AA (A) and AN (B) at the representative sites Fz and Fcz. The graphs show the waveforms for Hit responses (Hit, dashed black line) and for Correct Rejections (CR; solid black line). The target item onset is at 0 ms and the N2 is indicated by an arrow.
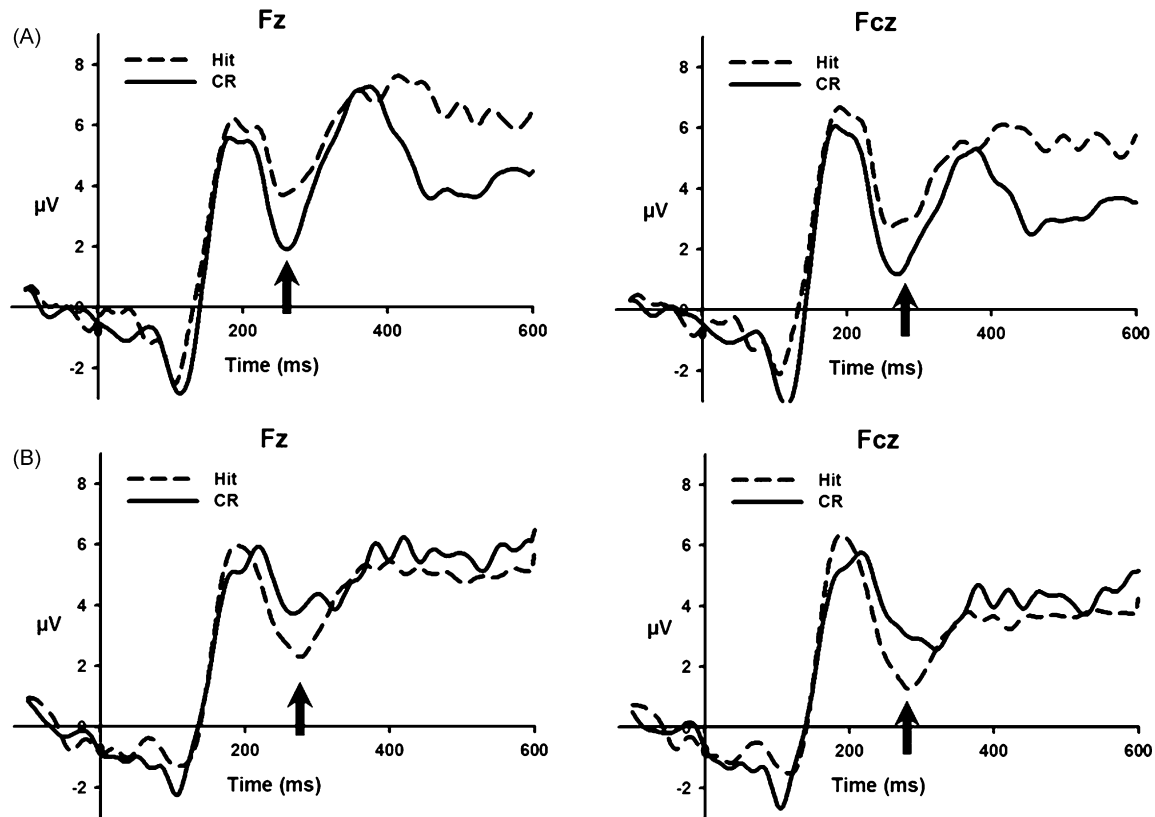
Fig. 3. Stimulus locked grand-average waveforms evoked by the appearance of the target item in the rule NA (A) and NN (B) at the representative sites Fz and Fcz. The graphs show the waveforms for Hit responses (Hit, dashed black line) and for Correct Rejections (CR; solid black line). The target item onset is at 0 ms and the N2 is indicated by an arrow.

the rule AA (2.64 μV; 1-mismatch) than when they had to confirm it (4.29 μV; 0-mismatch) (one-tailed paired t-test, $t(13) = 1.92$, $p < 0.05$) (Fig. 2A). A significant difference in the predicted opposite prediction was observed for the rule AN, where a Hit (3.60 μV; 1-mismatch) yielded a greater N2 than a CR (4.18 μV; 0-mismatch) (one-tailed paired t-test, $t(13) = 1.93$, $p < 0.05$) (Fig. 2B). We also observed a main effect of Left-to-Right scalp location (with smaller N2 for central sites as compared to more lateral ones; $F(2,26) = 4.09$, $p < 0.05$), an interaction between Left-to-Right and Anterior-to-Posterior scalp location ($F(4,52) = 3.04$, $p < 0.05$), but no interaction between either Rule or Response-Type and these factors. No other results were significant.

We then focus on the ERP data obtained for the rules NA and NN, i.e. the two rules entailing a negation in the antecedent (see Table 2). Again, we entered N2 amplitudes into an ANOVA using Rule (NA, NN), Response-Type (Hit, CR), Anterior-to-Posterior (frontal, frontocentral, central) and Left-to-Right scalp locations (left, midline, right) as within-subject factors. The interaction Rule × Response-type reached significance ($F(1,13) = 4.81$, $p < 0.05$), showing greater N2 for CR (1.75 μV; 1-mismatch) than Hit (3.14 μV; 0-mismatch) in the rule NA (one-tailed paired t-test, $t(13) = 1.29$, $p = 0.11$) (Fig. 3A), and greater N2 for Hit (1.69 μV; 1-mismatch) than CR (3.21 μV; 0-mismatch) in the rule NN (one-tailed paired t-test, $t(13) = 1.87$, $p < 0.05$) (Fig. 3B). The only other significant effect was an interaction between Rule and Anterior-to-Posterior, with greater N2 in anterior than posterior sites for the rule NN, but not for the rule NA.

In an additional analysis, we collapsed ERP data from the four rules (AA, AN, NA and NN) into three separate conditions of mismatch (0-mismatch, 1-mismatch, 2-mismatch) (Fig. 4). Within-subject ANOVAs on N2 amplitudes with Mismatch-Level (0, 1, 2), Anterior-to-Posterior (frontal, frontocentral, central) and Left-to-

Right scalp locations (left, midline, right) as factors revealed a main effect of mismatch ($F(2,26) = 10.75$, $p < 0.001$), a Mismatch-Level × Left-to-Right scalp location interaction ($F(4,52) = 2.99$, $p < 0.05$) and a Mismatch-Level × Anterior-to-Posterior × Left-to-Right scalp location interaction ($F(8,104) = 5.73$, $p < 0.001$). Post hoc paired t-tests revealed that the N2 was greater when subjects had to deal with 1-mismatch than 0-mismatch (two-tailed paired t-test, $t(13) = 2.26$, $p < 0.05$) and even greater when they had to deal with 2-mismatches than 1-mismatch (two-tailed paired t-test, $t(13) = 3.00$, $p < 0.05$).

Finally, we tested our prediction that there are inter-individual differences in the way participants process negations in conditional statements. More precisely, we focused on the rule AN because such a rule entails only a negation in the consequent (see Table 2). We found a significant negative correlation between N2 peak amplitude and RT only when the participants had to provide a Hit response (1-mismatch; $r = -0.53$; $p < 0.05$) (see Fig. 5).

4. Discussion

The present study aimed to record ERPs of participants who were required to give a correct logical response in the context of mismatches. Analysis of the behavioral data confirmed and replicated earlier studies showing that mismatches between a rule and a test-item negatively affect reasoning performance (Evans, 1972; Prado & Noveck, 2006, 2007).

We found that, although reaction times increased with the number of negations in the rules, the presence of mismatches between the rule and the target item was also predictive of the behavioral performance. That is, evaluating a rule that mismatched the target
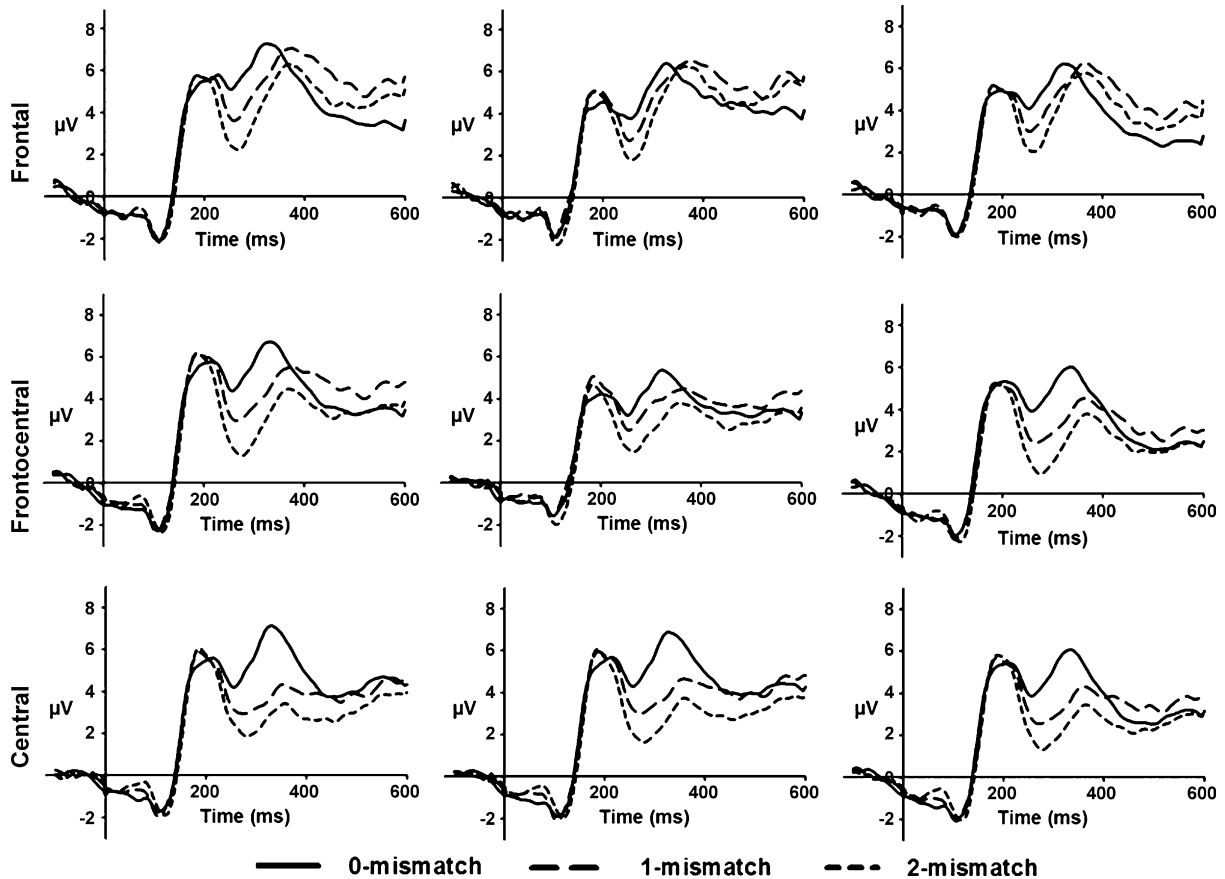
Fig. 4. Stimulus locked grand-average waveforms evoked by the appearance of the target item in the 0-mismatch (solid black line), the 1-mismatch (medium dashed black line), and the 2-mismatch (short dashed black line) conditions across the nine sites of interest. Left electrodes are shown in the left column, midline electrodes in the middle column, and right electrodes in the right column.

item was typically associated with longer reaction times than evaluating a rule that did not entail mismatch. This effect, persistent even when participants were not required to deal with a different number of negations in conditional rules, was observed across the four rules AA, AN, NA and NN.

Fig. 5. Scatterplot showing the N2 amplitude as a function of participants' reaction times when they provide a Hit response to the rule AN. Each plotted point represents the mean N2 amplitude (over the nine electrodes of interest) and the mean reaction time for a single subject.

Analyses of the ERPs revealed that a N2 component emerged at frontocentral sites 200–300 ms following the presentation of the target item (Figs. 2–4). More importantly, the amplitude of this component was linked to the presence of mismatching features between the conditional rule and the target item. N2 amplitude was greater in the rule AA (e.g. *If there is an H then there is a square*) when participants provided a CR (*H-in-a-circle*; 1-mismatch) than a Hit (*H-in-a-square*; 0-mismatch), but greater in the rule AN (e.g. *If there is an H then there is not a square*) when participants provided a Hit (*H-in-a-circle*; 1-mismatch) that a CR (*H-in-a-square*; 0-mismatch). Importantly, no difference was observed when the overall N2 amplitude in the rule AA was compared to the overall amplitude in the rule AN. That is, neither the Type of response (Hit vs. CR) nor the presence of a negation (AA vs. AN) appear to drive the N2, only the presence of mismatch.

This effect on the N2 amplitude was replicated using the rules NA (e.g. *If there is not an H then there is a square*) and NN (e.g. *If there is not an H then there is not a square*). Here, N2 was greater in the rule NA for CRs (e.g. *J-in-a-circle*; 2-mismatch) than Hits (e.g. *J-in-a-square*; 1-mismatch), and greater in the rule NN for Hits (e.g. *J-in-a-circle*; 2-mismatch) than CRs (e.g. *J-in-a-square*; 1-mismatch). Additionally, we collapsed all the conditions and found that the amplitude of this N2 was parametrically modulated by the number of mismatches between the rule and the target item, so that it was greater for 2-mismatches than for 1-mismatch, and even greater for 1-mismatch than for 0-mismatch (see Fig. 4). Taken together, the results above show (i) that giving a correct response on rules involving mismatching is associated with the emergence of a frontocentral N2, (ii) that the amplitude of this N2 is closely
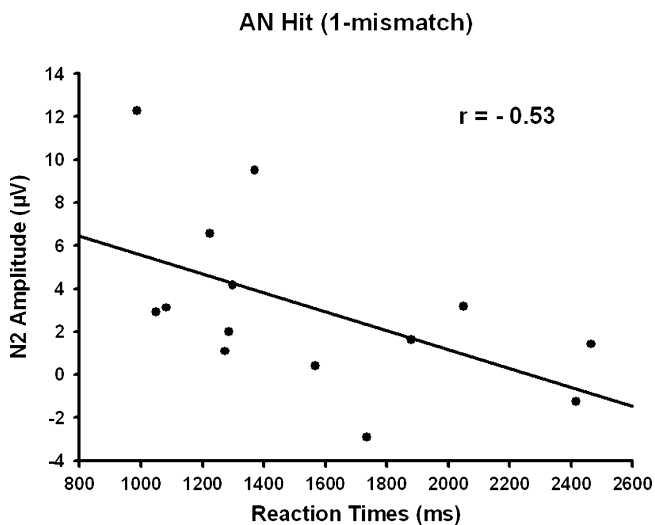
related to the level of mismatch between the rule and the target item, and (iii) that the mismatches are the basis of the N2.

These results are consistent with our previous behavioral and fMRI experiments. First, using the same paradigm, we showed that the mismatches themselves, not the negations, were predictive of reaction times (Prado & Noveck, 2006). That is, when faced with the rule AA, we found that participants were longer and less accurate to provide a CR (1-mismatch) than a Hit (0-mismatch). Here, we demonstrate greater N2 amplitude for CR than Hit when participants reasoned on the same rule AA.

This is consistent with a long-standing literature showing that the frontocentral N2 is highly sensitive to perceptual mismatches. For example, N2 amplitude has been found to be enhanced when the degree of perceptual mismatch between two sequential stimuli was increased (Wang et al., 2003), or when a cue differed perceptually from a target (Gehring et al., 1992). Although this effect has been shown in a variety of stimuli, our study is the first to show that varying the perceptual features between a logical rule and a target item trigger the same mismatch-related N2. Moreover, we found in our task that the amplitude of the N2 was parametrically modulated by the number of mismatching features between the rule and the target item. This effect is consistent with a previous study showing a similar parametric modulation of the N2 when different levels of perceptual conflict were manipulated in a sequential matching task (Wang et al., 2004).

Our findings also shed some light on the way negations are processed in conditional statements. We found a greater N2 amplitude when subjects successfully confirmed an AN rule (e.g. properly responded that *H-in-a-circle* confirms *If there is an H then there is not a square*) than when they had to correctly reject it. Giving a Hit response in that case (e.g. *H-in-a-circle*) involves dealing with more mismatching information than correctly rejecting the rule (*H-in-a-square*). The same effect was observed for the rule NN. This result strongly suggests that negations are not widely processed in conditional rules, but rather only serve to focus participant's attention on the negated constituent, as stated by the *heuristic* account (see Section 1). That is, when faced with the sentence *there is not a square*, the subjects do not consider other alternatives (i.e. other shapes) and the arrival of a target item involving another shape (e.g. *circle*) mismatches the initial statement, leading to longer reaction times and an enhanced N2. This interpretation is not consistent with the contrast-class (or rational) account of the mismatching effects. On the contrary, it supports the view that the mismatching effects result from a heuristic process. These findings are also consistent with the proposal that the N2 may reflect a mismatch between a target stimulus and a mental template (in this case the negated constituent), as suggested by Folstein and Van Petten (2008).

Can we claim that all participants prefer only a narrow reading view? Although much of our evidence defends a *narrow-reading* view of negations (see above), our findings also suggest that this strategy is subject to a reliable inter-individual variability. That is, participants who were the fastest to provide a Hit response in the rule AN (involving dealing with one mismatch if the negation is processed narrowly) were also the ones with the smallest N2. We propose that this effect arises because these participants process negations more widely than the participants exhibiting longer reaction times (and thus greater N2). Indeed, a participant who would compute a contrast-class when faced with the proposition *there is not a square* would be primed to evaluate the shape *circle* and would thus exhibit a smaller N2 (and a shorter RT) than a participant who would have a narrow-reading of negations (and would have to overcome a mismatch when faced to the shape *circle*). Such a hypothesis suggests that not every participant applies the same strategy when faced with a conditional rule involving negations, and could explain some discrepancies in the reasoning literature

concerning the role of negations in the mismatching effects (Evans, 1998; Oaksford, 2002).

More generally, these findings are in-line with influential dual-process theories of reasoning. Recently, several researchers in reasoning and decision making have converged on the idea that cognitive biases (such as the mismatching effect) could be explained in terms of an interaction between two different cognitive systems (Evans, 2003; Sloman, 1996; Stanovich, 2004). These *dual-systems* theories posit that heuristic processes are rooted in specific sub-components of a more general cognitive system, often referred to as the *heuristic* system (or System 1). Although the exact properties of the *heuristic* system can vary among theorists, it is in general believed to be largely automatic, preconscious, fast operating, and shared by humans and animals. Heuristic processes in System 1 are hypothesized to have their own neural bases and to be the source of many non-rational behaviors, including mismatching effects (Evans, 2003). To explain why people are nevertheless able to exhibit a rational behavior, System 1 is thought to compete with an *analytic* system, termed System 2. System 2 is more controlled, evolutionary more recent, slow operating and based on formal rules. It is believed to be the cornerstone of logical reasoning and hypothetical thinking (although it is assumed to be constrained by working memory capacity and related to measures of general intelligence).

On many occasions, such as when no mismatches are present in a conditional rule, both System 1 and System 2 will provide the same response on a given reasoning problem. In these cases, using a fast System 1 process will suffice in terms of computational resources needed. However, both analytic and heuristic systems may also provide different competing outputs. This is likely to be the case when the providing of a correct analytic response involves dealing with a mismatch between the items mentioned in the rule (i.e. T7) and those in the letter pair (i.e. P4). That is, in order to give a logical (i.e. task-relevant) response, subjects have to override the System 1 heuristic tendency to see these mismatching cases as irrelevant to the rule. One of the central claims of most dual-process theories is that it is precisely the role of System 2 to inhibit System 1 heuristic processes when an analytic response is needed (Evans, 2007; Stanovich & West, 2000). This is assumed to be the reason why participants are able to provide a rational and context-dependent response (Houde & Tzourio-Mazoyer, 2003).

Interestingly, the N2 has been also linked to cognitive and inhibitory control. Indeed, this ERP is classically observed in tasks involving the inhibition of a prepotent motor response, such as the so-called *go/no-go* paradigm. In this task, participants are required to quickly respond to a certain type of stimulus (*go*) but to withhold their motor response to another type of stimulus (*no-go*). Typically, the N2 is greater after *no-go* trials than after *go* trials (Eimer, 1993; Falkenstein, Hoormann, & Hohnsbein, 1999; Jodo & Kayama, 1992; Pfefferbaum, Ford, Weller, & Kopell, 1985). More recently, it has been proposed that the N2 is more likely to reflect the monitoring of a cognitive conflict than the inhibitory process *per se* (Botvinick, Cohen, & Carter, 2004; Yeung, Cohen, & Botvinick, 2004). For example, using a classic interference task (i.e. the Flanker task), (Van Veen & Carter, 2002) have observed that the frontocentral N2 was both elicited prior to the motor response and generated by the ACC, a structure we highlighted in our previous fMRI study on the mismatching effect. Although the exact role of the N2 is still under debate, there is a general consensus that this ERP component is closely related to effortful cognitive control processes: the greater its amplitude the more effort involved in controlling cognitive processes (and the smaller its amplitude, the better is participants' performance in interference tasks) (Lamm, Zelazo, & Lewis, 2006).

Our current results show that a similar N2 is evoked when participants are faced with mismatches in a reasoning task. Such a

finding could suggest that effortful cognitive control processes are engaged when participants are faced with mismatches in conditional rules. The N2 elicited when reasoners overcome mismatches could then be considered the electrophysiological signature of this inhibition of System 1 by System 2.

Acknowledgments

References

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences*, 8(12), 539–546.

Braine, M. D. S., & O'Brien, D. P. (1998). *Mental logic*. NJ: Lawrence Erlbaum.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews. Neuroscience*, 3(3), 201–215.

Courchesne, E., Hillyard, S. A., & Galambos, R. (1975). Stimulus novelty, task relevance and the visual evoked potential in man. *Electroencephalography and Clinical Neurophysiology*, 39(2), 131–143.

Daffner, K. R., Mesulam, M. M., Scinto, L. F., Calvo, V., Faust, R., & Holcomb, P. J. (2000). An electrophysiological index of stimulus unfamiliarity. *Psychophysiology*, 37(6), 737–747.

Eimer, M. (1993). Effects of attention and stimulus probability on ERPs in a Go/Nogo task. *Biological Psychology*, 35(2), 123–138.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception and Psychophysics*, 16, 143–149.

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. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology*, 75, 451–468.

Evans, J. S. B. T. (1989). *Bias in human reasoning: Causes and consequences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

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(10), 454–459.

Evans, J. S. B. T. (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking and Reasoning*, 13(14), 321–339.

Falkenstein, M., Hoormann, J., & Hohnsbein, J. (1999). ERP components in Go/Nogo tasks and their relation to inhibition. *Acta Psychologica (Amst)*, 101(2–3), 267–291.

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology*, 45(1), 152–170.

Gehring, W. J., Gratton, G., Coles, M. G., & Donchin, E. (1992). Probability effects on stimulus evaluation and response processes. *Journal of Experimental Psychology. Human Perception and Performance*, 18(1), 198–216.

Houde, O., & Tzourio-Mazoyer, N. (2003). Neural foundations of logical and mathematical cognition. *Nature Reviews. Neuroscience*, 4(6), 507–514.

Jodo, E., & Kayama, Y. (1992). Relation of a negative ERP component to response inhibition in a Go/No-go task. *Electroencephalography and Clinical Neurophysiology*, 82(6), 477–482.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lamm, C., Zelazo, P. D., & Lewis, M. D. (2006). Neural correlates of cognitive control in childhood and adolescence: Disentangling the contributions of age and executive function. *Neuropsychologia*, 44(11), 2139–2148.

Milham, M. P., Banich, M. T., & Barad, V. (2003). Competition for priority in processing increases prefrontal cortex's involvement in top-down control: An event-related fMRI study of the stroop task. *Brain Research. Cognitive Brain Research*, 17(2), 212–222.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2, 175–220.

Oaksford, M. R. (2002). Contrast classes and matching bias as explanations of the effects of negation on conditional reasoning. *Thinking and Reasoning*, 8, 135–151.

Oaksford, M., & Chater, N. (1995). Information gain explains relevance which explains the selection task. *Cognition*, 57(1), 97–108.

Oaksford, M., & Chater, N. (1996). Rational explanation of the selection task. *Psychological Review*, 103(2), 381–391.

Oaksford, M., & Moussakowski, M. (2004). Negations and natural sampling in data selection: Ecological versus heuristic explanations of matching bias. *Memory and Cognition*, 32(4), 570–581.

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(1), 97–113.

Pfefferbaum, A., Ford, J. M., Weller, B. J., & Kopell, B. S. (1985). ERPs to response production and inhibition. *Electroencephalography and Clinical Neurophysiology*, 60(5), 423–434.

Prado, J., & Noveck, I. A. (2006). How reaction times elucidate the matching bias and the way negations are processed. *Thinking and Reasoning*, 12(3), 309–328.

Prado, J., & Noveck, I. A. (2007). Overcoming perceptual features in logical reasoning: A parametric functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience*, 19(4), 642–657.

Rips, L. (1994). *The psychology of proof*. Cambridge, MA: MIT Press.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119, 3–22.

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? *The Behavioral and Brain Sciences*, 23(5), 645–665.

Van Veen, V., & Carter, C. S. (2002). The timing of action-monitoring processes in the anterior cingulate cortex. *Journal of Cognitive Neuroscience*, 14(4), 593–602.

Wang, H., Wang, Y., Kong, J., Cui, L., & Tian, S. (2001). Enhancement of conflict processing activity in human brain under task relevant condition. *Neuroscience Letters*, 298(3), 155–158.

Wang, Y., Tian, S., Wang, H., Cui, L., Zhang, Y., & Zhang, X. (2003). Event-related potentials evoked by multi-feature conflict under different attentive conditions. *Experimental Brain Research*, 148(4), 451–457.

Wang, Y., Cui, L., Wang, H., Tian, S., & Zhang, X. (2004). The sequential processing of visual feature conjunction mismatches in the human brain. *Psychophysiology*, 41(1), 21–29.

Yang, J., & Wang, Y. (2002). Event-related potentials elicited by stimulus spatial discrepancy in humans. *Neuroscience Letters*, 326(2), 73–76.

Yeung, N., Cohen, J. D., & Botvinick, M. M. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review*, 111(4), 931–959.