Cognition xxx (2009) xxx–xxx

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Belief inhibition during thinking: Not always winning but at least taking part

Wim De Neys *, Samuel Franssens

Lab Experimental Psychology, University of Leuven, Tiensestraat 102, 3000 Leuven, Belgium

ARTICLE INFO

Article history:
Received 19 June 2008
Revised 23 July 2009
Accepted 23 July 2009
Available online xxxx

Keywords:
Reasoning
Decision-making
Inhibition

ABSTRACT

Human thinking is often biased by intuitive beliefs. Inhibition of these tempting beliefs is considered a key component of human thinking, but the process is poorly understood. In the present study we clarify the nature of an inhibition failure and the resulting belief bias by probing the accessibility of cued beliefs after people reasoned. Results indicated that even the poorest reasoners showed an impaired memory access to words that were associated with cued beliefs after solving reasoning problems in which the beliefs conflicted with normative considerations (Experiment 1 and 2). The study further established that the impairment was only temporary in nature (Experiment 3) and did not occur when people were explicitly instructed to give mere intuitive judgments (Experiment 4). Findings present solid evidence for the postulation of an inhibition process and imply that belief bias does not result from a failure to recognize the need to inhibit inappropriate beliefs, but from a failure to complete the inhibition process. This indicates that people are far more logical than hitherto believed.

© 2009 Published by Elsevier B.V.

UNCORRECTED PROOF

1. Introduction

Human beings sometimes give the impression of being irrational. Consider, for example, people's puzzling preference for bottled water over tap water (Standage, 2005). Americans alone spend around $10 billion on bottled water each year. Although people cannot tell the difference between tap and bottled water in blind tastings, most of us nevertheless prefer to buy the bottled version. Water in a good-looking, sealed container seems to be automatically associated with purity and cleanliness. Although water from municipal water supplies is actually more stringently monitored and tightly regulated, people believe it is more likely to be contaminated. Despite numerous municipal projects promoting the benefits of tap water it seems hard for people to suppress the idea that bottled water is safer. Consequently, people keep on spending their money on the more expensive, more environmentally wasteful bottled alternative.

Scientific studies on reasoning and decision making confirm people's difficulty with discarding inappropriate beliefs. Over the last 50 years hundreds of studies have shown that in a wide range of reasoning tasks most educated adults fail to give the answer that is correct according to logic or probability theory. People seem to over-rely on intuitive gut feelings and stereotypical beliefs instead of on more demanding, deliberate reasoning when making decisions (Evans, 2003; Kahneman & Tversky, 1973; Sloman, 1996). Although this intuitive or so-called 'heuristic' thinking might sometimes be useful, it will often cue responses that conflict with more normative considerations. Just as in the bottled water example, it is assumed that sound reasoning in these cases requires that people temporarily suppress their intuitive beliefs and refrain from taking them into account. Such a belief inhibition plays a key role in theories of reasoning, decision-making, and social cognition and is considered one of the most fundamental higher-order cognitive abilities (e.g., Evans, 2008; Houdé, 1997, 2007; Stanovich & West, 2000).

Despite the popularity of the belief inhibition claim, it is surprising to note that the basic processing characteristics

* Corresponding author. Tel.: +32 16 326143; fax: +32 16 326099.
E-mail address: Wim.Deneys@psy.kuleuven.be (W. De Neys).

0010-0277/$ - see front matter © 2009 Published by Elsevier B.V.
doi:10.1016/j.cognition.2009.07.009

Please cite this article in press as: De Neys, W., & Franssens, S. Belief inhibition during thinking: Not always winning but at least taking part. *Cognition* (2009), doi:10.1016/j.cognition.2009.07.009

have hardly been examined. A crucial case in point is the nature of an inhibition failure. At least two different views can be contrasted. People might be biased because they are not aware that their beliefs conflict with more normative considerations and consequently do not even initiate an inhibition process (e.g., Ehrlinger, Johnson, Banner, Dunning, & Kruger, 2008; Kahneman, 2002). Alternatively, one might suggest that people do detect that cued intuitive beliefs are unwarranted and attempt to inhibit their beliefs, but simply fail to complete the process. The point is whether belief bias arises because of a failure to engage in an inhibition process or because of a failure to complete it. The answer to this question has far stretching implications for claims about human rationality (e.g., see De Neys, 2006a). Bluntly put, the first view suggests that people do simply not realize that their response is wrong. Reasoners would not know that their beliefs conflict with traditional logical or probabilistic norms or would not consider these norms to be relevant. The second view, however, implies that people's errors are less ignorant. If people actively try to block the belief-based response, this suggests that they know that it is not fully warranted and try to do something about it. This sketches a less bleak picture of human rationality. Not everybody might manage to win the inhibition struggle, but everybody would at least be taking part and try to adhere to the norms.

Based on the available reasoning data it is hard to decide between the different failure views (Evans, 2007, 2008). Much publicity has been given, for example, to recent brain-imaging studies showing that successfully overcoming belief bias during reasoning activates a specific region of the frontal lobes (i.e., the lateral prefrontal cortex, e.g., De Martino, Kumaran, Seymour, & Dolan, 2006; De Neys, Vartanian, & Goel, 2008; Goel & Dolan, 2003; Houdé et al., 2000; Prado & Noveck, 2007; Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003). This same region is also involved in responding to basic cognitive control tasks in which inhibition of a habituated, erroneous response is paramount (e.g., Aron, Robbins, & Poldrack, 2004). Although such studies are important to localize the brain-regions that are involved in sound reasoning, they do not help us to draw strong conclusions about the nature of the inhibition failure. An insufficient recruitment of the specific brain-areas that mediate the inhibition process fits both with the engagement failure and the completion failure view. In a similar vein, individual differences studies have shown that people highest in cognitive capacity (i.e., participants with high IQ's or working memory spans) manage to overcome belief bias and reason in line with normative standards (e.g., De Neys, 2006a; De Neys & Verschueren, 2006; Newstead, Handley, Harley, Wright, & Farrelly, 2004; Stanovich & West, 2000). Although such findings suggest that belief inhibition is a demanding process, they do not show us why people fail to inhibit. It might be that bad reasoners lack sufficient resources to complete the inhibition process or it might be that people with insufficient cognitive resources are simply not aware that inhibition is required.

A closer look at the belief inhibition studies in the reasoning field points to an even deeper problem. Evidence for the role of an inhibition process is typically quite indirect. The brain-imaging and individual differences studies, for example, do not show us that people actually discard their beliefs. They indicate that the postulated belief inhibition process is demanding and activates a brain region that is activated when people need to withhold prepotent responses, but this does not imply that the cued erroneous beliefs were actually blocked. This point is not trivial. In our opinion, a lot of the explanatory power and popularity of the belief inhibition claim rests on the analogy with classic findings in the memory field. It is well established in memory studies that when people have to suppress unwanted thoughts or actively neglect information, access to this information will be distorted (e.g., MacLeod, Dodd, Sheard, Wilson, & Bibi, 2003; Neill, 1997; Tipper, 1985). The inhibition concept basically refers to this temporary inaccessibility of initially discarded information. Reasoning theories assume that people go through a similar information discarding process during reasoning. However, in contrast with the memory studies, we are lacking any direct evidence with respect to the crucial impact of the postulated inhibition process on the accessibility of the beliefs. The present study will address this shortcoming. We adopt a classic procedure from the memory literature to probe the accessibility of cued beliefs after people engage in a reasoning task. The findings will provide a more solid ground for the postulation of a belief inhibition process during thinking and will help us to clarify the nature of an inhibition failure.

At this point one might note that there is some controversy in the memory field with respect to the theoretical status of the inhibition concept. It is debated whether an observed temporary inaccessibility of a memory trace entails that the information was simply tagged as inappropriate or literally deactivated at the neural level (see MacLeod et al., 2003, for a review). Some memory researchers have suggested that the inhibition label should only be used to refer to an actual neural deactivation. The present study does not speak to this issue. Both views imply that people have previously tried to disregard the impaired information. It is precisely such a discarding process that reasoning and decision making researchers traditionally envisage when referring to belief inhibition. We use the traditional label belief inhibition to refer to this postulated discarding process during reasoning. The key question for reasoning and decision-making theories is whether we can demonstrate that this postulated process impairs the accessibility of cued beliefs.

To test our hypotheses we first presented participants with classic reasoning problems in which intuitive beliefs and logical or probabilistic considerations conflicted or not (i.e., conflict and no-conflict problems). In the conflict problems sound reasoning required that people inhibited a cued belief-based response. In the no-conflict or control problems such inhibition was not required since beliefs and normative considerations cued the same response. For example, in one study we asked participants to evaluate the validity of deductive syllogisms. Intuitively, people will be tempted to base their response to these problems on the believability of the conclusion. In the conflict versions this is problematic because the believability of the conclusion conflicts with its logical status (e.g., an invalid

UNCORRECTED PROOF

Please cite this article in press as: De Neys, W., & Franssens, S. Belief inhibition during thinking: Not always winning but at least taking part. *Cognition* (2009), doi:10.1016/j.cognition.2009.07.009

syllogism with a believable conclusion). Consider the following example: "All flowers are plants. Roses are plants. Therefore, roses are flowers". Although the conclusion in the example is logically invalid and should be rejected, intuitively many people will nevertheless tend to accept it because it fits with their prior beliefs. Sound reasoning requires that this belief-based thinking is temporarily discarded. However, on no-conflict versions the believability of the conclusion was consistent with its logical status (e.g., an invalid syllogism with an unbelievable conclusion). Consider the following example: "All fruit can be eaten. Hamburgers can be eaten. Therefore, hamburgers are fruit". Both a priori beliefs and logical considerations will tell participants to reject the conclusion. In this case there is no conflict and no need to inhibit the cued beliefs. Accuracy on such control problems is typically uniformly high.

In the present study we always presented participants with a lexical decision task after they had solved a reasoning problem. In a lexical decision task participants have to determine whether a string of presented letters is a word or not (Meyer & Schvaneveldt, 1971). In our study, half of the strings that were presented were non-words (e.g., "braxzl"). Half of the presented words were so-called 'target' words that were closely related to the beliefs that were cued in the reasoning task (e.g., "rose" or "hamburger"). The other half of the words were completely unrelated to the cued beliefs (e.g., "pencil"). The time people need to decide whether a string is a word or not allows us to test the inhibition claims. The classic memory studies established that neglecting specific thoughts or information distorts recall of this information (e.g., MacLeod et al., 2003; Neill, 1997; Tipper, 1985). If people go through a similar information discarding process during reasoning, putting your beliefs aside during reasoning should also hinder subsequent recall of these beliefs: After belief inhibition, memory access to cued beliefs and associated knowledge should be temporarily impaired. However, people do not need to inhibit their beliefs on the no-conflict problems. Consequently, if people really attempt to discard their beliefs when solving conflict problems, one expects to see longer lexical decision times on the target words after conflict than after no-conflict problems.

The crucial question with respect to the nature of the inhibition failure concerns the lexical decision performance of people who typically fail to solve the conflict problems correctly. If people err because they do not detect that their beliefs are erroneous and fail to initiate an inhibition process, then their recall should not be distorted. However, if everybody always engages in an inhibition process, then lexical access to target words after presentation of a conflict problem should be impaired whether or not the participant managed to solve the reasoning problems correctly.

We tested the predictions with two infamous reasoning tasks. In Experiment 1 participants were presented with deductive syllogisms whereas participants in Experiment 2 reasoned about problems that were modeled after the classic base-rate neglect problems (Kahneman & Tversky, 1973). In these probabilistic judgment problems a belief-based response cued by a stereotypical personality description can conflict with the normative response cued by consideration of the base-rates in a sample. We specifically selected these two tasks because they instigated much of the debate on human (ir)rationality. Consistency of the findings across different reasoning tasks will give us an indication of the generality of the results. In Experiment 3 and 4 the findings will be validated further. Experiment 3 examines whether the predicted impaired memory access is temporary in nature. Experiment 4 tests whether the impaired access disappears when reasoning task instructions take away the need to engage in belief inhibition.

2. Experiment 1

2.1. Method

2.1.1. Participants

A total of 96 undergraduates studying at the University of Leuven (Belgium) participated in return for course credit. All participants were native Dutch speakers.

2.1.2. Material

Reasoning task: The syllogistic reasoning task was based on the work of Sá, West, and Stanovich (1999) and Markovits and Nantel (1989). Participants evaluated eight conditional syllogisms. Four of the problems had conclusions in which logic was in conflict with believability (i.e., conflict problems, two problems with an unbelievable-valid conclusion, and two problems with a believable-invalid conclusion). For the other four problems the believability of the conclusion was consistent with its logical status (i.e., no-conflict problems, two problems with an unbelievable-invalid conclusion, and two problems with a believable-valid conclusion). The following item format was adopted:

All fruits can be eaten.
Hamburgers can be eaten.

Therefore, hamburgers are fruits.

1. The conclusion follows logically from the premises.
2. The conclusion does not follow logically from the premises.

A complete overview of all eight problems can be found in the Appendix A.

Lexical decision task: After each problem a total of 24 letter strings was presented. Participants indicated whether the string was a word or not by pressing one of two response keys. Half of the letter strings were non-words, the other half were Dutch words. Six of the presented words were target words that were closely related to the beliefs that were cued in the reasoning task. Targets were core words from the conclusion or strongly associated words. The other six words were completely unrelated to the beliefs that the conclusion referred to.[1]

[1] Note that target words will always be recognized faster than unrelated words because the mere presentation of the reasoning problem will prime the related target words. Despite the general priming, the crucial prediction remains that if the information in the conclusion is inhibited in case of a conflict, accessing the target words should take longer after solving conflict vs. no-conflict problems.

All words were selected with the help of a Dutch word association index (De Deyne & Storms, 2008). After we had constructed an initial list of target and unrelated words two raters were asked to validate the classifications. In the few cases that judgments diverged the specific word was replaced with an alternative that all parties could agree on. A complete overview of the selected words can be found in the Appendix A.

The crucial prediction concerns the lexical decision time for target words after solving conflict versus no-conflict problems. Clearly, different target words were used in the lexical decision tasks for conflict and no-conflict problems. To establish that there were no a priori lexical differences between the selected target words for conflict and no-conflict problems, these words were included as a subset of the stimuli in an unrelated lexical decision study. In this pilot study the lexical decision task was not preceded by a reasoning task. A total of 79 participants evaluated the words. Results showed that the lexical decision times of the target words for conflict (M = 593 ms, SE = 8.61) and no-conflict (M = 591 ms, SE = 8.69) problems did not differ, $F(1, 78) < 1$.

2.1.3. Procedure

Participants were tested in small groups. Participants were first familiarized with the task-format. They were shown an example of a reasoning problem and practiced the lexical decision task. It was clarified that in the actual experiment both tasks would always alternate. Participants received standard deductive reasoning instructions that stressed that the premises should be assumed to be true, and that a conclusion should be accepted only if it followed logically from the premises. The eight reasoning problems were presented in random order. We used a serial presentation format for the syllogistic reasoning task (e.g., Goel & Dolan, 2003). First, each premise was presented for 3 s. After 6 s the conclusion and response options appeared. The complete problem remained on the screen until participants entered their response. Average response time in the present experiment was 6.1 s (SD = 2.9). Hence, each reasoning trial lasted about 12 s.

The lexical decision trials started after the response on the reasoning problem was entered. The 24 strings that had been selected for that problem were presented in random order. Words were presented in the center of the screen and participants were instructed to respond as quickly as possible while avoiding errors. A fixation cross was presented for 500 ms before each word was presented. After the lexical decision trials the experiment was briefly paused until the participant was ready to continue with the next reasoning problem.

2.2. Results

Reasoning task: Participants' performance on the reasoning task was as expected. People were typically biased when cued beliefs and logic conflicted. Overall, correct response rates reached 53% on the conflict problems and 87% on the no-conflict problems, $F(1, 95) = 78.17$, $p < .0001$, $\eta_p^2 = .45$. As Table 1 shows, no-conflict problems were also solved faster than conflict problems, $F(1, 95) = 9.3$, $p < .003$, $\eta_p^2 = .09$. These results closely replicate the findings in previous studies with similar syllogistic reasoning problems (e.g., De Neys, 2006a; Markovits & Nantel, 1989).

Lexical decision task: The central question concerned participants' lexical decision performance. Incorrect classifications of the letter strings were infrequent (less than 6% error rate across all trials) and where they did occur they were excluded from the analysis. Our main focus was the lexical decision time for target words that were associated with the beliefs that had been cued in the reasoning task. We also entered the lexical decision times for unrelated words in the analysis. These data were submitted to a 2 (problem type: conflict or no-conflict) × 2 (word type: target or unrelated) repeated measures ANOVA.

Results showed that there was a main effect of the word type factor, $F(1, 95) = 155.7$, $p < .001$, $\eta_p^2 = .62$. Not surprisingly, lexical decisions were always faster for the target words than for the unrelated words which had not been primed during reasoning. More crucial was the main effect of the problem type factor, $F(1, 95) = 4$, $p < .05$, $\eta_p^2 = .04$, and its interaction with the word type factor, $F(1, 95) = 13.9$, $p < .001$, $\eta_p^2 = .13$. Consistent with the claim that people inhibit their beliefs in case of a belief-logic conflict, simple effect tests indicated that lexical decision times for belief-related target words were longer after solving conflict problems than after solving no-conflict problems, $F(1, 95) = 15.95$, $p < .001$, $\eta_p^2 = .14$. As Fig. 1 indicates, the lexical decision times for unrelated words that had not been cued during reasoning did not differ, $F(1, 95) < 1$. Hence, it is not the case that memory access is generally impaired after solving conflict problems. As one might expect, only the

Table 1
Reasoning accuracy (% correct) and response latencies (s) in the different experiments.

Task	Accuracy		Response time	
	Conflict	No-conflict	Conflict	No-conflict
Syllogisms				
Experiment 1 – standard	53% (3.6)	87% (1.5)	6.7 s (.53)	5.5 s (.44)
Experiment 3 – delay	61% (3.8)	89% (1.6)	6.4 s (.56)	5.8 s (.47)
Experiment 4 – instructions	9% (3.4)	97% (1.3)	4.4 s (.55)	2.8 s (.44)
Base-rates				
Experiment 2 – standard	32% (3.5)	96% (1.4)	16.8 s (.52)	15.2 s (.43)
Experiment 3 – delay	34% (3.8)	96% (1.6)	17 s (.56)	14 s (.47)
Experiment 4 – instructions	23% (3.8)	96% (1.3)	15.1 s (.52)	13.7 s (.42)

Note: Standard errors are shown in parentheses.

UNCORRECTED PROOF

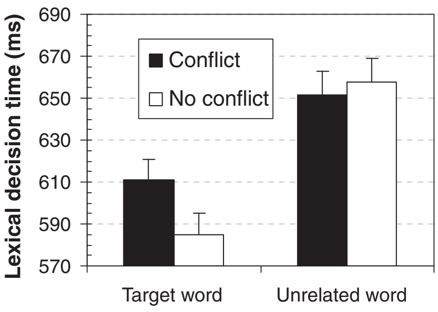

Fig. 1. Average lexical decision time for words that were related (i.e., targets) and unrelated to cued beliefs after solving conflict and no-conflict syllogisms. Error bars are standard errors.

access to words that were associated with conflicting beliefs was affected.

The above memory probing findings provide some of the first memory-based support for the postulation of a belief inhibition process during reasoning. However, they do not yet clarify the nature of an inhibition failure. Although average reasoning performance on the conflict problems was low, some participants did perform well. It might be suggested that these good reasoners are driving the observed effect. The crucial question with respect to the nature of the inhibition failure concerns the lexical decision performance of people who typically fail to solve the conflict problems correctly. To address this issue we compared the lexical decision findings of the best and worst scoring half of our participants (i.e., good and bad reasoners). If people typically err because they do not detect that their beliefs are erroneous and fail to initiate an inhibition process, then bad reasoners should not show the impaired lexical access after solving conflict problems. However, if everybody always engages in an inhibition process, then lexical access to target words after presentation of a conflict problem should be impaired whether or not the participant managed to solve the reasoning problems.

Based on a median split of the reasoning performance on the crucial conflict problems, participants who solved more than 50% of the conflict problems correctly were put in the good reasoners group (average score was 93%). Participant who scored 50% or less were put in the bad reasoners group (average score was 32%). This reasoning skill factor (bad vs. good reasoners) was entered as a between-subjects factor in the above 2 (problem type) × 2 (word type) ANOVA on the lexical decision times. Results were pretty straightforward. The skill factor, $F(1, 94) = 1.6$, $p = 2$, nor any of its interactions with the other factors in the design reached significance [Word × Skill, $F(1, 94) < 1$, Problem × Skill, $F(1, 94) = 2.4$, $p = .15$, Word × Problem × Skill, $F(1, 94) < 1$]. As Fig. 2 shows, both capacity groups clearly showed the same standard pattern with longer lexical decision times for target words after conflict problems had been solved. As Fig. 2 suggests, if anything, the increase even tended to be somewhat more pronounced for the bad reasoners.

The median split analysis gave us a powerful test to address the failure issue. However, in the bad reasoners group there were still some reasoners who solved some of the conflict problems correctly. Hence, an advocate of the inhibition-engagement-failure view might still argue that the engagement failure claim only concerns the very weakest group of reasoners who fail to solve any of the problems correctly. In this respect our "bottom half" selection criterion might have been too liberal. To eliminate such a confound we repeated the analysis with a smaller but more extreme capacity group. There were 18 participants in the present sample who failed to solve any of the conflict problems correctly. Lexical decision data for this group was compared with a group of 24 participants who solved all conflict problems correctly. However, results were completely consistent with the first analysis. Lexical decision times were not affected by reasoning skill [main effect Skill, $F(1, 40) = 1.85$, $p = .18$, Skill × Word, $F(1, 40) < 1$, Skill × Problem, $F(1, 40) = 3.53$, $p = .07$, Skill × Problem × Word, $F(1, 40) < 1$].

Finally, a correlational analysis also indicated that the observed impairment for the target words after solving conflict problems (i.e., lexical decision time for target words after conflict problems – lexical decision time for target words after no-conflict problems) did not depend on one's reasoning performance on the conflict syllogisms, $r(96) = -.19$, $p = .06$. If only good reasoners were to show the effect, the correlation should have been positive.

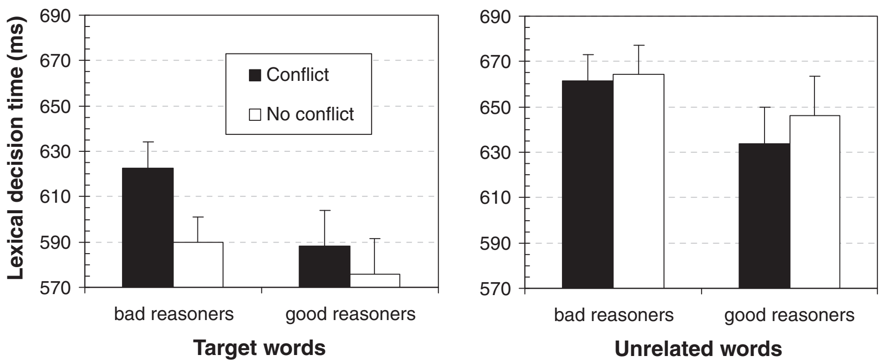

Fig. 2. Lexical decision times for the worst (bad reasoners) and best (good reasoners) scoring half of participants in Experiment 1. Error bars are standard errors.

2.3. Discussion

Consistent with the claim that people inhibit beliefs that conflict with logical knowledge during reasoning we observed that access to words associated with these beliefs was distorted after reasoning. When beliefs cued a response that was consistent with the logical status and inhibition was not required, lexical decisions for target words were made significantly faster than when beliefs and logic conflicted. All reasoners displayed this memory distortion after solving conflict problems. This suggests that even the poorest reasoners were at least trying to fight the biasing beliefs.

Given that we may assume that good reasoners are more successful at the inhibition, one might wonder why the observed distortion was not more pronounced for good than for bad reasoners. It is paramount to note here that our procedure only allows us to make a categorical claim about whether people engage in an inhibition process or not. If people engage in a belief discarding process, we can argue that they should show an impaired access to target words after solving conflict problems. However, the size of the impairment cannot be taken as measure of the extent or quality of the inhibition process. In essence, the memory inaccessibility is a negative by-product of the belief discarding process. It is possible, for example, that more gifted people pay a less severe price for the inhibition (e.g., accessibility is easier restored). Hence, the fact that good and bad reasoners show similar impairment does not necessarily imply that the inhibition was equally efficient or successful. The observed impairment does allow us to conclude that everyone at least engaged in an inhibition process. This implies that belief bias should not be attributed to a failure to engage an inhibition process but rather to a failure to complete it.

3. Experiment 2

In Experiment 2 we test whether our initial findings can be replicated with a different reasoning task. Participants in Experiment 2 were asked to solve problems that were modeled after Kahneman and Tversky's (1973) base-rate neglect problems.[2] Consider the following example:

> In a study 100 people were tested. Jo is a randomly chosen participant of this study. Among the 100 participants there were 5 men and 95 women.
> Jo is 23 years old and is finishing a degree in engineering. On Friday nights, Jo likes to go out cruising with friends while listening to loud music and drinking beer.
> What is most likely?
> a. Jo is a man
> b. Jo is a woman.

Given the size of the two groups in the sample, it will be more likely that a randomly drawn individual will be a woman. Normative considerations based on the group size or base-rate information cue response (b). However, many people will be tempted to respond (a) on the basis of stereotypical beliefs cued by the description. Just as in the deductive conflict problems in Experiment 1, normative considerations will conflict with our beliefs and sound reasoning requires inhibition of the compelling but erroneous belief-based response.

One can easily construct no-conflict or control versions of the base-rate problems. In the no-conflict version the description of the person will simply be composed of stereotypes of the larger group (e.g., De Neys & Glumicic, 2008; Ferreira, Garcia-Marques, Sherman, & Garrido, 2006). Hence, contrary to the classic problems, base-rates and description will not conflict and the response can be rightly based on the beliefs cued by the description without any need for inhibition.

3.1. Method

3.1.1. Participants

A total of 100 first-year psychology students from the University of Leuven (Belgium) participated in return for course credit. All participants were native Dutch speakers.

3.1.2. Material

Reasoning task: Participants solved a total of eight base-rate problems. Four of these were conflict problems in which the description of the person was composed of common stereotypes of the smaller population group tested (i.e., the description and the base-rates conflicted). In the four no-conflict problems the description and the base-rates agreed.

Problems were based on a wide range of stereotypes (e.g., involving gender, age, race). Descriptions were selected on the basis of an extensive pilot study (Franssens & De Neys, 2009). Selected descriptions for the conflict and no-conflict problems moderately but consistently cued one of the two groups. This point is not irrelevant. For convenience, we label responses that are in line with the base-rates as correct answers. However, if reasoners adopt a formal Bayesian approach (e.g., Gigerenzer, Hell, & Blank, 1988) and combine the base-rates with the diagnostic value of the description, this can lead to complications when the description is extremely diagnostic. Imagine that we have a sample of males and females and the description would state that the randomly drawn individual "is the pope of the catholic church". Now, by definition, no matter what the base-rates in the sample are, one would always need to conclude that the person is a man. We limited the impact of this problem by only selecting descriptions that were judged to have a moderate diagnostic value. By combining these with quite large base-rates (i.e., 95/100) one may generally conclude that the response that is cued by the base-rates should be selected if participants mange to refrain from giving too much weight to the intuitive beliefs cued by the description.

The order of the two response options ('a' and 'b') was counterbalanced. For half of the problems the correct response (i.e., the response consistent with the base-rates) was option 'a' whereas for the other half the second response option ('b') was the correct one. A complete overview of all eight problems can be found in the Appendix A.

[2] Syllogistic reasoning and base-rate task stem from two somewhat separated branches (i.e., the deductive reasoning branch and judgment and decision-making branch) of the psychology of thinking field. For convenience, we refer to both tasks as "reasoning" tasks.

UNCORRECTED PROOF

For the lengthy base-rate problems we used a slightly different presentation format than for the short syllogisms in Experiment 1. We tried to minimize the information that was presented at one time on the screen without altering the basic structure of the task. Hence, the general information on the first line of the problem (e.g., 'In a study 100 people were tested. Jo is a randomly chosen participant from this study.') was presented separately on the screen. When participants had read the sentences they pressed a key, and then the remaining part of the problem appeared. On average participants needed about 16 s (SD = 5.3) to solve the problems.

Lexical decision task: As in Experiment 1, after each problem a total of 24 letter strings was presented. Targets were core words that had been presented in the description or closely associated words. Material selection and presentation procedure was completely similar to Experiment 1. A complete overview of the selected words can be found in the Appendix A.

Note that in Experiment 1 we presented a different set of target words for conflict and no-conflict problems. We therefore established in a pilot study that there were no a priori lexical decision time differences for the two sets. The structure of the base-rate problems in Experiment 2 allowed us to control for possible word selection confounds more directly. Conflicting base-rate problems can be easily converted into no-conflict problems by switching the base-rates around. There is no need to alter the description and selected target words. Consequently, in Experiment 2, problems that were used as conflict problems for one half of the participants were used as no-conflict problems for the other half of the participants (and vice versa). Hence, the words in the lexical decision task were completely crossed. The exact same words that were used as targets for conflict problems for one half of the participants became targets for the no-conflict problems for the other half of the participants.

3.1.3. Procedure

As in Experiment 1, participants were tested in small groups and were first familiarized with the task-formats. Participants received the following instructions for the base-rate problems:

> In a big research project a number of studies were carried out where short personality descriptions of the participants were made. In every study there were participants from two population groups (e.g., carpenters and policemen). In each study one participant was drawn at random from the sample. You'll get to see the personality description of this randomly chosen participant. You'll also get information about the composition of the population groups tested in the study in question. You'll be asked to indicate to which population group the participant most likely belongs.

The eight base-rate problems were presented in random order. After each problem the corresponding lexical decision trials were presented. The procedure for the lexical decision task was completely similar to the one adopted in Experiment 1.

3.2. Results and discussion

Reasoning task: Reasoning performance on the base-rate problems replicated the findings in previous studies (e.g., De Neys & Glumicic, 2008). Participants seemed to neglect the base-rate information and erred on the vast majority of the conflict problems. On average, only 32% of the problems were solved correctly. However, as expected, people had far less difficulties when the stereotypical beliefs and base-rates pointed towards the same conclusion. Correct response rates on the no-conflict problems reached 96%, $F(1, 99) = 323.9$, $p < .0001$, $\eta_p^2 = .77$. No-conflict problems were also solved faster than conflict problems, $F(1, 99) = 10.55$, $p < .002$, $\eta_p^2 = .10$.

Lexical decision task: As in Experiment 1, lexical decision times were first submitted to a 2 (reasoning problem: conflict or no-conflict) × 2 (word type: target or unrelated) repeated measures ANOVA. As Fig. 3 shows, results replicated the findings of Experiment 1. Despite the quite low number of correct reasoning responses, overall people needed longer to identify words that were associated with cued beliefs after they had solved conflict problems, $F(1, 99) = 4.1$, $p < .05$, $\eta_p^2 = .05$. Lexical decision times for unrelated words did not differ, $F(1, 99) < 1$. As in Experiment 1, the effect of problem type and word type factors interacted, $F(1, 99) = 3.93$, $p < .05$, $\eta_p^2 = .04$. There was also a main effect of the word type factor, $F(1, 99) = 14.93$, $p < .001$, $\eta_p^2 = .13$, whereas the effect of the problem type factor itself was not significant, $F(1, 99) < 1$.

Next, the sample was split in two skill groups based on a median split of people's performance on the conflict problems. Participants who solved 50% or more of the conflict problems correctly were put in the high capacity group (average score was 74%). Participants who scored less than 50% were put in the low capacity group (average score was 10%). The reasoning skill factor (bad vs. good reasoners) was entered as a between-subjects factor in the above ANOVA. Results replicated the findings of Experiment 1. The skill factor, $F(1, 98) = 1.15$, $p = .28$, nor any of its interactions with the other factors reached significance [Word × Skill, $F(1, 98) < 1$, Problem × Skill, $F(1, 98) < 1$, Word × Problem × Skill, $F(1, 98) < 1$]. As Fig. 4 clarifies, the two capacity groups showed the same basic lexical decision impairment.

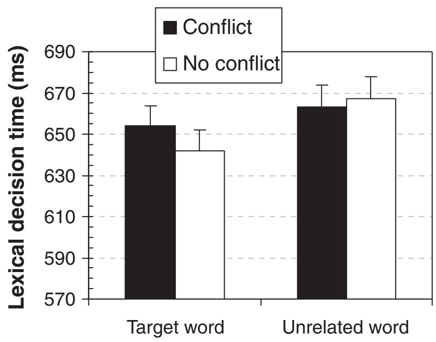

Fig. 3. Average lexical decision time for words that were related (i.e., targets) and unrelated to cued beliefs after solving conflict and no-conflict base-rate problems. Error bars are standard errors.

UNCORRECTED PROOF

Please cite this article in press as: De Neys, W., & Franssens, S. Belief inhibition during thinking: Not always winning but at least taking part. *Cognition* (2009), doi:10.1016/j.cognition.2009.07.009

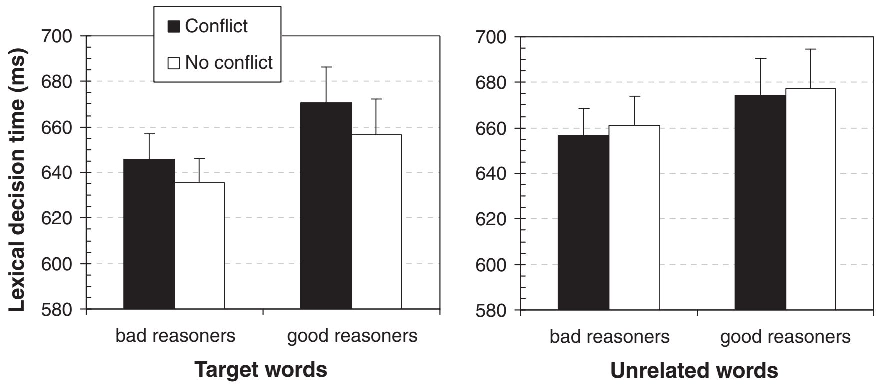

Fig. 4. Lexical decision times for the worst (bad reasoners) and best (good reasoners) scoring half of participants in Experiment 2. Error bars are standard errors.

We also repeated the analysis with more extreme skill groups. Thirty-nine participants failed to solve any of the conflict problems whereas 14 participants solved all of them correctly. However, as in Experiment 1, results were consistent with the median split analysis. Once again, the main effect of reasoning skill, $F(1, 51) < 1$, and its interactions with the other factors were not significant [Word × Skill, $F(1, 51) = 1.57$, $p = .22$, Problem × Skill, $F(1, 51) = 3.09$, $p = .09$, Word × Problem × Skill, $F(1, 51) < 1$].[3]

Finally, as in Experiment 1, a correlational analysis also indicated that the observed impairment for the target words after solving conflict problems (i.e., lexical decision time for target words after conflict problems – lexical decision time for target words after no-conflict problems) did not depend on one's reasoning performance on the conflict problems, $r(100) = .08$, $p = .44$.

Lexical decisions for syllogisms vs. base-rates: The pattern of lexical decision findings was consistent across the two experiments. For completeness, we also examined the impact of the reasoning task (syllogisms or base-rate problems) more directly by including it as a between-subjects factor in the 2 (problem type) × 2 (word type) ANOVA. Results showed that the main effect of Task, $F(1, 194) = 5.99$, $p < .025$, $\eta_p^2 = .03$, and its interaction with the Word factor, $F(1, 194) = 42.87$, $p < .0001$, $\eta_p^2 = .18$, were both significant. Simple effect tests indicated that lexical decision times for target words were overall faster after solving syllogisms than after solving base-rate problems, $F(1, 194) = 17.06$, $p < .0001$, $\eta_p^2 = .08$. Lexical decision times for unrelated words did not differ, $F(1, 194) < 1$. This finding makes sense if one takes into account that a simple syllogistic conclusion will prime the target words more strongly than the lengthier description in the base-rate problems. The crucial finding was that the type of reasoning task did not interact with the problem type, $F(1, 194) < 1$, or Problem Type × - Word Type interaction, $F(1, 194) = 1.92$, $p = .17$. A planned contrast established that the lexical decision time increase on the target words after solving conflict vs. no-conflict problems did not differ for the two types of reasoning tasks, $F(1, 194) = 2.58$, $p = .11$. Whether one solved syllogisms or base-rate problems, lexical decisions for target words took about 18 ms longer after solving the conflict problems.

A final analysis established that the median-split Skill factor, $F(1, 192) = 2.75$, $p = .1$, and its interactions with the other factors was also not affected by the type of reasoning task [Reasoning Task × Word × Skill, $F(1, 192) < 1$, Reasoning Task × Problem × Skill, $F(1, 192) = 1.68$, $p = .2$, Reasoning Task × Word × Problem × Skill, $F(1, 192) < 1$].[4] Planned contrasts showed that even when combing the two experiments and contrasting the performance of about 200 participants, the crucial lexical decision time increase on the target words after solving conflict problems did not differ for the best and worst group of reasoners [worst vs. best scoring half, $F(1, 192) = 1.01$, $p = .31$; all wrong vs. all correct, $F(1, 91) < 1$]. The worst scoring half of the participants, $F(1, 192) = 16.93$, $p < .0001$, $\eta_p^2 = .08$, and even participants who failed to solve any syllogism or base-rate problem correctly, $F(1, 91) = 5.39$, $p < .025$, $\eta_p^2 = .06$, still showed significantly longer lexical decision times after solving the conflict problems.

4. Experiment 3

The observed impaired access to target words in Experiment 1 and 2 supports the claim that all reasoners attempt to inhibit cued beliefs when they conflict with logical or probabilistic norms. However, inhibition refers to a *temporary* inaccessibility of stored information. When we inhibit information it does not stay inhibited forever. After a brief period of time the inhibition will start to fade out and the information will become accessible again. In Experiment 3 we focussed on this temporal characteristic of the inhibition process to validate our findings. Participants were pre-

[3] Since there were only 14 participants who never erred, we also contrasted the group who always erred with the best scoring half of reasoners. However, results were consistent [Skill, $F(1, 71) = 1.62$, $p = .21$, Word × Skill, $F(1, 71) = 1.36$, $p = .25$, Problem × Skill, $F(1, 71) = 2.67$, $p = .11$, Word × Problem × Skill, $F(1, 71) < 1$].

[4] Results were similar with the more extreme capacity groups of participants who failed or succeeded on all conflict problems [Reasoning Task × Skill, $F(1, 91) = 1.2$, $p = .28$, Reasoning Task × Word × Skill, $F(1, 91) < 1$, Reasoning Task × Problem × Skill, $F(1, 91) = 1.68$, $p = .2$, Reasoning Task × Word × Problem × Skill, $F(1, 91) < 1$].

UNCORRECTED PROOF

sented with the same reasoning problems and lexical decision task as in Experiment 1 and 2. The only difference was that after participants had entered their response for the reasoning problem, they did not start the lexical decision task immediately but were presented with a one-minute filler task (i.e., they solved easy math problems). After a one-minute delay the initially inhibited beliefs should become accessible again. If the impaired access to target words in Experiment 1 and 2 results from an inhibition process, the impairment should tend to disappear in Experiment 3.

4.1. Method

4.1.1. Participants

A total of 170 first-year psychology students from the University of Leuven (Belgium) participated in return for course credit. None of these participated in Experiment 1 or 2. All participants were native Dutch speakers. Lexical decision performance of participants in Experiment 1 and 2 was used as a baseline to test the impact of the delay factor.

4.1.2. Material

Reasoning tasks: Participants solved the same reasoning tasks as in Experiment 1 and 2. Half of the participants were presented with the syllogistic reasoning task whereas the other half solved the base-rate problems.

Lexical decision task: Participants were presented with the same lexical decision task as in Experiment 1 and 2. The only difference was that after participants had entered their response for the reasoning problem, they did not start the lexical decision task immediately but were presented with a one-minute filler task. In the filler task participants were asked to solve easy math problems (e.g., $(9 \times 3) + 2 = ?$).

4.1.3. Procedure

As in Experiment 1 and 2, participants were tested in small groups and were first familiarized with the task-formats. Participants practiced the lexical decision and filler task and were told that the tasks would alternate in the actual experiment. Remaining instructions and procedure were completely similar to Experiment 1 and 2.

4.2. Results and discussion

Reasoning tasks: As Table 1 shows, reasoning performance in Experiment 3 was in line with the previous experiments. Accuracy on the conflict, $F(1, 179) = 2.1$, $p = .15$, and no-conflict syllogisms, $F(1, 179) < 1$, did not differ from the syllogistic performance in Experiment 1. Likewise, conflict, $F(1, 183) < 1$, and no-conflict base-rate problems, $F(1, 183) < 1$, were solved equally well with and without delay. Response times on the conflict, $F(1, 179) < 1$, and no-conflict syllogisms, $F(1, 179) < 1$, and conflict $F(1, 183) < 1$, and no-conflict base-rate problems, $F(1, 183) = 2.57$, $p = .12$, were also not affected by the delay. This clearly establishes that the inclusion of the filler task did not alter reasoning performance per se.

Lexical decision task: Lexical decision times were submitted to a 2 (problem type: conflict or no-conflict) × 2 (word type: target or unrelated) × 2 (delay: filler task or no-filler task) × 2 (reasoning task: syllogisms or base-rates) ANOVA. This design partially repeats the analysis in Experiment 1 and 2. We focus here on the crucial effect of the delay factor. We tested the key effects of interest with planned contrasts.

As Fig. 5 shows, results supported the inhibition account. After a one-minute delay accessing belief-related target words did no longer take more time for conflict than for no-conflict problems, both when solving syllogisms, $F(1, 362) < 1$, and base-rate problems, $F(1, 362) = 1.58$, $p = .21$. Fig. 5 further clarifies that the delay tended to increase the lexical decision time for target words of no-conflict problems, whereas lexical decisions for the target words of conflict problems showed the opposite trend and tended to speed-up after the delay. This interaction was overall significant, $F(1, 362) = 7.22$, $p < .01$, $\eta_p^2 = .02$, and did not differ for the two types of reasoning tasks, $F(1, 362) < 1$. The longer lexical decision times on the no-conflict problems after the delay are not surprising given that the delay will result in less efficient priming. After one-minute, lexical decisions will benefit less from the initial cueing of the beliefs. However, on the conflict problems we predicted that the access to cued beliefs was initially inhibited. Since the inhibition should only be temporary in nature, access will start to be restored and lexical decisions will consequently benefit from the delay.

The observed pattern helps us to discard a possible alternative explanation for the findings in Experiment 1 and 2. One could argue that because conflict problems are more complex than no-conflict problems, people will always engage in some additional processing after reading the preambles of the conflict problems. Whatever the nature of this additional processing might be, it will already result in some delay between the initial cueing of the beliefs and the lexical decision task. This delay could lead to a less efficient priming of target words for conflict problems and consequently explain the longer lexical decision times without any need to postulate an inhibition process. Experiment 3 discards this account. If less efficient priming after solving conflict problems were to explain the impairment findings of Experiment 1 and 2, the additional delay in Experiment 3 should result in even more impaired lexical decision times. The inhibition account, however, specifically predicts that after the delay from the filler task, the initially blocked beliefs should become accessible again. Therefore, accessing target words for conflict problems should be faster and not slower after the delay. The fact that the delay tended to speed-up the lexical decisions for conflict problems establishes that the memory access was initially distorted because of an inhibition process.

For completeness, we also examined the impact of the delay on the unrelated words. Planned contrast established that contrary to the target words, the delay impact on unrelated words did not differ for conflict and no-conflict problems, neither when solving base-rates, $F(1, 362) < 1$, nor syllogisms, $F(1, 362) < 1$. The only indication for an impact of the delay on the unrelated words was that when solving syllogisms, lexical decisions seemed to be overall

UNCORRECTED PROOF

Please cite this article in press as: De Neys, W., & Franssens, S. Belief inhibition during thinking: Not always winning but at least taking part. *Cognition* (2009), doi:10.1016/j.cognition.2009.07.009

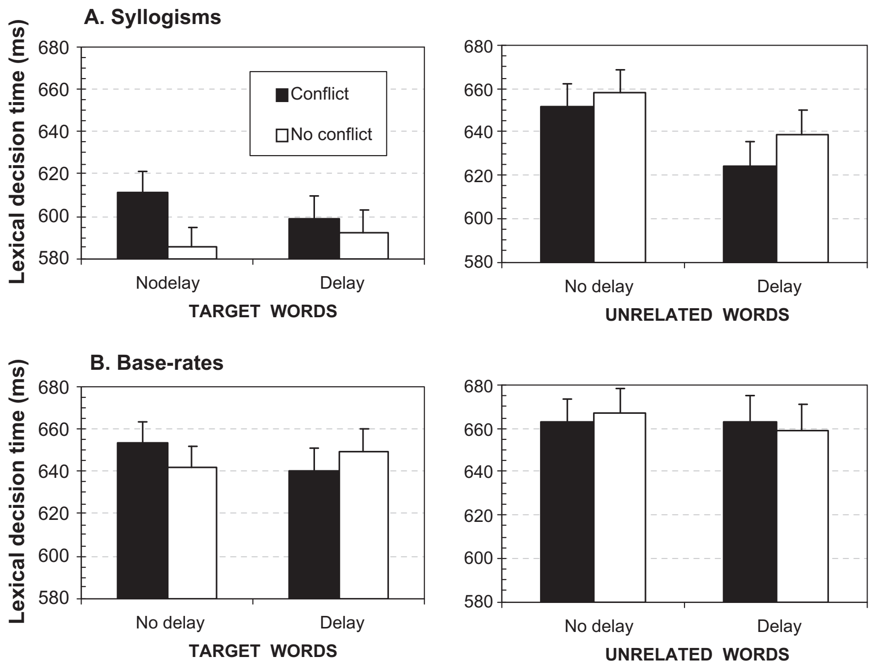

Fig. 5. The impact of a one-minute delay on the average lexical decision times after solving syllogisms (top panel) and base-rate problems (bottom panel). Error bars are standard errors.

somewhat faster after the delay. However, this trend did not reach significance, F(1, 362) = 2.51, $p = .11$. Hence, as one might expect, the delay had no impact on the accessibility of information that had not been cued initially.

A final analysis established that the impact of the delay did not differ for good and bad reasoners. Results showed that the crucial speeding-up of the lexical decisions for conflict problems and slowing-down for no-conflict problems after the delay did not differ for the worst and best scoring half of the participants [Syllogisms, F(1, 358) < 1; Base-rates, F(1, 358) = 2.59, $p = .11$; Combined, F(1, 358) < 1] or participants who solved none or all of the conflict problems correctly [Syllogisms, F(1, 173) < 1; Base-rates, F(1, 173) = 1.23, $p = .26$; Combined, F(1, 173) < 1].

5. Experiment 4

Experiment 3 established that the observed memory impairment in Experiment 1 and 2 was only temporary in nature. In Experiment 4 we validated the findings further by changing the nature of the reasoning task. We tried to eliminate the tendency to engage in an inhibition process by explicitly instructing participants to respond rapidly and select the response that seemed intuitively most plausible. Under these intuitive thinking instructions, there is no longer any need to inhibit the cued beliefs and consequently access to the target words should simply not become impaired. If the longer lexical decision times after solving conflict problems in Experiment 1 and 2 result from the postulated inhibition process, we should no longer observe them under the intuitive instructions in Experiment 4.

5.1. Method

5.1.1. Participants

A total of 178 first-year psychology students from the University of Leuven (Belgium) participated in return for course credit. None of these participated in the previous experiments. All participants were native Dutch speakers. Lexical decision performance of participants in Experiment 1 and 2 was used as a baseline to test the impact of the instruction factor.

5.1.2. Material

Reasoning tasks: Participants were presented with the same items as in Experiment 1 and 2. About half of the participants were presented with the syllogisms ($n = 85$) whereas the others were presented with the base-rate problems ($n = 93$). Instructions and task-format were modified to cue mere belief-based thinking.

Syllogisms: The task was introduced to participants as a pilot study in which the believability of a number of statements needed to be evaluated. Any references to logical reasoning in the task instructions were avoided. Participants were told that they would see short stories consisting of three sentences and simply needed to indicate whether they believed the final sentence or not. The two response alternatives were rephrased as "1. The sentence is believable" and "2. The sentence is not believable". Instructions stressed that we were "interested in people's initial response and did not want participants to think too long about their response". Previous studies indicated that some participants spontaneously engage in logical reasoning when presented with conditional syllogisms,

Please cite this article in press as: De Neys, W., & Franssens, S. Belief inhibition during thinking: Not always winning but at least taking part. *Cognition* (2009), doi:10.1016/j.cognition.2009.07.009

even when they are not explicitly instructed to do so (e.g., De Neys, Schaeken, & d'Ydewalle, 2005). The present task modifications minimized such a possible confound.

Base-rates: The task was introduced as a study on "gut feelings". Participants were given the general task instructions as in Experiment 2 but were asked to respond rapidly and select the response that seemed intuitively most plausible. Instructions again stated explicitly that we were "interested in people's initial response and did not want participants to think too long about their response".

Lexical decision task: Participants were presented with the same lexical decision task as in Experiment 1 and 2.

5.1.3. Procedure

Except for the specific reasoning task instructions the procedure was completely similar to Experiment 1 and 2.

5.2. Results and discussion

Reasoning tasks: Accuracy and response latencies established that the instruction manipulation was successful. As expected, participants gave overall more belief-based responses under intuitive thinking instructions in Experiment 4 than under standard instructions in Experiment 1, $F(1, 179) = 45.6$, $p < .0001$, $\eta_p^2 = .20$, and Experiment 2, $F(1, 191) = 3.08$, $p < .085$, $\eta_p^2 = .02$. Both for the syllogisms, $F(1, 179) = 128.51$, $p < .0001$, $\eta_p^2 = .42$, and base-rate problems, $F(1, 191) = 3.08$, $p < .085$, $\eta_p^2 = .02$, this tendency was more pronounced on the conflict than on the no-conflict problems. As Table 1 indicates, participants hardly ever gave the original "correct"[5] logical or base-rate response on the conflict problems when instructed to reason intuitively. Overall, responses were also given faster under intuitive thinking instructions in Experiment 4 than under standard instructions in Experiment 1, $F(1, 179) = 34.22$, $p < .0001$, $\eta_p^2 = .16$, and Experiment 2, $F(1, 179) = 4.87$, $p < .03$, $\eta_p^2 = .03$. These faster responses were equally clear for conflict and no-conflict problems, both for syllogisms, $F(1, 179) < 1$, and base-rate problems, $F(1, 191) < 1$. The trends towards faster and more frequent belief-based responses indicate that participants indeed engaged in a more intuitive type of thinking.

Lexical decision task: Lexical decision times were submitted to a 2 (problem type: conflict or no-conflict) × 2 (word type: target or unrelated) × 2 (instructions: standard or intuitive) × 2 (reasoning task: syllogisms or base-rates) ANOVA. This design partially repeats the analysis in Experiment 1 and 2. We focus here on the crucial effect of the instruction factor. We tested the key effects of interest with planned contrasts.

As Fig. 6 shows, results supported the inhibition account. When people were reasoning intuitively and did not need to engage in an inhibition process, accessing belief-related target words immediately after the reasoning task did no longer take more time for conflict than for no-conflict problems, both when solving syllogisms, $F(1, 370) < 1$, and base-rate problems, $F(1, 370) < 1$. As Fig. 6 indicates, this effect resulted from a speeding-up of the lexical decisions for conflict problems and a slight slowing-down for no-conflict problems under intuitive thinking instructions. This interaction effect was overall significant, $F(1, 370) = 5.48$, $p < .025$, $\eta_p^2 = .02$, and did not differ for the two types of reasoning tasks, $F(1, 370) < 1$.

As expected, contrary to the target words, the instruction impact on unrelated words did not depend on whether participants had solved conflict or no-conflict problems, neither for syllogistic, $F(1, 370) < 1$, nor base-rate problems, $F(1, 370) < 1$. The only indication for an impact of the instructions on the unrelated words was a small trend towards faster lexical decisions under intuitive thinking instructions when solving syllogisms, but the effect was not significant, $F(1, 370) < 1$. As one might expect, this indicates that taking away the need to engage in belief inhibition when dealing with conflict problems does not affect the accessibility of unrelated words.

Note that Experiment 4 helps us to rule out another specific alternative account for our initial findings. One might suggest that the observed memory impairments in Experiment 1 and 2 did not result from an active, thinking-related belief inhibition process but rather from a more basic encoding process related to the inability to form a coherent representation when reading the problems. That is, the observed effects might be explained by processes that are independent of whether or not a subject uses this information to draw a conclusion. For example, while reading the base-rate information (e.g., study with 5 men and 95 women) people might start to activate stereotypes associated with the largest group because they expect to read a description that is consistent with it. When the description subsequently contradicts this expectation the simultaneous activation of these two conflicting representations (e.g., of a man and a woman) might result in some interference.[6] Hence, the point is that it might be the presence of such incoherent representations during encoding that drives the observed memory impairments in our experiments and not the type of thinking-related belief inhibition process that reasoning theories typically envisage. Experiment 4 argues against this alternative encoding account. Participants were presented with the exact same base-rates and descriptions as in our first experiments. Hence, at the more basic encoding level the representation formation processes will keep on cueing conflicting representations when reading them. However, under intuitive instructions there was no longer any need to prevent belief-based reasoning and engage in the more active belief inhibition process that is postulated by the reasoning community. Hence, if the longer lexical decision times in Experiment 1 and 2 merely resulted from encoding interference during reading and not from the postulated thinking-related inhibition process, we should still

[5] For consistency we keep on referring to the logical and base-rate response in Experiment 4 as correct responses.

[6] Note that the alternative encoding account is far less appealing for the syllogistic problems. One might argue that reading unbelievable conclusions per se results in encoding of conflicting representations (i.e., the conclusion would conflict with what is expected on the basis of semantic knowledge). However, since conclusion believability was crossed with problem type this factor cannot account for the observed difference between conflict and no-conflict problems.

UNCORRECTED PROOF

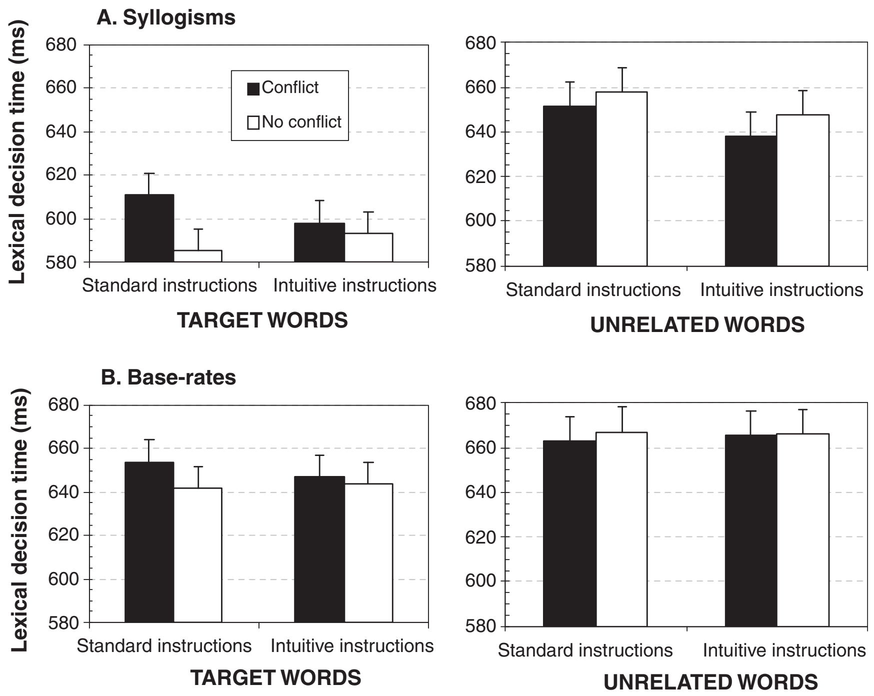

Fig. 6. The impact of the explicit instruction to think intuitively on lexical decision times after solving syllogisms (top panel) and base-rate problems (bottom panel). Error bars are standard errors.

have observed the effect under the intuitive instructions in Experiment 4.

6. General discussion

Probing people's memory for beliefs that were cued during reasoning provided direct evidence for the postulation of a belief inhibition process during thinking. Consistent with the claim that people discard beliefs that conflict with more normative considerations during reasoning, we observed that access to words associated with these beliefs was distorted after reasoning: When beliefs cued a response that conflicted with the appropriate logical or probabilistic response, lexical decisions for target words associated with the cued beliefs took significantly more time than when beliefs and normative considerations did not conflict and inhibition was not required. The study further established that the impairment was only temporary in nature and did not occur when people were explicitly instructed to give mere intuitive judgments.

All reasoners displayed the crucial memory distortion. Even the poorest reasoners in our sample needed more time to access the belief-related target words after solving conflict problems. This clarifies that the widespread belief bias we observed does not result from a failure to initiate an inhibition process but rather from a failure to complete it. As noted, these results help to sketch a less bleak picture of human rationality. If people were biased because they did not detect that their beliefs were not warranted and failed to initiate an inhibition process, memory access to the cued beliefs should not have been distorted. Hence, the present accessibility findings establish that people are far more logical than their answers suggest. Although people's judgments are often biased they are no mere intuitive, illogical thinkers who disregard normative considerations. All reasoners try to discard beliefs that conflict with normative considerations. The problem is simply that not everyone manages to complete the process.

The inhibition findings have important implications for the status of logic and probability theory as normative standards. Faced with the omnipresence of belief bias some authors have questioned the validity of these norms (e.g., Oaksford & Chater, 2007; Todd & Gigerenzer, 2000). Bluntly put, it was argued that if the vast majority of well-educated, young adults fail to solve a simple reasoning task, this might indicate that there is something wrong with the task scoring norm rather than with the participants. The basic point of these authors was that people might interpret the tasks differently and adhere to other norms than the classic ones. This debate has raged through the field for decades without clear solution (e.g., Stein, 1996). Clarifying the nature of an inhibition failure helps to break the stalemate. The fact that people tried to block the intuitive beliefs when they conflicted with the traditional norms not only implies that people know the norms but also that they judge them to be relevant. If people did not believe that base-rates or logical validity mattered, they would not waste time trying to block the conflicting response. People might not always manage to adhere to the norm but they are at least trying to and are clearly not simply discarding it or treating it as irrelevant. This should at least give pause for thought before rejecting the validity of the traditional norms.

UNCORRECTED

Please cite this article in press as: De Neys, W., & Franssens, S. Belief inhibition during thinking: Not always winning but at least taking part. *Cognition* (2009), doi:10.1016/j.cognition.2009.07.009

The present memory-based behavioural findings allow us to complement the growing number of brain-imaging studies on the neural substrate of belief bias. As we noted, overcoming belief bias has been shown to result in an increased activation of the lateral prefrontal brain-areas (e.g., De Martino et al., 2006; Goel & Dolan, 2003; Prado & Noveck, 2007; Sanfey et al., 2003). The memory-accessibility data lend credence to the idea that the recruitment of these areas actually reflects the operation of a belief inhibition process. In addition, our data imply that the less clear activation of these lateral prefrontal areas when people are biased needs to be attributed to the incomplete nature of this inhibition process.

Our findings also validate a recent imaging study that monitored the activation of a more medial frontal brain-area (i.e., the anterior cingulate cortex) believed to be involved in conflict detection (De Neys et al., 2008). De Neys et al. showed that this medial "conflict detection area" was always activated when people were trying to solve reasoning problems, even when people were biased by their beliefs and failed to select the correct response in the end. De Neys et al. argued that this finding indicated that people always detected that their belief-based response was erroneous and conflicted with the normative considerations (see also De Neys & Glumicic, 2008). The present findings support this claim. If people were not detecting the conflict first, they would also see no need to initiate an inhibition process. The present findings clarify, however, that people do not simply stop at detecting the conflict. People also try to do something about it and start fighting the inappropriate beliefs. This point is important with respect to the debate on the validity of the classic norms. Successful conflict detection per se does not suffice to establish that people are also adhering to the norm. An advocate of the invalidity view could rightly argue that knowing that a response conflicts with some norm does not imply that you also belief that the norm is appropriate or should be respected. A psychopath, for example, might also know that murder conflicts with moral standards. The problem is that he does not feel any intention to adhere to these norms. The finding that people are trying to fight the conflicting beliefs clarifies that people are no rational psychopaths and intend to adhere to the logical norm.

Our lexical decision findings were consistent across the two reasoning tasks we presented. We specifically selected the syllogistic reasoning and base-rate task because of the central role they play in the reasoning and decision making field. The replication of the findings across these popular tasks lends credence to the generality of the results. However, it should be clear that the reasoning and decision making fields study hundreds of tasks and numerous variants of one and the same task. Hence, some caution is needed when drawing general conclusion from the present study. Obviously, people might face other difficulties in other tasks (Stanovich & West, 2008). We do believe that the study more broadly serves as a key illustration of the importance of introducing processing measures (i.e., measure that clarify "how" people are arriving at an answer) in the psychology of thinking. It has been argued that a general shortcoming of classic reasoning and decision-making research, as well as the central debate on human rationality, is that scholars have almost exclusively focused on people's response accuracy (i.e., whether or not people manage to give the correct response) and not on the underlying cognitive processes (De Neys, 2009; Hertwig & Gigerenzer, 1999; Hoffrage, 2000; Reyna, Lloyd, & Brainerd, 2003). The present study demonstrates how this approach is bound to bias any conclusions about human rationality or the validity of classic logical norms. Looking at how people are arriving at an erroneous response sketches a more optimistic picture of the human reasoning machinery. Our data clearly indicate that people can be far more normative than their answers suggest. Although we might not always win the inhibition struggle and avoid belief bias, we do seem to know that we are being biased and try to fight the unwarranted beliefs.

It will be clear that the present findings raise some interesting questions for further study. For example, our key finding was that after a conflict between beliefs and normative considerations memory access to information associated with the cued heuristic beliefs was impaired. However, one might also wonder what happens with the information that is associated with the normative considerations (e.g., the base-rates) in these cases. One possibility is that this information becomes more accessible. Consistent with this idea, De Neys and Glumicic (2008) already observed that people have little difficulty in recalling the base-rate information of conflict problems after they finish the reasoning task. The present methodology could be used to test this idea more directly by examining the lexical decision times for cued normative information. Likewise, one might wonder why people do only inhibit their beliefs in case of a conflict. In theory, one could always block belief-based reasoning and rely on mere logical reasoning. This point underscores the fact that the human reasoning engine respects the principle of cognitive economy (e.g., Evans, 2008). It is well-established that belief-based reasoning is much less demanding than logical thinking (e.g., De Neys, 2006a, 2006b). Hence, simply inhibiting one's beliefs throughout would be quite costly and inefficient. If we are not to waste scarce cognitive resources, overriding beliefs needs to be restricted to the conflict cases. This does imply that it is paramount that reasoners monitor for such a conflict. As we clarified, the fact that people always initiate an inhibition process in case of a conflict implies that reasoners are doing this and are remarkably good at it too. One might remark that the quite flawless nature of the monitoring in turn suggests that it cannot be very demanding. We simply want to note here that Franssens and De Neys (2009) recently presented direct empirical evidence that supports this idea (a further discussion of more theoretical implications can be found in De Neys & Glumicic, 2008).

A last comment we want to make is related to the status of the inhibition concept in memory research. As we pointed out in the introduction, the ultimate origin of an observed temporary inaccessibility of a memory trace is still debated by memory researchers. It is not clear whether it results from a literal deactivation of the memory trace at the neural level or from a competition between competing responses after one of them has been flagged as inappropriate (see MacLeod et al., 2003, for a review). In

UNCORRECTED PROOF

Please cite this article in press as: De Neys, W., & Franssens, S. Belief inhibition during thinking: Not always winning but at least taking part. *Cognition* (2009), doi:10.1016/j.cognition.2009.07.009

the present study we made no claims about this issue. As we noted, both conceptualizations share the general idea that at some level the information is being disregarded (i.e., it is not having it is normal impact on our behaviour). It is this process that reasoning and decision-making researchers have typically subsumed under the general header "belief inhibition". The present study demonstrates for the first time that we find the hallmark memory trace of such a discarding process after solving conflict problems (i.e., access to belief-related knowledge is distorted after solving conflict problems). However, just as in the memory field, the adopted methodology does not allow us to specify the exact origin of the observed memory impairment. We cannot conclude whether belief-related target words were flagged as inappropriate, whether their activation threshold was literally deactivated, or whether, as one reviewer suggested, people undermined their beliefs after conflict detection and attached a higher degree of uncertainty to them. Note, however, that the ultimate origin of the memory impairment is not the crucial issue here. The different accounts would point to the exact same bold conclusions for the rationality debate. Let's say that our reviewer is right and people undermine their beliefs and become less certain about them after solving conflict problems. The higher associated uncertainty would then distort subsequent memory access. The fact that people start to question their beliefs would still be prima facie evidence for the claim that they detect the conflict and try to do something about it. If people were not to believe that the classic norms were relevant, there would be no reason whatsoever to start questioning their intuitive beliefs and attach more uncertainty to them. Hence, the point we want to stress is that whether people literally inhibit their beliefs, label them as inappropriate, or become less certain about them does not affect the crucial conclusions for the reasoning field. Of course, this does not imply that such a more fine-grained future clarification of the memory mechanism (e.g., literal neural deactivation or not) behind the belief inhibition phenomenon is useless. What matters at this stage, however, is that just as in the memory field, we can provide basic evidence for the claim that information has been disregarded during thinking in the first place. It is this crucial evidence that the present study looked for and found.

Acknowledgement

Wim De Neys's research is funded by the Flemish Fund for Scientific Research (FWO-Vlaanderen).

Appendix A. Overview of the reasoning problems and selected target and unrelated words (translated from Dutch)

A.1. Syllogisms (Experiment 1 – 3 – 4)

Conflict problems

1.
All flowers need water
Roses need water
Roses are flowers

Target words: rose, petal, garden, flower, plant, bush
Unrelated words: wolf, competition, date, stone, axe, cooked

2.
All things with an engine need oil
Cars need oil
Cars have engines

Target words: car, steer, drive, engine, train, fire
Unrelated words: smart, annoying, tea, slum, mint, wheat

3.
All mammals can walk
Whales are mammals
Whales can walk

Target words: whale, dolphin, ocean, run, marathon, walk
Unrelated words: firm, head, enough, story, flexible, rattle

4.
All vehicles have wheels
A boat is a vehicle
A boat has wheels

Target words: boat, canal, ship, wheel, drive, tire
Unrelated words: circle, forever, curve, night, pants, people

No-conflict problems

5.
All things that are smoked are bad for your health
Cigarettes are smoked
Cigarettes are bad for your health

Target words: cigarette, smoke, cancer, health, doctor, ill
Unrelated words: ball, optimum, monastery, tender, difference, sketch

6.
All African countries are warm
Spain is warm
Spain is an African country

Target words: Spain, sea, beach, Africa, sun, lion
Unrelated words: telephone, shoe, hole, joke, spoon, bed

7.
All meat products can be eaten
Apples can be eaten
Apples are meat products

Target words: apple, pear, fruit, meat, food, cow
Unrelated words: child, cloud, idol, psychologist, elite, fashion

Please cite this article in press as: De Neys, W., & Franssens, S. Belief inhibition during thinking: Not always winning but at least taking part. *Cognition* (2009), doi:10.1016/j.cognition.2009.07.009

8.
All birds have wings
Crows are birds
Crows have wings

Target words: crow, raven, black, wing, fly, feathers
Unrelated words: war, alphabet, calf, aniseed, room, video

A.2. Base-rate problems (Experiment 2–4)

Conflict problems

1.
In a study 100 people were tested. Among the participants there were five people who drive a used Nissan and 95 people who drive a BMW. Etienne is a randomly chosen participant of the study.

Etienne is 38 years old. He works in a steel plant. He lives in a small apartment in the outskirts of Charleroi. His wife has left him.

What is most likely?

Etienne drives a BMW.
Etienne drives a used Nissan.

Target words: factory, apartment, abandoned, machine, alone, lonely
Unrelated words: issue, ridiculous, proposal, welcome, speech, opt

2.
In a study 100 people were tested. Among the participants there were five sixteen-year-olds and 95 forty-year-olds. Els is a randomly chosen participant of the study.

Els likes to listen to techno and electro music. She often wears tight sweaters and jeans. She loves to dance and has a small nose piercing.

What is most likely?

Els is 16 years old.
Els is 40 years old.

Target words: techno, dance, party, jeans, drugs, feast
Unrelated words: ready, ring, humour, go, hand, rumour

3.
In a study 100 people were tested. Among the participants there were 95 Swedes and five Italians. Mario is a randomly chosen participant of the study.

Mario is 25 years old. He is a charming young man and is a real womanizer. His favourite dish is the spaghetti his mother makes.

What is most likely?

Mario is a Swede.
Mario is an Italian.

Target words: charming, seduce, spaghetti, handsome, sweet, macaroni
Unrelated words: bathroom, diagnosis, weight, month, activity, strike

4.
In a study 100 people were tested. Among the participants there were 95 Muslims and five Buddhists. Sarah is a randomly chosen participant of the study.

Sarah is 19 years old. She likes to philosophize and she hates materialism. She wears second-hand clothes and would love to go to India one day.

What is most likely?

Sarah is a Buddhist.
Sarah is a Muslim.

Target words: philosopher, India, wisdom, China, second-hand, religion
Unrelated words: deviation, episode, participant, very, parade, hear

No-conflict problems

5.
In a study 100 people were tested. Among the participants there were 95 people who like to watch Canvas and five people who like to watch VTM. Aline is a randomly chosen participant of the study.

Aline is 35 years old. She writes reviews for a magazine. Her husband works at the university. She loves painting and photography.

What is most likely?

Aline likes to watch Canvas.
Aline likes to watch VTM.

Target words: magazine, paint, photography, newspaper, movie, illustration
Unrelated words: goal, favourite, attainable, attempt, medical, assignment

Note: VTM is a popular, commercial ("Fox"-like) Flemish TV channel. Canvas is a more educational, publicly-funded ("PBS"-like) channel.

(continued on next page)

6.
In a study 100 people were tested. Among the participants there were 95 people who live in the country and five people who live in the city. Debby is a randomly chosen participant of the study.

Debby is 22 years old. She rides a horse. After school she takes care of the animals at home. In the weekends she rises early and visits her grandparents.

What is most likely?

Debby lives in the country.
Debby lives in the city.

Target words: horse, nurse, cattle, grandparent, grassland, elderly
Unrelated words: jury, father, sophisticated, call, pull, felt-tip

7.
In a study 100 people were tested. Among the participants there were five people who vote for the green party and 95 people who vote for the Flemish Interest party. Jeanine is a randomly chosen participant of the study.

Jeanine is 67 years old. She worked as an assembly line packer. She believes that traditional values are important and lives in an area where there's a lot of crime.

What is most likely?

Jeanine votes for Flemish Interest
Jeanine votes for the green party

Target words: assembly, grind, crime, register, wrap, boring
Unrelated: intention, population, convention, breakthrough, record, hope

Note: Flemish Interest is a conservative, anti-immigrant, far right party.

8.
In a study 100 people were tested. Among the participants there were 5 women and 95 men. Dominique is a randomly chosen participant of the study.

Dominique is 32 years old and is a self-confident and competitive person. Dominique's goal is building a career. Dominique does a lot of sport and is well-muscled.

What is most likely?

Dominique is a woman.
Dominique is a man.

Target words: self-confident, career, muscled, job, power, strong
Unrelated words: tempo, paste, episode, sandal, system, corn

References

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences, 8*, 170–177.
De Deyne, S., & Storms, G. (2008). Word Associations: Norms for 1424 Dutch words in a continuous task. *Behavior Research Methods, 40*, 198–205.
De Martino, B., Kumaran, D., Seymour, B., & Dolan, R. J. (2006). Frames, biases, and rational decision making in the human brain. *Science, 313*, 684–687.
De Neys, W. (2006a). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science, 17*, 428–433.
De Neys, W. (2006b). Automatic-heuristic and executive-analytic processing in reasoning: Chronometric and dual task considerations. *Quarterly Journal of Experimental Psychology, 59*, 1070–1100.
De Neys, W. (2009). Beyond response output: More logical than we think. *Behavioral and Brain Sciences, 32*, 87–88.
De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition, 106*, 1248–1299.
De Neys, W., Schaeken, W., & d'Ydewalle, G. (2005). Working memory and everyday conditional reasoning: Retrieval and inhibition of stored counterexamples. *Thinking and Reasoning, 11*, 349–381.
De Neys, W., Vartanian, O., & Goel, V. (2008). Smarter than we think: When our brains detect that we are biased. *Psychological Science, 19*, 483–489.
De Neys, W., & Verschueren, N. (2006). Working memory capacity and a notorious brain teaser: The case of the Monty Hall Dilemma. *Experimental Psychology, 53*, 123–131.
Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., & Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and human Decision Processes, 105*, 98–121.
Evans, J. St. B. T. (2003). In two minds: Dual process accounts of reasoning. *Trends in Cognitive Sciences, 7*, 454–459.
Evans, J. St. B. T. (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking and Reasoning, 13*, 321–329.
Evans, J. St. B. T. (2008). Dual-processing accounts of reasoning, judgement and social cognition. *Annual Review of Psychology, 59*, 255–278.
Ferreira, M. B., Garcia-Marques, L., Sherman, S. J., & Garrido, M. (2006). Automatic and controlled components of judgment under uncertainty. *Proceedings of the Cognitive Science Society, 28*, 1293–1298.
Franssens, S., & De Neys, W. (2009). The effortless nature of conflict detection during thinking. *Thinking and Reasoning, 15*, 105–128.
Gigerenzer, G., Hell, W., & Blank, H. (1988). Presentation and content: The use of base-rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 513–525.
Goel, V., & Dolan, R. J. (2003). Explaining modulation of reasoning by belief. *Cognition, 87*, B11–B22.
Hertwig, R., & Gigerenzer, G. (1999). The 'conjunction fallacy' revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making, 12*, 275–305.
Hoffrage, U. (2000). Why the analyses of cognitive processes matter. *Behavioral and Brain Sciences, 23*, 679–680.
Houdé, O. (1997). Rationality in reasoning: The problem of deductive competence and the inhibitory control of cognition. *Current Psychology of Cognition, 16*, 108–113.
Houdé, O. (2007). First insights on "neuropedagogy of reasoning". *Thinking and Reasoning, 13*, 81–89.
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.
Kahneman, D. (2002). Maps of bounded rationality: A perspective on intuitive judgement and choice. *Nobel Prize Lecture*. Retrieved January 11, 2006, from http://nobelprize.org/nobel_prizes/economics/laureates/2002/kahnemann-lecture.pdf [December].

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237–251.

MacLeod, C. M., Dodd, M. D., Sheard, E. D., Wilson, D. E., & Bibi, U. (2003). In opposition to inhibition. In B. H. Ross (Ed.). *The psychology of learning and motivation* (Vol. 43, pp. 163–206). San Diego, CA: Academic Press.

Markovits, H., & Nantel, G. (1989). The belief bias effect in the production and evaluation of logical conclusions. *Memory and Cognition, 17*, 11–17.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence upon retrieval operations. *Journal of Experimental Psychology, 90*, 227–234.

Neill, W. T. (1997). Episodic retrieval in negative priming and repetition priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 1291–1305.

Newstead, S. E., Handley, S. J., Harley, C., Wright, H., & Farrelly, D. (2004). Individual differences in deductive reasoning. *Quarterly Journal of Experimental Psychology, 57A*, 33–60.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford, UK: Oxford University Press.

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

Reyna, V. F., Lloyd, F. J., & Brainerd, C. J. (2003). Memory, development, and rationality: An integrative theory of judgement and decision-making. In S. Schneider & J. Shanteau (Eds.), *Emerging perspectives on judgment and decision research*. New York: Cambridge University Press.

Sá, W., West, R. F., & Stanovich, K. E. (1999). The domain specificity and generality of belief bias: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology, 91*, 497–510.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision making in the ultimatum game. *Science, 300*, 1755–1758.

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

Standage, T. (2005). *Bad to the last drop*. New York Times Editorial. Retrieved January 31, 2008, from http://www.nytimes.com/2005/08/01/opinion/01standage.html [August].

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences, 23*, 645–726.

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology, 94*, 672–695.

Stein, E. (1996). *Without good reason: The rationality debate in philosophy and cognitive science*. Oxford, England: Oxford University Press.

Tipper, S. P. (1985). The negative priming effect: Inhibitory priming by ignored objects. *Quarterly Journal of Experimental Psychology, 37A*, 571–590.

Todd, P. M., & Gigerenzer, G. (2000). Precis of simple heuristics that make us smart. *Behavioral and Brain Sciences, 23*, 727–780.

UNCORRECTED PROOF

Please cite this article in press as: De Neys, W., & Franssens, S. Belief inhibition during thinking: Not always winning but at least taking part. *Cognition* (2009), doi:10.1016/j.cognition.2009.07.009