UNIVERSITÉ PARIS DESCARTES
INSTITUT DE PSYCHOLOGIE

Détection de Conflits Durant le Raisonnement: Résultats Empiriques et Perspective Développementale

Mémoire de synthèse en vue d'une Habilitation à Diriger des Recherches

Wim De Neys

2013

Conflict Detection During Thinking: Empirical Findings and Developmental Perspective

Summary

Human judgment is often biased by erroneous intuitions. Consider, for example, the fears of the H1N1 virus that recently swept the world. The media commonly referred to the new virus as swine or Mexican flu although it no longer harbored in swine and had already spread over the world at the time of the outbreak. Hence, eating pork or having dinner at your local Mexican restaurant did not pose any clear health risks. The World Health Organization tried hard to inform the public but the mere intuitive association with the name of the virus nevertheless seemed to have an irresistible pull on people's behavior: A lot of us stopped eating at Mexican restaurants, Haitian officials rejected an aid ship with Mexican food aid, pork-belly futures collapsed at Wall Street, and the Egyptian government even ordered their farmers to kill all of their pigs (Alexander, 2009). From a logical point of view, none of these measures was effective to stop the spread of the virus or avoid contamination but intuitively people nevertheless felt they were better off by simply avoiding contact with Mexicans or pork meat.

Decades of reasoning and decision-making research has shown that similar intuitive thinking is biasing people's judgment in a wide range of situations and tasks (Evans, 2003; Kahneman, Slovic, & Tversky, 1982). In general, human reasoners seem to have a strong tendency to base their judgment on fast intuitive impressions rather than on more demanding reasoning. Although this intuitive or so-called "heuristic" thinking might often be useful, it will sometimes cue responses that conflict with normative logical considerations and bias our decision-making. Hence, for sound reasoning it is paramount that reasoners monitor their heuristic intuitions for conflict with logical principles and inhibit the tempting intuitions in case such a conflict is detected (Houdé, 2000, 2007; Evans, 2007; Stanovich & West, 2008; De Neys & Glumicic, 2008; Barrouillet, 2011).

The conflict detection process is a key component of any theory of reasoning and decision-making. Unfortunately, the process is poorly understood and there are

some quite different views on its efficiency. For example, a number of authors have argued that conflict detection during thinking is quite unsuccessful (e.g., Evans, 1984; Kahneman & Frederick, 2002). According to these authors the widespread heuristic bias can be attributed to a failure to monitor our intuition. Because of lax monitoring people would simply fail to detect that the intuitive response conflicts with more normative considerations. Bluntly put, people would be biased because they do not notice that their intuition is wrong.

However, others have suggested that conflict detection during thinking is actually pretty flawless (e.g., Epstein, 1994; Sloman, 1996). According to these authors, there is nothing wrong with the detection process. People do notice that the intuitive response conflicts with more normative considerations. The problem, however, is that despite this knowledge they will not always manage to inhibit and discard the tempting intuitive beliefs. Thus, people "behave against their better judgment" (Denes-Raj & Epstein, 1994, p. 1) when they give an unwarranted heuristic response: They detect that they are biased but simply fail to block the biased response. In sum, in this view biased decisions are attributed to an inhibition failure rather than a conflict detection failure per se (see also Houdé, 2000, 2007).

Clarifying the efficiency of the conflict detection process and the resulting nature of the heuristic bias is paramount for the development of reasoning and decision-making theories. The issue also has far-stretching implications for our view of human rationality (De Neys, 2012). The problem, however, is that it is hard to decide between the alternative views based on traditional reasoning data (Evans, 2007, 2008). My research over the last couple of years has dealt with this shortcoming. In a nutshell, I have tried to use subtle processing measures to test the efficiency of conflict detection during reasoning. In the first part of this thesis I review this work and point to the implications for the debate on human rationality. In the second part of the thesis I look to the future and present a detailed sketch of the developmental research program that I would like to pursue in the coming years.

Overview of conflict detection findings

My conflict detection studies have typically contrasted people's processing of classic "incongruent" reasoning problems in which a cued heuristic response conflicts with the logically correct response, with processing of no-conflict or "congruent" control versions in which this conflict is removed (see Table 1, p. XX, for examples). Accuracy rates on the control versions are typically very high whereas they are notoriously low on the conflict versions. However, the key contribution of the conflict detection studies is that they started to look under the accuracy surface and focused on more subtle measures that made it possible to test whether people processed the two types of problems any differently. For example, one basic procedure has been to simply look at people's response latencies: I have observed that people need more time to solve the conflict than the control versions (e.g., De Neys & Glumicic, 2008; see also Bonner & Newell, 2010, and Villejoubert, 2009, for replications). Now, clearly, the only difference between the two versions is whether the cued heuristic response is consistent with the traditional normative principles or not. If people were mere heuristic thinkers that did not monitor for conflict with these normative considerations and disregarded them, they should not process the two types of problems any differently. Hence, the latency findings supported the idea that people detect the conflict between their heuristic answer and the normative response.

Further support for the detection claim has come from gaze-tracking data that showed that the longer latencies are specifically accompanied by a longer inspection of normatively critical problem information (De Neys & Glumicic, 2008). This "reviewing" has also been shown to result in a better recall of the normatively critical problem information (e.g., Franssens & De Neys, 2009).

The behavioral conflict findings have been validated with a brain-based approach. For example, in one study (De Neys, Vartanian, & Goel, 2008) I have used fMRI to monitor the activation of a specific brain area, the anterior cingulate cortex (ACC), which is believed to mediate conflict detection during thinking (e.g., Botvinick, Cohen, & Carter, 2004; see also discussion between Yeung et al., 2011 and Grindband et al., 2011). Participants were given classic conflict and the no-conflict control versions. In line with the behavioral findings, results showed that the ACC was much more activated when people solved the conflict versions than when they solved the control versions.

In a subsequent study, participants' skin-conductance was recorded to monitor autonomic nervous system activation while they were solving conflict and no-conflict syllogisms (De Neys, Moyens, & Vansteenwegen, 2010). Results showed that solving the conflict problems resulted in a clear electrodermal activation spike. Hence, in addition to the ACC activation, solving conflict problems literally aroused participants. These neural conflict signals have also been shown to decrease people's subjective response confidence in their heuristic answer: Participants typically indicate that they feel less confident about their answer after solving conflict problems than after solving the control problems (e.g., De Neys, Cromheeke, & Osman, 2011).

Taken together, the conflict detection studies suggest that reasoners are detecting the biased nature of their judgment: Although people are typically tempted to give the heuristic response, they at least seem to notice that it conflicts with logical norms and indicate that it is not fully warranted (e.g., De Neys, 2012). These conflict detection findings have been taken as support for the idea that heuristic bias typically results from an inhibition failure (De Neys & Franssens, 2009). That is, the problem does not seem to be that people do not detect that the heuristic response is questionable and needs to be discarded, but rather that people fail to complete the demanding inhibition process. This idea fits with the literature in the reasoning field that has stressed the critical role of inhibitory processing skills to override erroneous heuristic responses (e.g., Houdé, 1997, 2000, 2007; Moutier, Plagne-Cayeux, Melot, & Houdé, 2006; Stanovich & West, 2000). This suggests that people are more logical than their biased answers suggest (De Neys, 2012).

Developmental Research Perspective

Although decades of reasoning and decision-making research have shown that human judgment is often biased by intuitive heuristics, my previous work on conflict detection during thinking nevertheless indicate that despite their biased response, adults typically do detect that their answer is not fully warranted and conflicts with logical considerations. This conflict sensitivity suggests that people are biased because they fail to override the tempting intuitions and are smarter than their errors suggest.

However, it is crucial here to note that my conflict detection studies have been typically run with adult participants. The development of the conflict detection process during thinking has received little attention. Hence, it cannot be excluded that bias detection failures play a more crucial role earlier on in our reasoning development. This hypothesis receives some support from basic neurological studies that suggest that the Anterior Cingulate Cortex (ACC), the critical brain structure that is supposed to be mediating elementary conflict monitoring, is quite slow to mature and would not reach full functionally until middle adolescence (e.g., Davies, Segalowitz, & Gavin, 2004; Fitzgerald et al., 2010; Santesso & Segalowitz, 2008). This tentatively indicates that there might be a critical transition with respect to the locus of heuristic bias in human development. That is, whereas adults would be primarily biased because they fail to inhibit the heuristic response after successful conflict detection, younger reasoners would be biased because they fail to detect the need to inhibit the heuristic response in the first place.

Obviously, from a theoretical point of view it is important to identify possible changes in the nature or locus of heuristic bias throughout our development (e.g., Barrouillet, 2011). However, at a more applied level, establishing whether or not heuristic bias results from a bias detection failure is perhaps even more critical. Exploring the development of conflict detection during thinking is paramount to develop efficient intervention programs to de-bias children's thinking. Note that influential existing intervention programs with adults and older adolescents have focused on training reasoners' inhibitory processing capacities (e.g., Houdé, 2007; Houdé et al., 2000; Moutier, Angeard, & Houdé, 2002; Moutier & Houdé, 2003). However, if younger children do not yet detect that their heuristic response is erroneous, the inhibition training will have less than optimal results. Indeed, any increase in inhibitory processing capacity per se is rather pointless if one is not able to determine whether or not it is needed to inhibit in the first place. Hence, examining children's detection skills is really indispensable to determine which component(s) future intervention programs need to target in order to efficiently de-bias children's thinking. Despite the clear educational and societal importance of this issue, the lack of developmental conflict detection research does currently not allow us to tackle it. The core objective of my planned future research will be to address this shortcoming in the reasoning field and fully document the development of children's conflict detection efficiency.

To reach my objective, I envisage to contrast detection efficiency in age groups ranging from the preadolescent elementary school age to young adulthood (i.e., focused age range of 8 to 18 years). Selection of this age range is inspired by the neurological studies that suggest that the Anterior Cingulate Cortex would not reach full functionally until middle adolescence (i.e., about 14 years of age, see Davies et al., 2004; Fitzgerald et al., 2010; Santesso & Segalowitz, 2008). Hence, within the selected age range I should manage to identify the critical developmental changes in conflict detection during reasoning.

In the planned studies participants will always be presented with sets of incongruent and congruent reasoning tasks. Conflict detection will be assessed with a wide range of behavioural and neuroscientific procedures. The planned studies can be grouped in three phases or workpackages that will chronologically build upon one another. In the initial phase or Workpackage 1, I will construct a database with age appropriate material that will be used for my actual developmental conflict detection experiments. In Workpackage 2, I will rely on behavioral testing procedures (i.e., latency, recall, and response confidence measures) to contrast the conflict detection efficiency in different age groups. In Workpackage 3, I will fine-tune and validate the findings with neuroscientific test procedures (i.e., fMRI, EEG, and SCR measures) that will directly focus on the role of the Anterior Cingulate Cortex.

Taken together the planned studies will result in a full behavioural and neurological specification of the development of the conflict detection process during thinking. Thereby the project will directly identify possible changes in the locus of biased thinking throughout our development and will serve as a much needed building block to develop efficient training programs to help children avoid biased thinking.

Introduction

My dad runs a beer store. When buying a case of fancy Belgian beer, customers often ask whether they can buy a couple of matching glasses. My dad usually gets these glasses for free from his suppliers so he actually doesn't mind giving them away. However, he does not like to be easy on his customers and enjoys putting their decision-making skills to the test. When people ask him how much they owe him for the glasses, he tells them he is charging 5 euro for a glass but he also informs them that if they take a full box of six glasses instead of the one or two they asked for, they will get a 100% reduction. From a rational, economical point of view it is pretty obvious what people need to do. Two glasses will cost them 10 euro (2 * 5 euro = 10 euro). Six glasses would normally cost them 30 euro (6 * 5 euro = 30 euro) but thanks to the 100% reduction they will not be paying anything if they take the full box (100% of 30 euro is 30 euro, of course). This is a very basic calculation that most elementary school children would have little trouble solving. Nevertheless, what my dad typically observes is that although he is catering to well-educated middle-class families, the vast majority of his customers decide to reject his offer. Even when he warns them that they are missing out on the 100% reduction they still decide to stick to (and pay for!) the original number of glasses they asked for. Hence, people prefer to pay for glasses they could easily get for free. As my dad puts it, his customers' striking "failure to think" forces one to conclude that humans are ignorant, irrational beings.

Interestingly, the scientific study of human thinking might seem to confirm my dad's observations. Since psychological studies of reasoning and decision-making started booming in the late 1950s, numerous studies have shown that in a wide range of reasoning and decision-making tasks most educated adults are biased and fail to give the answer that is correct according to logic or probability theory (Evans & Over, 1996; Kahneman, Slovic, & Tversky, 1982). The general problem seems to be that reasoners overrely on intuitions and gut feelings instead of on more demanding, deliberative reasoning when making decisions (Evans, 2003; Kahneman, 2002). Although this intuitive or so-called "heuristic" thinking might sometimes be useful, it will often cue responses that are not warranted from a normative point of view. Consequently, people's reasoning and decision-making is often biased.

It is not hard to see how such intuitive or heuristic thinking is biasing my dad's customers in his store. Intuitively, people's gut feeling might simply be telling them that by offering an additional reduction my dad is trying to persuade them to buy more than they asked for. In general, such a heuristic might be a useful tool to prevent falling prey to sales tricks. However, in my dad's store this mere intuitive reasoning is costing people good money. Hence, the point is not that heuristics or intuitions are necessarily bad. The point is rather that during reasoning and decision-making it is crucial to check whether one's intuitions conflict with more normative considerations. As my dad would claim, the omnipresence of heuristic bias suggests that people are not very good at detecting such conflicts.

The conflict detection process is a key component of any theory of reasoning and decision-making. Unfortunately, the process is poorly understood and there are some quite different views on its efficiency. Consistent with my dad's view, for example, a number of authors have argued that conflict detection during thinking is quite unsuccessful (e.g., Evans, 1984; Kahneman & Frederick, 2002). According to these author,s the widespread heuristic bias can be attributed to a failure to monitor our intuition. Because of lax monitoring people would simply fail to detect that the intuitive response conflicts with more normative considerations. Bluntly put, people would be biased because they do not notice that their intuition is wrong.

However, others have suggested that conflict detection during thinking is actually pretty flawless (e.g., Epstein, 1994; Sloman, 1996). According to these authors, there is nothing wrong with the detection process. People do notice that the intuitive response conflicts with more normative considerations. The problem, however, is that despite this knowledge they will not always manage to inhibit and discard the tempting intuitive beliefs. Thus, people "behave against their better judgment" (Denes-Raj & Epstein, 1994, p. 1) when they give an unwarranted heuristic response: They detect that they are biased but simply fail to block the biased response. In sum, in this view biased decisions are attributed to an inhibition failure rather than a conflict detection failure per se (see also Houdé, 2000, 2007).

Clarifying the efficiency of the conflict detection process and the resulting nature of the heuristic bias is paramount for the development of reasoning and decision-making theories. The issue also has far-stretching implications for our view of human rationality. If the popular bias-as-detection-failure view is right and reasoners do not detect that their heuristic response is wrong, this implies that reasoning errors are indeed quite "dumb". The second view, however, implies that people's errors are less ignorant. If people detect that their intuitive response is not fully warranted, this implies that people did not simply neglect the normative considerations. Contrary to my dad's conclusion, this would suggest that people are no mere heuristic thinkers and might be more rational than their actual responses show.

The problem, however, is that it is hard to decide between the alternative views based on traditional reasoning data (Evans, 2007, 2008). My research over the last couple of years has dealt with this shortcoming. In a nutshell, I have tried to use subtle processing measures to test the efficiency of conflict detection during reasoning. In the first part of this thesis I will review this work and point to the implications for the debate on human rationality. In the second part of the thesis I look to the future and present a detailed sketch of the research program that I would like to develop in the coming years.

For clarity, it is probably not a bad idea to note that the reader should bear some general points in mind with respect to the nomenclature that I will be using in this thesis. First, the labels "correct", "normative" or "logical" response are used as a handy shortcut to refer to "the response that has traditionally been considered as correct or normative according to standard logic or probability theory". The appropriateness of these traditional norms has sometimes been questioned in the reasoning field (e.g., see Stanovich & West, 2000). Under this interpretation, the heuristic response should not be labelled as “incorrect” or “biased”. I will point to implications of the conflict detection work for this debate later on in this thesis but for the sake of simplicity and consistency I stick to the traditional labelling here. In the same vein, I use the term “logical” as a general header to refer both to standard logic and probability theory.

Part A: Overview of empirical conflict detection findings

To detect or not to detect? *(De Neys & Glumicic, 2008, Cognition)*

In a first study that I ran to start testing the efficiency of the conflict detection process during thinking (see De Neys and Glumicic, 2008), Tamara Glumicic (one of my students) and I clarified that the classic claims about the detection process were typically anecdotal in nature. Epstein (1994; Denes-Raj & Epstein, 1994; Epstein & Pacini, 1999), for example, repeatedly noted that when picking an erroneous answer his participants spontaneously commented that they did "*know*" that the response was wrong but stated they picked it because it "*felt*" right. Such comments do seem to suggest that people detect that their intuition conflicts with normative considerations. The problem, however, is that spontaneous self-reports and anecdotes are no hard empirical data. This is perhaps best illustrated by the fact that Kahneman (2002, p. 483) also refers to "casual observation" of his participants to suggest that only in "some fraction of cases, a need to correct the intuitive judgements and preferences will be acknowledged". Therefore, in a first experiment De Neys and Glumicic decided to adopt a thinking aloud procedure (e.g., Ericsson & Simon, 1993). The thinking aloud procedure has been designed to gain reliable information about the course of cognitive processes. Participants are simply instructed to continually speak aloud the thoughts that are in their head as they are solving a task. Thinking aloud protocols have been shown to have a superior validity compared to interpretations that are based on retrospective questioning or people's spontaneous remarks (Payne, 1994).

De Neys and Glumicic (2008) asked their participants to solve problems that were modelled after Kahneman and Tversky's classic (1973) base-rate neglect problems. These base-rate neglect problems are among the most (in)famous tasks in the field. In the problems people first get information about the composition of a sample (e.g., a sample with 995 females and 5 males). People are also told that short personality descriptions are made of all the participants and they will get to see one description that was drawn randomly from the sample. Consider the following example:

> A psychologist wrote thumbnail descriptions of a sample of 1000 participants consisting of 995 females and 5 males. The description below was chosen at random from the 100 available descriptions.
>
> 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.
>
> Which one of the following two statements is most likely?
> a. Jo is a man
> b. Jo is a woman

From a logical point of view, given the size of the two groups in the sample, it is more likely that a randomly drawn individual will be a female[1]. However, intuitively many people will be tempted to respond that the individual is a male based on stereotypical beliefs cued by the description ("Jo is an engineer and drinks beer").

The crucial question for De Neys and Glumicic was whether verbal protocols would indicate that when people selected the intuitive response option ("a. Jo is a man") they at least referred to the group size information during the reasoning process (e.g., " … because Jo's drinking beer and loud I guess Jo'll be a guy, *although there were more women* …"). In this task such basic sample size reference during the reasoning process can be considered as a minimal indication of successful conflict monitoring. It indicates that this information is not simply neglected.

Results were pretty straightforward. People who gave the correct response typically also referred to the base-rate information and reported they were

[1] Consistent with previous work, responses that are in line with the base rates (i.e., selection of the largest group as most likely answer) are labelled as correct answers. As I noted in the introduction, the actual normative status of the 'correct' response in these problems is sometimes debated (Gigerenzer, Hell, & Blank, 1988). The present thesis is concerned with the empirical question as to what extent people take the base rates into account during thinking whether or not the base rates ultimately turn out to be "normative" or not. Therefore, one can adopt a nominalist stance towards the use of the terms 'correct' and 'error'.

experiencing a conflict (e.g., "… it sounds like he's a guy, *but because they were more women*, Jo must be female so I'll pick option b …"). However, people who gave the intuitive response hardly ever (less than 6% of the cases) mentioned the base-rate information (e.g., a typical protocol would read something like " … This person is a guy … drinks, listens to loud music … yeah, must be a guy … so I'll pick a … "). Hence, consistent with my dad claims and the error-as-detection-failure view, the verbal protocols seemed to indicate that people are indeed mere intuitive reasoners who do not detect that they are biased.

De Neys and Glumicic noted, however, that it could not be excluded that conflict detection was successful at a more implicit level. It might be that the conflict detection experience is not easily verbalized. People might notice that there is something wrong with their intuitive response but they might not always manage to put their finger on it. Such more implicit conflict detection would still indicate that people detect that their response is not fully warranted, of course. To capture such implicit detection De Neys and Glumicic also presented participants with a surprise recall test. After a short break following the thinking-aloud phase participants were asked to answer questions about the group sizes in the previous reasoning task. Participants were not told that recall would be tested while they were reasoning but De Neys and Glumicic reasoned that the detection of the conflict should result in some additional scrutinising of the normative base-rate information. This deeper processing of the base-rate information should subsequently benefit recall.

To validate the recall hypothesis participants were also presented with additional control problems. In the classic base-rate problems the description of the person is composed of common stereotypes of the smaller group so that the response cued by the base-rates and the intuitive response that is cued by the description disagree. In addition to these classic problems De Neys and Glumicic also presented problems in which the base-rates and description both cued the same response. In these *congruent* problems the description of the person was composed of stereotypes of the *larger* group. Hence, contrary to the classic (i.e., *incongruent*) problems the intuitive response did not conflict with more normative considerations and the response could be rightly based on mere intuitive processing. For a reasoner who neglects the base-rates and does not detect the conflict on the classic problems both types of problems will be completely similar and base-rate recall should not differ. However, if one does detect the conflict, the deeper processing of the base-rates in case of a conflict should result in a better recall for the classic problems than for the congruent control problems.

Recall results showed that participants had indeed little trouble recalling the base-rates of the classic conflict problems. People easily remembered which one of the two groups in each problem was the largest. On the congruent control problems, however, recall performance was merely at chance level. Interestingly, the superior recall was obvious even for those people who never mentioned the base-rates while thinking-aloud and failed to solve any of the presented classic conflict problems correctly. Since the only difference between the classic and control problems was the conflicting nature of the base-rates and description, De Neys and Glumicic concluded that people had little difficulty in detecting the conflict per se.

In an additional experiment De Neys and Glumicic examined the conflict detection issue further by introducing a "moving window" procedure (e.g., Just, Carpenter, & Wooley, 1982). In the experiment the base-rates and the description were presented separately. First, participants saw the base-rate information on a computer screen. Next, the description and question were presented and the base-rates disappeared. Participants had the option of visualizing the base-rates afterwards by holding a specific button down. Such base-rate reviewing can be used as an additional conflict detection index. De Neys and Glumicic explained their recall findings by assuming that when people detect that the description conflicts with the previously presented base-rates they will spend extra time scrutinizing or "double checking" the base-rates. With the "moving window" procedure the time spent visualizing the base-rates can be used as a measure of this reviewing tendency. If conflict detection is indeed successful, people should show longer response latencies and a stronger tendency to visualize the base-rates when solving classic incongruent vs. congruent control problems. This is exactly what De Neys and Glumicic observed. Once again the stronger base-rate reviewing was present for the least-gifted reasoners in the sample who consistently gave the intuitive response on all presented incongruent problems.

To the brain and beyond *(De Neys, Vartanian, & Goel, 2008, Psych Science)*

In a further attempt to clarify the nature of heuristic bias, I decided to focus on the neural basis of conflict detection and response inhibition during thinking (see De Neys, Vartanian, & Goel, 2008). Together with Oshin Vartanian and Vinod Goel, I noted that numerous imaging studies established that conflict detection and actual response inhibition are mediated by two distinct regions in the brain. Influential work in the cognitive control field (e.g., Botvinick, Cohen, & Carter, 2004; Ridderinkhof, Ullsperger, Crone, & Nieuwenhuis, 2004), for example, showed that detection of an elementary conflict between competing responses is among the functions of the medial part of the frontal lobes, more specifically the Anterior Cingulate Cortex (ACC). While the ACC signals the detection, correct responding and actually overriding the erroneous, prepotent response has been shown to depend on the recruitment of the more lateral part of the frontal lobes (more specifically the right lateral prefrontal cortex or RLPFC).

De Neys, Vartanian, and Goel (2008) therefore suggested that turning to the brain might help to address the dispute about the nature of heuristic bias. Solving classic reasoning and decision-making problems that cue a salient but inappropriate intuitive response requires that reasoners detect that the intuitive response conflicts with normative considerations, first. In addition, the intuitive responses will need to be successfully inhibited. If the ACC and RLPFC mediate this conflict detection and inhibition process, respectively, correct reasoning should be associated with increased activation in both areas. De Neys, Vartanian, and Goel reasoned that the crucial nature of the intuitive bias could be clarified by contrasting ACC and RLPFC activation for intuitive and normative responses. The bias-as-inhibition-failure and bias-as-detection-failure views make differential predictions with respect to the activation of the conflict detection region. If De Neys and Glumicic's initial behavioural findings were right and people at least detect that the intuitive response conflicts with more normative considerations, the ACC should be activated whether or not people are biased. However, if biased decisions arise because people fail to detect that the intuitive response is inappropriate, people will not detect a conflict when they give an intuitive response and consequently the ACC should not be activated.

De Neys, Vartanian, and Goel tested these predictions in an fMRI study in which participants were asked to solve base-rate problems while the activation of the ACC and RLPFC was monitored. As expected, results showed that for trials in which people selected the correct base-rate response on the classic, incongruent problem versions both the conflict detection (ACC) and inhibition region (RLPFC) showed increased activation. When people were biased and selected the intuitive response on these problems, the RLPFC inhibition region was not recruited. The conflict detection ACC region, however, did show clear activation when the intuitive response was selected. On congruent control trials in which the cued intuitive and normative response did not conflict, the ACC was not significantly activated.

In sum, De Neys, Vartanian, and Goel's crucial finding was that biased and unbiased responses on the classic base-rate problems only differed in RLPFC recruitment. Solving incongruent problems did engage the ACC region but the activation did not differ for intuitive or base-rate responses. Consistent with De Neys and Glumicic's behavioural findings this suggested that the intuitive bias should not be attributed to a detection failure but rather to an inhibition failure.

The effortless nature of conflict detection *(Franssens & De Neys, 2009, TAR)*

Taken together the De Neys and Glumicic (2008) and De Neys, Vartanian, and Goel studies (2008) supported the view of authors such as Epstein (1994) who claimed that conflict detection during thinking is pretty flawless. However, the absence of any verbally expressed conflict experience suggested that the popular characterization of this process as an explicitly experienced struggle in which people are actively deliberating between two different options ("I know it's wrong but it feels right") is not very accurate. Hence, I started to be convinced that the conflict detection process itself might be better conceived as an intuitive process that simply warns people that more deliberate reasoning is required. Although the conflict detection would suffice to inform people that their heuristic conclusion is not fully warranted and needs to be scrutinized, it would not guarantee that further deliberate reasoning is actually engaged to override and inhibit the heuristic response. Bluntly put, it looks like people intuitively feel that "something" is wrong but, without more demanding deliberate thinking, cannot exactly specify what.

Together with my student Sam Franssens, I therefore decided to set up a study to test the claim that conflict detection is an intuitive process directly (see Franssen & De Neys, 2009). One of the key characteristics of intuitive, implicit processing is that it is effortless and does not draw on people's limited executive working memory resources that are required for controlled processing (e.g., Moors & De Houwer, 2006). Franssens and De Neys (2009) therefore decided to burden these executive resources during reasoning. In their study participants were asked to memorize spatial dot patterns while they were trying to solve base-rate problems. This dot memorization task had been previously shown to burden the executive resources (Miyake, Friedman, rettinger, Shah, & Hegarty, 2001). Franssens and De Neys reasoned that if conflict detection during thinking was indeed intuitive, it should not be affected by the executive memorization load. The efficiency of the conflict detection process was measured by presenting the participants with the surprise base-rate recall task that was introduced in the De Neys and Glumicic (2008) studies. Results showed that reasoning performance per se decreased under memorization load. Participants gave more heuristic responses when their executive resources were burdened. However, the recall performance was not affected. Even under load base-rate recall was still better for classic incongruent than for congruent control problems and the percentage correct recall for the incongruent problems did not differ under load and no-load conditions. Hence, the study nicely supported the characterization of conflict detection as a flawless and intuitive process.

Gut conflict feelings *(De Neys, Moyens, & Vansteenwegen, 2010, CABN)*

Further support for the intuitive nature of the conflict detection process comes from a study that I ran together with Elke Moyens and Deb Vansteenwegen in which we decided to measure people's autonomic nervous system activation during thinking (see De Neys, Moyens, & Vansteenwegen, 2010). The inspiration for this study came from basic cognitive control studies (e.g., Botvinick, Cohen, & Carter, 2004; Ridderinkhof, Ullsperger, Crone, & Nieuwenhuis, 2004). In these basic studies people are typically presented with very elementary conflict tasks in which they need to withhold an inappropriate but dominant response. As I mentioned, previous work in this field showed that the anterior cingulate cortex (ACC) is especially sensitive to the presence of conflict between competing responses (e.g., Van Veen & Carter, 2006). The fMRI study of De Neys et al. (2008) that I presented above established that this same cortical conflict region was activated when people gave biased responses during high-level reasoning. Interestingly, it has been shown in the cognitive control field that besides ACC activation, the elementary conflicts also elicit global autonomic arousal (Kobayashi, Yoshino, Takahashi, & Nomura, 2007). In other words, at least in the elementary control tasks, the presence of conflict seems to be accompanied by visceral arousal as reflected, for example, in increased skin conductance (Hajcak, McDonald, & Simons, 2003). This suggests that basic measures of electrodermal activation can be used as a biological index of conflict detection in the reasoning field. Based on the cognitive control findings one can expect that if conflict detection during thinking is indeed flawless, solving reasoning tasks in which intuitions conflict with logic will elicit increased skin conductance responses. Hence, measuring participants skin conductance during reasoning allowed me to validate my previous fMRI findings. In addition, establishing a possible link between autonomic modulation and conflict detection would help to provide more solid ground for conceptualization of conflict detection as an intuitive process. That is, people might indeed literally "feel" the presence of conflict.

Hence, in the study we presented participants with classic incongruent and congruent reasoning problems that were used in the previous studies and attached electrodes to the palm of participants' hands to measure skin conductance (SCR) fluctuations. Results were very straightforward. As expected, we observed a clear SCR boost when participants were solving the incongruent problems. Consistent with my fMRI and behavioural findings, this SCR boost was present even when participants failed to solve the incongruent problem correctly.

Biased but in doubt *(De Neys, Cromheeke, & Osman, 2011, Plos One)*

The conflict detection work that I presented so far indicated that although it is clear that people do not explicitly detect that they are erring, they do seem to be sensitive to the presence of conflict between cued intuitive and normative logical or probabilistic principles at a more implicit level. The lack of explicitation has been explained by arguing that the neural conflict detection signal should be conceived as an implicit "gut" feeling. The signal would inform people that their intuition is not fully warranted but people would not always manage to verbalize the experience and explicitly label the logical principles that are being violated. Although this hypothesis is not unreasonable, it faces a classic caveat. Without discarding the possible value of implicit processing, the lack of explicit evidence does open the possibility that the implicit conflict signal is a mere epiphenomenon. That is, the studies reviewed above clearly established that some part of our brain is sensitive to the presence of conflict in classic reasoning tasks. However, this does not necessarily imply that this conflict signal is also being used in the reasoning process. In other words, showing that the presence of conflict is detected does not suffice to argue that reasoners also "know" that their intuition is not warranted. Indeed, a critic might utter that the fact that despite the clear presence of a conflict signal people do not report experiencing a conflict and keep selecting the erroneous response, questions the value of this signal. Hence, what is needed to settle the bias debate is some minimal (nonverbal) indication that this signal is no mere epiphenomenon but has a functional impact on the reasoning process. I have tried to pass this last hurdle in a set of experiments that I ran with Sofie Cromheeke and Magda Osman (see De Neys, Cromheeke, & Osman, 2011).

We reasoned that a straightforward way to assess the functional relevance of the implicit conflict signal is to examine people's decision confidence after they solve a reasoning problem. If the detection signal is not merely epiphenomenal, but actually informs people that their intuitive response is not fully warranted, people's decision confidence should be affected. That is, if people detect that they are biased but simply fail to verbalize the experience, we should at the very least expect to see that they do not show full confidence in their judgments.

Of course, people might never show full confidence and there might be myriad reason for why individuals differ in their confidence ratings (e.g., Kruger & Dunning, 1999; Shynkaruk & Thompson, 2006). Note, however, that our main research question does not concern people's absolute confidence level. As in the initial detection studies, we give participants classic conflict problems and congruent control problems. To recap, the only difference between the two types of problems is that cued intuitions conflict with traditional normative principles in the conflict versions while intuition and normative principles cue the same response in the congruent or no-conflict versions. The aim of the confidence contrast for the two types of problems is to help decide the detection debate. If detection of the intrinsic conflict on the classic versions is functional for the reasoning process and informs people that their intuitive response is questionable, participants should show lower confidence ratings after solving conflict problems as compared to no-conflict problems. If people do not detect the presence of conflict or the signal has no impact on the reasoning process, confidence ratings for the two types of problems should not differ.

To test our predictions participants were given a set of incongruent and congruent reasoning problems. Note that we used a range of different classic problems (e.g., not only the base-rate problems that I used in previous work but also incongruent and congruent syllogistic reasoning problems and conjunction fallacy problems, for example[2]) to test the generality of the findings. After participants solved a problem we showed them a confidence rating scale that ranged from 100% ("Very confident that my answer is correct") to 0% ("Very unconfident that my answer is correct"). Participants were asked to indicate how confident they were that the response they just gave was correct.

Results confirmed our predictions. For all the different problem types that we used, participants who failed to solve the incongruent versions correctly and selected the heuristic response were significantly less confident in their answer after solving the incongruent than after solving the congruent problems (i.e., on average we observed about a 10% drop in confidence). This directly establishes that reasoners detect that their heuristic response is literally questionable. Hence, the previously established neural and behavioural conflict signals are not merely epiphenomenal. Although people might not manage to explain why their answer conflicts with logical principles, they do know that their answer is not fully appropriate.

Not always winning, but taking part *(De Neys & Franssens, 2009, Cogntion)*

The evidence that I presented with respect to the successful nature of the conflict detection process, lends credence to the view that heuristic bias results from a failure to override the inappropriate but salient heuristic response. An interesting follow-up question is whether this inhibition failure needs to be conceived as a

[2] The interested reader can find examples of these different problems in Table 1 (p. 43)

failure to engage in inhibitory processing or as a failure to complete the process. That is, do people after they detect the initial conflict at least try to inhibit the heuristic response too? To answer this question De Neys and Franssens (2009) presented participants with a lexical decision task after they solved reasoning problems. In the lexical decision task participants have to say whether a string of presented letters (e.g., "DETXXC" or "BALL") forms an existing word or not. Classic memory studies have shown that when people try to inhibit certain information, memory access to this information is temporary impaired afterwards (e.g., MacLeod et al., 2003; Neill, 1997; Tipper, 1985). Lexical decision tasks are used to test this memory accessibility. For example, if you inhibit the word "BALL" and are asked whether "BALL" is a word or not, you will need a couple ms longer to make your decision.

De Neys and Frannsens used this procedure in a reasoning setting. Participants solved a range of incongruent and congruent reasoning problems. After each problem they were presented with a lexical decision task. The critical manipulation was that half of he presented words (i.e., so-called target words) were strongly associated to the heuristic response that was cued in the reasoning tasks. For example, in the base-rate problem with "Jo" who was drawn from a sample with males and female possible target words associated with the heuristic response ("male") would be "TIE", "FOOTBALL" or "TRUCK" etc. De Neys and Franssens reasoned that if people indeed tried to inhibit the heuristic response when it conflicted with the normative response, then lexical decision times for the target words should be longer after solving incongruent vs. congruent problems. This was exactly what they observed. Even participants who failed to answer the incongruent problem correctly showed a slightly impaired memory access, suggesting that although they did not succeed in inhibiting the heuristic response, they at least engaged in inhibitory processing and tried to do so. Obviously, this further establishes that people detect that the heuristic response is not warranted.

Implications for the Rationality Debate *(De Neys, 2012, PIPS)*

The studies reviewed above suggest that people are quite good at detecting the conflict between cued heuristic intuitions and more normative considerations when solving classic reasoning problems. Although people's responses are typically biased they do seem to have an intuitive gut feeling that is telling them that their heuristic answer is not fully warranted. Even though it is hard for people to verbalize this intuitive conflict feeling, its flawless manifestation indicates that normative considerations are not simply neglected. If people were not to know the normative principles (e.g., the fact that base-rates matter) or would not consider these normative principles to be relevant, there would simply be no conflict to be detected in the first place and congruent and incongruent problem versions should be processed in the exact same manner. Clearly, conflict can only occur when both the intuitive response and normative considerations are taken into account during thinking. The fact that people are particularly sensitive to the presence of this conflict when solving classic decision-making problems implies that people are no mere heuristic thinkers who simply neglect normative considerations. In this section I will try to clarify that this point has some profound implications for the debate about the rationality of the human species (e.g., Stanovich & West, 2000; Stein, 1996).

The so-called "rationality debate" has raged through the reasoning and decision-making field for more than four decades without clear solution. In essence, the debate centers around two related questions: a) whether human reasoning is rational and b) whether the traditional normative systems (such as logic and probability theory) against which the rationality of our inferences and decisions are measured are actually valid. The initial findings in the 1960's that pointed to the omnipresence of heuristic bias led some theorists to question the rationality of the human species (e.g., Wason, 1968, 1983; see Evans, 2002, for a nice review). Just like my dad in his store, these theorists concluded that people's widespread failure to reason in line with the logical or probabilistic norm indicated that humans are irrational beings. However, later on this pessimistic conclusion was rejected by theorists who started questioning the validity of the classic norms. 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 (e.g., Hertwig & Gigerenzer, 1999; Oaksford & Chater, 1998; Todd & Gigerenzer, 2000). For example, in the base-rate problems participants might interpret the task as a simple social classification task and would therefore not keep track of the base-rate information. These authors clarified that the rationality of our behaviour depends on the goals we try to fulfil. If our goal is making a social classification judgment, neglecting the base-rates is the rational thing to do and cannot be considered a bias. Hence, according to this "alternative norm" view, people's behaviour in the classic reasoning and decision-making experiments is perfectly rational but has simply been measured against the wrong standards.

One might note that the opposite rationality views are trading-off rationality and norm validity. People like my dad take the validity of the classic norms for granted and conclude that the failure to reason in line with these norms points to human irrationality. The "alternative norm" view on the other hand saves human rationality but at the cost of the validity of the classic norms. I believe that studying the conflict detection process during thinking presents an opportunity to resolve this debate and unify the two views. The conflict detection data that I reviewed suggest that both human rationality and the validity of the classic norms can be saved. If people were really to interpret classic reasoning and decision-making tasks as social classification tasks and were to believe that normative considerations such as sample sizes do not matter, their task processing should not be affected by the presence of a conflict between cued social intuitions and the very same normative principles. Hence, contrary to the "alternative norm" view this indicates that people do not consider the classic norms to be irrelevant. On the other hand, the fact that people pick up this conflict shows that they take normative considerations into account and are no mere intuitive thinkers. In sum, people might not always manage to reason in line with the classic norms but this does not imply that they do not know the norms or consider them to be irrelevant. The initial conflict detection studies suggest that all reasoners are at least trying to adhere to the classic norms and detect that their intuition is not warranted.

Conclusion

In this first part of the thesis I highlighted my research framework that started focusing on the efficiency of the conflict detection process during thinking. I hope to have clarified the potential and importance of this line of research. The key point is that a failure to characterize the conflict detection process during thinking is bound to bias any conclusions about human rationality or the validity of the classic norms. My conflict processing data indicates that people are pretty good at detecting their bias. Contrary to popular views in the decision-making field and the opinion of at least one Belgian beer expert, this suggests that people are far more rational and normative than their biased answers suggest.

Part B: Developmental research perspective

Background and objectives of research program

Taken together, my previous work on conflict detection that I reviewed in Part A suggests that reasoners are detecting the biased nature of their judgment: Although people are typically tempted to give the heuristic response, they at least seem to notice that it conflicts with logical norms and indicate that it is not fully warranted (e.g., De Neys, 2012). These conflict detection findings have been taken as support for the idea that heuristic bias typically results from an inhibition failure (e.g., Bonner & Newell, 2010; De Neys & Glumicic, 2008). That is, the problem does not seem to be that people do not detect that the heuristic response is questionable and needs to be discarded, but rather that people fail to complete the demanding inhibition process (De Neys & Franssens, 2009). This idea fits with the vast literature in the reasoning field that has stressed the critical role of inhibitory processing skills to override erroneous heuristic responses (e.g., Brainerd & Reyna, 2001; De Neys & Van Gelder, 2008; Handley et al., 2004; Houdé, 1997, 2000, 2007; Moutier, Plagne-Cayeux, Melot, & Houdé, 2006; Simoneau & Markovits, 2003; Stanovich & West, 2000). However, it is crucial here to note that the bias detection studies have been typically run with adult participants. Unfortunately, the development of this bias detection or awareness process has received little attention.

In general, developmental psychologists have already argued that reasoning is a multi-component process and that biased responses might have multiple causes (e.g., Brainerd & Reyna, 2001; Jacobs & Klaczynski, 2002; Markovits & Barrouillet, 2004; Stanovich, West, & Toplak, 2011). Hence, it cannot be excluded that bias detection failures play a more crucial role earlier on in our reasoning development. This hypothesis receives some support from basic neurological studies that suggest that the Anterior Cingulate Cortex (ACC), the critical brain structure that is supposed to be mediating elementary conflict monitoring, is quite slow to mature and would not reach full functionally until middle adolescence (e.g., Davies, Segalowitz, & Gavin, 2004; Fitzgerald et al., 2010; Santesso & Segalowitz, 2008). This tentatively indicates that there might be a critical transition with respect to the locus of heuristic bias in human development. That is, whereas adults would be primarily biased because they fail to inhibit the heuristic response after successful conflict detection, younger reasoners would be biased because they fail to detect the need to inhibit the heuristic response in the first place.

Obviously, from a theoretical point of view it is important to identify possible changes in the nature or locus of heuristic bias throughout our development (e.g., Barrouillet, 2011). However, at a more applied level, establishing whether or not heuristic bias results from a bias detection failure is perhaps even more critical. Exploring the development of conflict detection during thinking is paramount to develop efficient intervention programs to de-bias children's thinking. Note that influential existing intervention programs with adults and older adolescents have focused on training reasoners' inhibitory processing capacities (e.g., Houdé, 2007; Houdé et al., 2000; Moutier, Angeard, & Houdé, 2002; Moutier & Houdé, 2003). However, if younger children do not yet detect that their heuristic response is erroneous, the inhibition training will have less than optimal results. Indeed, any increase in inhibitory processing capacity per se is rather pointless if one is not able to determine whether or not it is needed to inhibit in the first place. Moreover, training inhibition in the absence of good conflict detection might even have unwanted negative side effects. Note that heuristic thinking is not always wrong. Often, the heuristic response will reside with more deliberate and elaborate logical analysis. Since heuristic thinking is typically fast and effortless (e.g., Stanovich & West, 2000; but see also Morsanyi & Handley, 2008), it can be highly beneficial and even life-saving in these cases (e.g., Gigerenzer & Todd, 1999). Making optimal use of available inhibitory capacities requires that one monitors for conflict first and shuts down the heuristic route only when it is needed. Training inhibition in the absence of efficient bias detection might therefore result in a general shutdown of the heuristic route. That is, children might simply start to mistrust their heuristic intuitions throughout. In many cases this could deprive children from the advantages of heuristic thinking. Hence, examining children's detection skills is really indispensable to determine which component(s) future intervention programs need to target in order to efficiently de-bias children's thinking.

In sum, my existing research on conflict detection during thinking has focused on adults' performance. The primary objective of the research that I plan in the coming years is to fully document critical developmental changes in this conflict detection efficiency over the elementary school and secondary school years (i.e., preadolescence to young adulthood). This planned research will be paramount to identify possible changes in the locus of heuristic bias throughout our development and will be indispensible to develop efficient intervention programs.

In the following sections I sketch the details of the program and planned experiments.

Program structure

To address my objectives and document the development of children's conflict detection during thinking I will present participants in the preadolescent to young adulthood age range with a set of classic reasoning tasks and measure their detection efficiency. As in my studies with adults, the participants will always be presented with standard, "incongruent" or "conflict" versions of the tasks and control "incongruent" or "no-conflict" versions. The classic, standard versions cue a strong heuristic response that conflicts with the logically correct response. In the no-conflict versions this conflict is removed and the heuristic response is consistent with the logical response. Examples of these tasks can be found in Table 1. The selected tasks (i.e., conjunction fallacy task, base-rate neglect task, belief bias syllogisms tasks) have been used in hundreds of studies and inspired much of the theorizing in the field. Adopting these so-called "fruit flies" tasks of the reasoning and decision-making field (Bonnefon, 2011) will allow me to test the generality of my findings.

Participants' conflict detection will be tested with a range of behavioral and neuroscientific procedures that allow testing whether reasoners process the standard and control versions differently and show sensitivity to the intrinsic conflict in the standard versions. I envisage to contrast detection efficiency in age groups ranging from the preadolescent elementary school age to young adulthood (i.e., focused age range of 8 to 18 years). Selection of this age range is inspired by basic neurological studies that suggest that the Anterior Cingulate Cortex (ACC), the critical brain structure that is supposed to be mediating elementary conflict monitoring, is quite slow to mature and would not reach full functionally until middle adolescence (i.e., about 14 years of age, see Davies et al., 2004; Fitzgerald et al., 2010; Santesso & Segalowitz, 2008). Hence, within the selected age range I should manage to identify the critical developmental changes in conflict detection during reasoning.

The studies that I am envisaging can be grouped in three integrated workpackages (WP) that will chronologically build upon one another. In the initial Workpackage 1 I will construct a database with age appropriate material that will be used for the actual conflict detection experiments. In Workpackage 2, I will rely on behavioral testing procedures to explore the conflict detection efficiency in the targeted preadolescent to young adulthood age range. In Workpackage 3, I will fine-tune and validate the findings with neuroscientific test procedures.

Database construction *(Workpackage 1)*

The classic reasoning problems that will be adopted in my planned studies typically cue a strong heuristic response based on a stereotypical personality description (e.g., base-rate and conjunction tasks) or the fit with general world-knowledge (e.g., believable or unbelievable syllogistic conclusions). A critical prerequisite to study conflict detection in a developmental context is of course that all age groups are familiar with the problem material. For example, if elementary schoolers have not yet acquired the typical "accountant" stereotype, the base-rate problem in Table 1 will simply not cue a heuristic response and conflict detection or heuristic inhibition will be redundant for sound reasoning. In general, if younger age groups are not familiar with the intended stereotypical association, the problem will not cue an intuitive heuristic response, and conflict detection will by definition not be possible. To avoid such a confound, I will start my developmental project with an extensive material validation study.

In an initial material generation phase, I will draw on a limited number of available studies (e.g., De Neys & Vanderputte, 2011), additional literature search, and interviews with parents and teachers to construct a wide range of stereotypical descriptions and world-knowledge statements that can be expected to be highly familiar for young elementary schoolers. In a second phase, participants from the different age groups of interest will be presented with the generated material and will be asked to rate its familiarity (e.g., rate how believable the syllogistic conclusions are or rate how likely it is that the described person belongs to category X on a 5-point scale). This will allow me to construct appropriate material for the planned developmental conflict detection studies.

For completeness, note that at the end of all my planned experiments, I also envisage to present participants with a post-experimental questionnaire with abstract-content problems. These abstract problems contain stereotype or belief neutral material that does not cue a heuristic response (e.g., a personality description that states that a person has "black hair" or a syllogistic conclusion like "All X are Y"). Abstract problems allow us to check whether participants show some basic familiarity with the elementary logical "fruit flies" principles in the absence of interfering beliefs (e.g., see De Neys & Vanderputte, 2011). Although developmental studies have established that this is typically the case in the current age range of interest (e.g., Brainerd & Reyna, 2001; Morris, 2000; Téglás, Girotto, Gonzalez, & Bonatti, 2007), inclusion of the problems allows me to identify the few participants whose lack of conflict detection might result from a knowledge gap (i.e., a "mind" gap, Stanovich & West, 2008) rather than monitoring failure per se. Note that the neutral status of the selected abstract problem content will also be validated in the planned rating study.

Planning. For the rating study I envisage to test large groups of participants from grade 3 (i.e., average age 8+ years) to grade 12 (i.e. average age +17 years) and a group of young adult university students. Selected material will be printed in booklets and ratings will be collected in the classrooms in group sessions. Participants will be recruited through the large network of schools that my new host group (LaPsyDE) at Paris Descartes University is collaborating with.

Behavioral detection studies *(Workpackage 2)*

In second phase of the project I will rely on behavioral testing procedures to start exploring the conflict detection efficiency in the targeted preadolescent to young adulthood age range. I will use the material database build in WP1 to construct appropriate conflict and no-conflict versions of the three classic reasoning tasks I plan to test (i.e., base-rate neglect task, conjunction fallacy task, and belief-bias syllogisms task). I envisage two large scale studies with different behavioral detection measures that have been used extensively in my conflict detection work with adults. The first study will combine a latency based detection index with a recall measure and the second one will back-up the latency findings with a confidence measure. I describe these studies in more detail bellow.

Latency & Recall study. In a first study I will record response latencies to measure conflict detection efficiency in the different age groups. As I noted, conflict detection studies established that although adults typically fail to solve conflict problems correctly and give the heuristic response, they nevertheless take slightly longer to answer the conflict than the no-conflict problems (e.g., Bonner & Newell, 2010; De Neys & Glumicic, 2008; Stupple & Ball, 2008; Villejoubert, 2009). Basic cognitive control research suggests that such increased processing time results from the decision uncertainty that accompanies successful conflict detection (Botvinick, 2007). Now, given that accuracy rates for adults only hover around 20% we can expect that the heuristic response will be dominant in all age groups. However, if in contrast with adults, younger reasoners do not yet detect that their heuristic answer conflicts with logical considerations, younger reasoners should not show the increased processing time. Hence, contrasting the response latencies for conflict and no-conflict problems will allow me to test the presence of conflict detection across the preadolescent to young adulthood age range of interest. Note that the critical prediction here concerns the interaction between age group and problem version. Clearly, there are numerous factors that might result in developmental changes in children's general cognitive processing speed (e.g., older children might simply read faster etc.). However, such general factors will equally affect the processing of conflict and no-conflict problems. By focusing on the conflict contrast (i.e., conflict latency minus no-conflict latency) across age groups I sidestep such possible confounds.

As I documented in part A of this thesis, the conflict detection studies with adults established that the longer processing time also boosts recall of conflict problem information after the experiment (e.g., De Neys & Glumicic, 2008; Franssens & De Neys, 2009). In my planned developmental studies I will validate the latency findings with such a recall based measure. After participants have finished solving all reasoning problems they will be presented with an unannounced recall task in which they will be asked to recall critical problem information of the conflict and no-conflict problems. For example, for the base-rate problems participants will be asked to recall the base-rate information. For the conjunction problems and syllogisms participants will be presented with a list of conclusions and conjunctive statements and will need to indicate whether each statement was presented in the reasoning task or not. Successful conflict detection and longer problem processing time should result in a better memorization performance for conflict than no-conflict problems. Hence, comparing memory performance across age groups will allow me to validate the latency findings. Once again, as in all my studies, the critical question concerns the interaction between age and problem type. Clearly, one can envisage that general cognitive maturation will boost the overall memory performance in older age groups. However, since such general memory improvement will equally affect the recall on conflict and no-conflict problems, the relative recall difference across age groups will still give us a clear indication of specific improvements in conflict detection.

Confidence study. Successful conflict detection has also been shown to result in a decreased response confidence (Botvinick, 2007; De Neys et al., 2011). Adults typically indicate that they are less confident that their response is correct after solving conflict than no-conflict problems. This measure directly indicates that adults detect that their heuristic response on the conflict problems is not fully warranted (De Neys et al., 2011). In an additional behavioral study, I will use this response confidence measure as an extra developmental conflict detection index. After each reasoning problem participants will be asked to indicate how confident they are that their response is correct on a simple rating scale. If younger reasoners do not yet detect that their heuristic response conflicts with the logical norm, they should not show the decreased confidence after solving conflict vs. no-conflict problems. Note that in line with my hypothesis, De Neys et al. (2011) presented some initial confidence pilot data that suggested that a convenience sample of 13-year old adolescents showed less conflict sensitivity than an adult control group. The planned large scale confidence study will allow me to extend and validate these findings.

Planning: Both for the latency/recall and confidence studies I envisage to test participants from grade 3 (i.e., average age 8+ years) to grade 12 (i.e. average age +17 years) and a group of young adult university students. Children will be tested at their local school. For the latency studies I will use a computerized testing procedure. Reasoning problems will be presented on laptops and participants will be tested in small groups during course breaks. The confidence studies will be paper-and-pencil experiments and will be run in group sessions in each classroom. I plan to test about 30 participants at each grade level. In both studies each participant will be presented with 8 items (4 conflict and 4 no-conflict problem versions) of each of my three reasoning tasks of interest. Hence, each participants will solve a total of 24 reasoning problems. Presentation order of the reasoning tasks and problem order in each task will be fully counterbalanced.

Neuroscientific studies *(Workpackage 3)*

In a planned third phase of the project I will use neuroscientific testing procedures to extend the behavioral findings and start exploring the neurological basis of development changes in conflict detection efficiency during thinking. My planned studies will primarily focus on the Anterior Cingulate Cortex (ACC), the critical brain structure that is assumed to mediate conflict detection during thinking (e.g., Botvinick et al., 2004; De Neys et al., 2008; Yeung et al., 2011) and which is expected to operate less efficiently for young adolescents (e.g., Davies et al., 2004; Fitzgerald et al., 2010; Santesso & Segalowitz, 2008). I plan a set of functional magnetic imaging (fMRI) studies and combined electroencephalogram (EEG) and electrodermal skin conductance response (SCR) recording studies to test the hypotheses. I describe these protocols in more detail below.

fMRI studies. As I noted, in my work with adults I already used fMRI to monitor ACC activation during thinking (De Neys, Vartanian, & Goel, 2008). Our results showed that the ACC activation peaked when participants solved conflict vs. no-conflict base-rate problems even though participants typically failed to give the correct response. Hence, the ACC activation pattern suggested that adults detect the conflict between their heuristic response and the logically cued response when solving classic conflict problems. In my future developmental research I plan to use fMRI to contrast the ACC activation during solving of conflict and no-conflict problems in different age groups. If younger reasoners do not yet manage to monitor their heuristic inferences for conflict with logical considerations, they should not show the differential ACC activation for conflict vs. no-conflict problems that has been observed in the studies with adults. This will allow me to validate my behavioral findings and directly document the hypothesized developmental changes in ACC functionality (e.g., Davies et al. , 2004; Fitzgerald et al., 2010). Clearly, although I will a priori focus on ACC region-of-interest analyses, I will also perform whole brain-analyses to identify possible other conflict sensitive brain areas that might show developmental modulation. The study design and stimuli presentation format will be modeled after the De Neys et al. (2008) study but obviously I will use the age appropriate problem content from the database I plan to construct in the first project phase (WP 1).

ERP/SCR studies. To validate and refine my planned neuoroscientific fMRI work I also want to run combined EEG/SCR studies. Classic EEG recordings in basic cognitive control studies suggest that conflict detection in the ACC generates a very specific brain potential (i.e., the Error Related Negativity or ERN, e.g., Gehring & Fencsik, 2001). The ERN is a sharp negative voltage deflection in the EEG that is maximal at frontomedial recording sites (e.g., Cz) and typically peaks about 50 ms after an erroneous response. The ERN is believed to reflect executive control activity associated with the monitoring of conflict and error (Amodio et al., 2004; Compton et al., 2008; but see Burle et al., 2008). Now, if the claim that adult reasoners always detect that their heuristic response on conflict problems deviates from the logical response and is not fully warranted, one predicts to find the ERN for incorrectly solved conflict problems. If younger reasoners do not yet detect this intrinsic conflict, they should not show the ERN (or show decreased ERN amplitude). Hence, contrasting the ERN amplitude across age groups will allow me to find converging evidence for the fMRI findings.

In addition, while participant's EEG is recorded I will also simultaneously record changes in their skin conductance response (SCR) from electrodes attached to the palm of their hand. As I clarified in Part A of the thesis, the ACC has been implicated in autonomic nervous system modulation (Critchley, Tang, Glaser, Butterworth, & Dolan, 2005; Hajcak, McDonald, & Simons, 2003). More specifically, it has been argued that autonomic arousal associated with conflict detection gives rise to an increased skin conductance response (e.g., Kobayashi et al., 2007). In their work with adults, De Neys et al. (2010) indeed established that adult reasoners also show an SCR amplitude spike when solving conflict vs. no-conflict syllogisms. Bluntly put, adults seem to literally "feel" the conflict they are detecting. Obviously, a potential conflict detection failure in younger age groups should result in an absence of this SCR spike.

Note that the SCR/EEG co-registration will also allow me to identify a potential dissociation between the behavioral and neuroscientific detection measures. Scarce research with lesion patients has suggested that a neurological conflict signal and its behavioral expression might diverge (Naccache et al., 2005). More specifically, it might be that the ACC is already signaling the presence of conflict at the neurological level but that this signal is not getting translated, for example, into a decreased behavioral confidence. Such genuine dissociation is an unlikely, theoretical possibility, but note that in case that it would occur the SCR recording will allow me to clearly identify it. The lack of a behavioral conflict marker has been specifically attributed to the absence of an autonomic (i.e., SCR) conflict response (e.g., Naccache et al., 2005). Hence, in case I do find fMRI/EEG evidence pointing towards increased ACC activation at age X in the absence of a behavioral conflict signal, this should be accompanied by a lack of SCR boost in my neuroscientific studies.

Planning. In the planned behavioral studies in WP1 and WP2 I envisage to test a large number of participants (+300) at each grade level across the 8-18 year age span of interest. Obviously, the financial and time-limitations associated with neuroscientific research will prevent me from adopting this approach in this phase. Hence, in this third neuroscientific phase or workpackage I will narrow down groups of interest based on my behavioral testing in the second phase. I intend to restrict the studies to three specific age groups of interest (e.g., one group that shows no detection, one age group that shows intermediate detection and on group that show perfect/adult like detection) and test approximately 15 participants in each group. This allows me to meet my objective while keeping the planning realistic. Further note that I will need to present a far greater number of items in the fMRI and EEG studies (i.e., +50 conflict and +50 no-conflict items) to get a reliable signal. Therefore, it is not feasible to present multiple reasoning tasks to the same participants. To test the generalizibility of the findings I will initially select two reasoning tasks of interest and run separate studies with each task.

Scanner time for the fMRI studies will be booked at the CYCERON institute in Caen with which the LaPsyDE group has a privileged working relation. The EEG will be run at the local LaPsyDE lab which recently acquired a state-of-the-art high density, 256 channel EEG system dedicated to developmental testing. Participants will be recruited through the LaPsyDE network.

Conclusion

In Part A of this thesis I documented how my previous work on conflict detection during thinking with adults suggests that biased thinking primarily results from an inhibition failure and not from a mere failure to detect that cued intuitions conflict with normative logical considerations. However, younger children might still have difficulties detecting this mere conflict. This implies that children might benefit from a different type of intervention program. Unfortunately, the development of the critical conflict detection process during thinking is not well-documented. As I stipulated here, the primary aim of my planned research program in the coming years is to address this shortcoming. I believe that the planned experimental workpackages that I have in mind will result in a full specification of the development of conflict detection during reasoning at the behavioral and neurological level. Thereby the project will directly identify possible changes in the locus of biased thinking throughout our development. At the theoretical level this research will have important implications for our view of human rationality. At an applied level the fundamental research in the project will be a critical and much needed building block to develop efficient training programs to help children avoid biased thinking.

Table 1. Illustrations of the classic reasoning tasks that will be used in the planned studies. The left panel (A) shows the classic, standard versions and the right panel (B) the control versions. The standard versions cue a heuristic response that conflicts with the correct logical response (i.e., the response considered correct according to standard logic or probability theory principles). In the control versions small content transformations guarantee that the cued heuristic response is consistent with the logical response.

A. Standard “Conflict” versions	**B. Control “No conflict” versions**
Conjunction fallacy task: Bill is 34. He is intelligent, punctual but unimaginative and somewhat lifeless. In school, he was strong in mathematics but weak in social studies and humanities. Which one of the following statements is most likely? a. Bill plays in a rock band for hobby* b. Bill is an accountant and plays in a rock band for hobby+	**Conjunction fallacy task:** Bill is 34. He is intelligent, punctual but unimaginative and somewhat lifeless. In school, he was strong in mathematics but weak in social studies and humanities. Which one of the following statements is most likely? a. Bill is an accountant*+ b. Bill is an accountant and plays in a rock band for hobby
Base-rate neglect task: A psychologist wrote thumbnail descriptions of a sample of 1000 participants consisting of 995 females and 5 males. The description below was chosen at random from the 1000 available descriptions. 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. Which one of the following two statements is most likely? a. Jo is a man+ b. Jo is a woman*	**Base-rate neglect task:** A psychologist wrote thumbnail descriptions of a sample of 1000 participants consisting of 995 males and 5 females. The description below was chosen at random from the 1000 available descriptions. 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. Which one of the following two statements is most likely? a. Jo is a man*+ b. Jo is a woman
Belief bias syllogisms task: Premises: All vehicles have wheels Boats are vehicles Conclusion: Boats have wheels a. The conclusions follows logically* b. The conclusion does not follow logically+	**Belief bias syllogisms task:** Premises: All vehicles have wheels Bikes are vehicles Conclusion: Bikes have wheels a. The conclusions follows logically*+ b. The conclusion does not follow logically

** = logical response, + = heuristic response*

References

Alexander, B. (2009). *Amid swine flu outbreak, racism goes viral*. NBC Los Angeles. Retrieved June 11, 2009, from http://www.msnbc.msn.com/id/30467300/.

Amodio, D. M., Harmon-Jones, E., Devine, P. G., Curtin, J. J., Hartley, S. L., & Covert, A. E. (2004). Neural signals for the detection of unintentional race bias. *Psychological Science*, 15, 88-93.

Ball, L. J., Philips, P., Wade, C. N., & Quayle, J. D. (2006). Effects of belief and logic on syllogistic reasoning: Eye-movement evidence for selective processing models. *Experimental Psychology, 53*, 77-86.

Barrouillet, P. (2011). Dual proces theories of reasoning: The test of development. *Developmental Review, 31*, 151-179.

Bonner, C., & Newell, B. R. (2010). In conflict with ourselves? An investigation of heuristic and analytic processes in decision making. *Memory & Cognition, 38*, 186-196.

Bonnefon, J. F. (2011). Le raisonneur et ses modèles [The reasoner and his models]. Grenoble, France : Presses Universitaires de Grenoble.

Botvinick, M. (2007). Conflict monitoring and decision making: Reconciling two perspectives on anterior cingulate function. *Cognitive, Affective and Behavioral Neuroscience, 7*, 356-366

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

Burle, B., Roger, C., Allain, S., Vidal, F., & Hasbroucq, T. (2008). Error negativity does not reflect conflict: A reappraisal of conflict monitoring and anterior cingulate cortex activity. *Journal of Cognitive Neuroscience, 20,* 1637-1655.

Brainerd, C. J., & Reyna, V. F. (2001). Fuzzy-trace theory : Dual processes in memory, reasoning, and cognitive neuroscience. In H. W. Reese & R. Kail (Eds.), *Advances in child development and behavi*or (Vol. 28, pp. 41-100). San Diego: Academic Press.

Compton, R. J., Robinson, M. D., Ode, S., Quandt, L. C., Fineman, S. L., & Carp, J. (2008). Error-monitoring ability predicts daily stress regulation. *Psychological Science, 19*, 702-708.

Critchley, H. D., Tang, J., Glaser, D., Butterworth, B., & Dolan, R. J. (2005). Anterior cingulated activity during error and autonomic response. *Neuroimage, 27*, 885-895.

Davies, P. L., Segalowitz, S. J., & Gavin, W. J. (2004). Development of response-monitoring ERPs in 7-to 25-year-olds. Developmental Neuropsychology, 25(3), 355-376.

Denes-Raj, V., & Epstein, S. (1994). Conflict between intuitive and rational processing: When people behave against their better judgement. *Journal of Personality and Social Psychology, 66,* 819-829.

De Neys, W. (2012). Bias and conflict: A case for logical intuitions. *Perspectives on Psychological Science, 7,* 28-38.

De Neys, W., Cromheeke, S., & Osman, M. (2011). Biased but in doubt: Conflict and decision confidence. *PLoS ONE, 6*, e15954.

De Neys, W., & Franssens, S. (2009). Belief inhibition during thinking: Not always winning but at least taking part. *Cognition, 113*, 45-61.

De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition, 106*, 1248-1299.

De Neys, W., Moyens, E., & Vansteenwegen, D. (2010). Feeling we're biased: Autonomic arousal and reasoning conflict. *Cognitive, Affective, & Behavioral Neuroscience, 10*, 208-216.

De Neys, W., & Van Gelder, E. (2008). Logic and belief across the life span: The rise and fall of belief inhibition during syllogistic reasoning. *Developmental Science, 12*, 123-130.

De Neys, W., & Vanderputte, K. (2011). When less is not always more: Stereotype knowledge and reasoning development. *Developmental Psychology, 47*, 432-441.

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.

Epstein, S. (1994). Integration of the cognitive and psychodynamic unconscious. *American Psychologists, 49*, 709-724.

Epstein, S., & Pacini, R. (1999). Some basic issues regarding dual-process theories from the perspective of cognitive-experiential self-theory. In S. Chaiken & Y. Trope (Eds.), *Dual Process Theories in Social Psychology* (pp. 462-482). New York: Guilford Press.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data.* Cambridge, MA: MIT Press.

Evans, J. St. B. T. (1984). Heuristic and analytic processing in reasoning. *British Journal of Psychology, 75*, 451-468.

Evans, J. St. B. T. (2002). Logic and human reasoning: An assessment of the deduction paradigm. *Psychological Bulletin, 128*, 978-996.

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.

Evans, J. St. B. T. (in press). How many dual process theories do we need: One, two or many? In J. St. B. T. Evans & K. Frankish (Eds.), *In Two Minds: Dual Processes and Beyond*. Oxford: Oxford University Press.

Evans, J. St. B. T., & Over, D. E. (1996). *Rationality and reasoning*. Hove, UK: Psychology Press.

Fitzgerald, K. D., Perkins, S. C., Angstadt, M., Johnson, T., Stern, E. R., Welsh, R. C., et al. (2010). The development of performance-monitoring function in the posterior medial frontal cortex. Neuroimage, 49, 3463-3473.

Franssens, S., & De Neys, W. (2009). The effortless nature of conflict detection during thinking. *Thinking & Reasoning, 15*, 105-128.

Gehring, W. S., & Fencsik, D. E. (2001). Functions of the medial frontal cortex in the processing of conflict and errors. *Journal of Neuroscience, 21*, 9430-9437.

Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Grinband, J., Savitskaya, J., Wager, T. D., Teichert T., Ferrera, V.P., & Hirsch J. (2011). The dorsal medial frontal cortex is sensitive to time on task, not response conflict or error likelihood. *Neuroimage, 57*, 303-311.

Hajcak, G., McDonald, N., & simons, R. F. (2003). To err is autonomic: Error-related brain potentials, ANS activity, and post-error compensatory behaviour. *Psychophysiology, 40,* 895-903.

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

Handley, S. J., Newstead, S. E., & Trippas, D. (2011). Logic, beliefs, and instruction: A test of the default interventionist account of belief bias. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 37*, 28-34.

Hertwig, R., & Gigerenzer, G. (1999). The 'conjunction fallacy' revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision-making, 12*, 275-305.

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. (2000). Inhibition and cognitive development: Object, number, categorization, and reasoning. *Cognitive Development, 15,* 63-73.

Houdé, O. (2007). First insights on neuropedagogy of reasoning. *Thinking & Reasoning, 13*, 81-89.

Houdé, O., Zago, L., Mellet, E., Moutier, S., Pineau, A., Mazoyer, B., & Tzourio-Mazoyer, N. (2000). Shifting from the perceptual brain to the logical brain: The neural impact of cognitive inhibition training. *Journal of Cognitive Neuroscience, 12*, 721-728.

Jacobs, J. E., & Klaczynski, P. A. (2002). The development of decision making during childhood and adolescence. *Current Directions in Psychological Science, 4*, 145-149.

Just, M. A., Carpenter, P. A., & Wooley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 111*, 228-238.

Kahneman, D. (2002, December). 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

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgement. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), Heuristics and Biases: The Psychology of Intuitive Judgement (pp. 49-81).

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgement under Uncertainty*: *Heuristics and Biases*. Cambridge, MA: Cambridge University Press.

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

Kobayashi, N., Yoshino, A., Takahashi, Y., & Nomura, S. (2007). Autonomic arousal in cognitive conflict resolution. *Autonomic Neuroscience: Basic and Clinical, 132*, 70-75.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*, 1121-1134.

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., & Barrouillet, P. (2004). Why is understanding the development of reasoning important?. *Thinking and Reasoning, 10*, 113-121.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General, 130*, 621-640.

Moors, A., & De houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin, 132*, 297-326.

Morris, A. K. (2000). Development of logical reasoning: Children's ability to verbally explain the nature of the distinction between logical and nonlogical forms of argument. *Developmental Psychology, 36*, 741-758.

Morsanyi, K., Handley, S. J. (2008). How smart do you need to be to get it wrong? The role of cognitive capacity in the development of heuristic-based judgment. *Journal of Experimental Child Psychology, 99*, 18-36.

Morsanyi, K., Handley, S. J. (2012). Logic feels so good-I like it! Evidence for intuitive detection of logicality in syllogistic reasoning. *J Exp Psychol Learn Mem Cogn.*

Moutier, S., Angeard, N., & Houdé, O. (2002). Deductive reasoning and matching-bias inhibition training : Evidence from a debiasing paradigm. *Thinking and Reasoning, 8,* 205-224.

Moutier, S., & Houdé, O. (2003). Judgment under uncertainty and conjunction fallacy inhibition training. *Thinking and Reasoning, 9*, 185-201.

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

Naccahe, L., Dehaene, S., et al. (2005). Effortless control: executive attention and conscious feeling of mental effort are dissociable. *Neuropsychologia, 43*, 1318-1328.

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

Oaksford, M., & Chater, N. (1998). *Rationality in an uncertain world: Essays on the cognitive science of human reasoning*. Hove, UK: Psychology Press.

Payne, J. W. (1994). Thinking aloud: Insights into information processing. *Psychological Science, 5*, 241-248.

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., Nieuwenhuis, S. (2004). The role of the medial frontal cortec in cognitive control. *Science, 306*, 443-447.

Santesso, D. L., & Segalowitz, S. J. (2008). Developmental differences in effor-related ERPs in middle- to late-adolescent males. Developmental Psychology, 44(1), 205-217.

Simoneau, M., & Markovits, H. (2003). Reasoning with premises that are not empirically true : Evidence for the role of inhibition and retrieval. *Developmental Psychology, 39*, 964-975.

Shynkaruk, J. M., & Thompson, V. A. (2006). Confidence and accuracy in deductive reasoning. *Memory and Cognition, 34*, 619-632.

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

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.

Stanovich, K. E., West, R. F., & Toplak, M. E. (2011). The complexity of developmental predictions from dual process models. *Developmental Review, 31*, 103-118.

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

Stupple, E. J. N., & Ball, L. J. (2008). Belief-logic conflict resolution in syllogistic reasoning: Inspection-time evidence for a parallel-process model. *Thinking & Reasoning, 14*, 168-181.

Téglas, E;, Girotto, V., Gonzalez, M., & Bonatti, L. L. (2007). Intuitions of probabilities shape expectations about the future at 12 months and beyond. *Proceedings of the National Academy of Sciences, 104*, 19156-19159.

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

Thompson, V. A., Striemer, C. L., Reikoff, R., Gunter, R. W. & Campbell, J. I. D. (2003). Syllogistic reasoning time: Disconfirmation disconfirmed. *Psychonomic Bulletin & Review 10*, 184–189.

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

Villejoubert, G. (2009). Are representativeness judgments automatic and rapid? The effect of time pressure on the conjunction fallacy. *Proceedings of the Annual Meeting of the Cognitive Science society, 30*, 2980-2985.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology, 60*, 273-281.

Wason, P. C. (1983). Realism and rationality in the selection task. In J. St. B. T. Evans (Ed.), *Thinking and Reasoning* (pp. 45-75). London: Routledge.

Yeung N., Cohen J. D. , & Botvinick, M. M. (20111). Errors of interpretation and modeling: a reply to Grinband et al. *Neuroimage, 57*, 316-319.

Appendix

Résumé (French summary)

Détection de conflits durant le raisonnement: Résultats empiriques et perspective développementale

Le jugement humain est souvent biaisé par des intuitions erronées. Considérons, par exemple, les peurs vis-à-vis du virus H1N1 qui s'est répandu récemment à travers le monde. Les médias ont communément mis sur ce virus le nom de grippe porcine ou de grippe mexicaine, bien qu'il ne soit plus porté par les espèces porcines et qu'il se soit déjà propagé dans le monde entier au moment de l'annonce de l'épidémie. Par conséquent, consommer du porc ou dîner dans votre restaurant mexicain local ne posait pas de risques sanitaires clairs. L'Organisation Mondiale de la Santé s'est efforcée d'informer le public mais la simple association intuitive avec le nom du virus semblait néanmoins posséder une attraction inébranlable sur le comportement des individus : beaucoup d'entre nous ont arrêté de manger dans les restaurants mexicains, les responsables haïtiens ont refusé l'aide d'un navire remplit d'aliments provenant du Mexique, le marché des poitrines de porc s'est effondré à Wall Street, et le gouvernement égyptien a même ordonné à ses agriculteurs de tuer tous leurs porcs (Alexander, 2009). D'un point de vue logique, aucune de ces mesures n'était efficace pour enrayer la propagation du virus ou éviter la contamination mais les individus ont intuitivement estimé qu'ils feraient simplement mieux d'éviter tout contact avec les Mexicains ou la viande de porc.

Des décennies de recherches sur le raisonnement et la prise de décision ont montré que des pensées intuitives similaires biaisaient les jugements des individus dans un large éventail de situations et de tâches (Evans, 2003; Kahneman, Slovic, & Tversky, 1982). De manière générale, les raisonneurs humains ont une forte de tendance à baser leurs jugements sur des impressions intuitives, rapides, plus que sur un raisonnement cognitivement coûteux. Bien que cette pensée intuitive, ou appelée autrement « heuristique », soit souvent utile, elle délivre parfois des réponses qui rentrent en conflit avec des considérations logiques, normatives, pour finalement biaiser notre prise de décision. Dès lors, pour un raisonnement solide, il est primordial que les raisonneurs contrôlent leurs intuitions heuristiques lorsqu'elles rentrent en conflit avec les principes logiques, et inhibent ces intuitions tentantes dans le cas où un tel conflit est détecté (Houdé, 2000, 2007; Evans, 2007; Stanovich & West, 2008; De Neys & Glumicic, 2008; Barrouillet, 2011).

Le processus de détection de conflit est un élément clé de toutes les théories du raisonnement et de la prise de décision. Malheureusement, le processus est mal compris et les points de vue divergent quant à son efficacité. Par exemple, un certain nombre d'auteurs ont fait valoir que la détection de conflit au cours de la pensée était infructueuse (e.g. Evans, 1984; Kahneman & Frederick, 2002). Selon ces auteurs, ce biais répandu peut être attribué à un échec du contrôle de notre intuition. En raison de ce manque de contrôle, les individus échoueraient simplement à détecter que la réponse intuitive entre en conflit avec les considérations normatives. Dit simplement, les individus seraient biaisés parce qu'ils ne remarquent pas que leur intuition est mauvaise.

Toutefois, d'autres auteurs ont suggéré que la détection de conflit était tout à fait réussie (e.g. Epstein, 1994; Sloman, 1996). Selon ces auteurs, il n'y a aucune faille du processus de détection. Les individus remarquent que la réponse intuitive est en conflit avec les considérations normatives. Le problème, cependant, est que malgré cette connaissance, ils ne réussissent pas toujours à empêcher et à écarter les croyances intuitives tentantes. Ainsi, les individus «agissent à l'encontre de leur meilleur jugement » (traduit de l'anglais, voir Denes-Raj & Epstein, 1994, p. 1) lorsqu'ils donnent une réponse heuristique non justifiée : ils détectent qu'ils sont biaisés, mais échouent simplement à bloquer la réponse biaisée. En résumé, dans cette perspective, les décisions biaisées sont attribuées à l'échec de l'inhibition plus qu'à un échec de détection de conflit (voir également Houdé, 2000, 2007).

La clarification concernant l'efficacité du processus de détection de conflit ainsi que la nature du biais est primordiale pour le développement des théories du raisonnement et de la prise de décision. Cette question a également des implications qui s'étendent à notre conception de la rationalité humaine (De Neys, 2012). Le problème, cependant, est qu'il est difficile de décider entre les différentes perspectives basées sur les données traditionnelles dans le champ du raisonnement (Evans, 2007, 2008). Mes travaux de recherche au cours des deux dernières années se sont penchés sur cette lacune. Brièvement, j'ai essayé d'utiliser des mesures de traitement subtiles afin de tester l'efficience de la détection de conflit durant le raisonnement. Dans la première partie de cette synthèse d'HDR, je passe en revue ces travaux et met en avant leurs implications dans le débat sur la rationalité humaine. Dans la seconde partie, j'insiste sur les directions futures et je présente un carnet de route détaillé du programme de recherche développemental que j'envisage de poursuivre dans les années à venir.

Vue générale des découvertes sur la détection de conflit

Mes études portant sur la détection de conflit ont typiquement opposé les traitements des individus faisant face à des problèmes classiques de raisonnement « incongruent », dans lesquels la réponse heuristique est en conflit avec la réponse correcte logique, aux traitements mis en jeu face aux problèmes non-conflit ou « congruents », dans lesquels le conflit n'apparaît pas (voir Table 1, p. 43, pour des exemples). Le taux d'exactitude sur les versions contrôles est typiquement très élevé alors qu'il est, notoirement, faible sur les versions conflit. Toutefois, la contribution essentielle des études portant sur la détection de conflit est qu'elles ont commencé à regarder au-delà de l'exactitude des réponses et se sont centrées sur des mesures plus subtiles, qui ont permis de tester si les individus traitaient les deux types de problèmes différemment. Par exemple, une procédure de base a été de regarder simplement les latences des réponses des individus: j'ai observé que les gens ont besoin de plus de temps pour résoudre les problèmes conflit que les problèmes contrôle (e.g. De Neys & Glumicic, 2008; see also Bonner & Newell, 2010, et Villejoubert, 2009, pour réplications). Clairement, la seule différence entre les deux versions se porte sur la réponse heuristique qui est congruente ou non avec les principes normatifs traditionnels. Si les gens n'étaient que de simples penseurs heuristiques qui ne surveillent pas le potentiel conflit avec ces considérations normatives et les négligent, ils ne devraient pas traiter les deux types de problèmes différemment. Au contraire, les observations des latences soutenaient l'idée que les individus détectaient le conflit entre leur réponse heuristique et la réponse normative.

Un soutien supplémentaire dans le débat sur la détection a été apporté par les données oculométriques qui ont montré que les latences les plus importantes étaient, spécifiquement, accompagnées par une inspection plus longue de l'information normative cruciale du problème (De Neys & Glumicic, 2008). Cette "exploration" a également été observée comme aboutissant à un meilleur rappel de l'information normative cruciale du problème. (e.g. Franssens & De Neys, 2009).

Les résultats comportementaux sur le conflit ont été validés par une approche basée sur l'imagerie cérébrale. Par exemple, dans une étude (De Neys, Vartanian, & Goel, 2008), j'ai utilisé un protocole d'IRMf pour mesurer l'activation d'une zone cérébrale spécifique, le Cortex Cingulaire Antérieur (CCA), considéré comme entrant en jeu lors de la détection de conflit durant les activités de pensée (e.g. Botvinick, Cohen, & Carter, 2004; voir aussi la discussion entre Yeung et al., 2011 et Grindband et al., 2011). Les participants faisaient face à des problèmes conflit classiques et des versions contrôle non-conflit. En accord avec les résultats comportementaux, les résultats ont montré que le CCA était plus activé lorsque les individus résolvaient les versions conflit que lorsqu'ils résolvaient les versions contrôle.

Dans une étude subséquente, la conductance électrique de la peau des participants était enregistrée pour surveiller l'activation du système nerveux autonome pendant qu'ils résolvaient des syllogismes conflit et non-conflit (De Neys, Moyens, & Vansteenwegen, 2010). Les résultats montraient que faire face aux problèmes conflit donnait lieu à un pic d'activation électrodermale clair. Par conséquent, en plus de l'activation du CCA, résoudre des problèmes conflit excite littéralement les patients. Ces signaux de conflit neural ont aussi été observés comme diminuant la confiance subjective des participants en leurs réponses heuristiques : typiquement, les participants indiquaient qu'ils se sentaient moins confiant à propos de leur réponse après avoir répondu à un problème conflit que lorsqu'ils répondaient à un problème contrôle (e.g. De Neys, Cromheeke, & Osman, 2011).

Prises toutes ensembles, les études sur la détection de conflit suggèrent que les raisonneurs détectent le caractère biaisé de leur jugement : bien que les individus soient généralement tentés de donner la réponse heuristique, ils semblent toutefois remarquer que cette dernière est en conflit avec les normes logiques et indiquent que leur réponse n'est pas pleinement justifiée (e.g. De Neys, 2012). Ces découvertes sur la détection de conflit ont été utilisées comme support à l'idée que les biais heuristiques résultaient typiquement d'un échec de l'inhibition (De Neys & Franssens, 2009). Ainsi, le problème ne semble pas être que les individus ne détectent pas que les réponses heuristiques sont contestables et nécessitent d'être écartées, mais plutôt que les individus échouent à compléter le processus d'inhibition cognitivement coûteux. Cette idée est consistante avec la littérature dans le champ du raisonnement, qui a souligné le rôle crucial des capacités de traitements d'inhibition, afin d'outrepasser les réponses heuristiques erronées (e.g. Houdé, 1997, 2000, 2007; Moutier, Plagne-Cayeux, Melot, & Houdé, 2006; Stanovich & West, 2000). Ceci suggère que les individus sont plus logiques que ce que leurs réponses biaisées suggèrent (De Neys, 2012).

Perspectives des Recherches Développementales

Bien que des décennies de recherche sur le raisonnement et la prise de décision ont montré que le jugement humain est souvent biaisé par les heuristiques intuitives, mon précédent travail sur la détection de conflit durant les activités de réflexion a néanmoins indiqué que malgré leur réponse biaisée, les adultes détectent que leur réponse n'est pas pleinement justifiée et rentre en conflit avec des considérations logiques. Cette sensibilité au conflit suggère que les individus sont biaisés parce qu'ils échouent à outrepasser l'intuition tentante et sont plus intelligents que ce que leurs erreurs suggèrent.

Cependant, il est crucial ici de noter que mes études portant sur la détection de conflit n'ont été conduites seulement qu'avec des participants adultes. Le développement du processus de détection de conflit pendant les activités de pensée n'a reçu que peu d'attention. Par conséquent, nous ne pouvons exclure que les échecs de détection de biais ont joué un rôle crucial plus tôt dans le développement de notre raisonnement. Cette hypothèse a reçu un soutien d'études neurologiques de base qui suggèrent que le Cortex Cingulaire Antérieur (CCA), la structure cérébrale cruciale supposée intervenir dans la surveillance de conflits élémentaires, est relativement lente dans sa maturation et n'atteindrait sa pleine fonctionnalité qu'au milieu de l'adolescence (e.g. Davies, Segalowitz, & Gavin, 2004; Fitzgerald et al., 2010; Santesso & Segalowitz, 2008). Ceci semblerait indiquer qu'il pourrait y avoir une transition critique vis à vis du site du biais d'heuristique durant le développement humain. Ainsi, alors que les adultes seraient biaisés parce qu'ils échouent à inhiber la réponse heuristique après avoir correctement détecté le conflit, les jeunes raisonneurs seraient eux biaisés parce qu'ils échouent à détecter la nécessité d'inhiber la réponse heuristique qui apparaît en premier lieu.

Bien sûr, il est particulièrement important d'un point de vue théorique d'identifier les changements possibles dans la nature ou le site des biais heuristiques tout au long de notre développement (e.g. Barrouillet, 2011). Cependant, à un niveau plus appliqué, établir si oui ou non les biais d'heuristiques sont dus à un échec de détection du biais est peut être encore plus crucial. Explorer le développement de la détection de conflit au cours de la pensée est primordiale pour développer des programmes d'intervention efficaces visant à dé-biaiser les activités de pensée des enfants. Notons que les programmes d'interventions efficaces existants pour les adultes et les adolescents plus âgés se sont focalisés sur l'exercice des capacités de traitements inhibiteurs des raisonneurs (e.g. Houdé, 2007; Houdé et al., 2000; Moutier, Angeard, & Houdé, 2002; Moutier & Houdé, 2003). Toutefois, si les jeunes enfants ne détectent pas que leur réponse heuristique est erronée, l'entraînement à l'inhibition donnera lieu à moins de résultats optimaux. En effet, toute augmentation de capacité des traitements inhibiteurs semble inutile si l'on n'est pas capable de déterminer si oui ou non il est nécessaire d'inhiber en premier lieu. Par conséquent, examiner les compétences des enfants dans la détection est indispensable pour déterminer quels composants les programmes d'intervention pédagogique futurs doivent cibler pour dé-biaiser correctement les activités de pensée des enfants. Malgré la claire importance éducative et sociétale de cette question, le manque de recherches portant sur le développement de la détection de conflit ne nous permet pas actuellement d'y répondre. L'objectif principal de mes recherches futures sera de combler ce manque dans le domaine du raisonnement et de documenter exhaustivement le développement de l'efficacité de la détection de conflit chez les enfants.

Pour atteindre mon objectif, je prévois de différencier l'efficacité de la détection dans des groupes d'âge allant de l'âge scolaire élémentaire préadolescent aux jeunes adultes (i.e., tranche d'âge ciblée de 8 à 18 ans). Le choix de cette tranche d'âge est inspiré par les études de neurosciences qui suggèrent que le Cortex Cingulaire Antérieur n'atteindrait son plein fonctionnement qu'au milieu de l'adolescence (i.e. vers les 14 ans, voir Davies et al., 2004; Fitzgerald et al., 2010; Santesso & Segalowitz, 2008). Ainsi, dans la tranche d'âge sélectionnée, je devrais réussir à identifier les changements développementaux cruciaux dans la détection de conflit lors du raisonnement.

Dans les études prévues, les participants se verront toujours présenter des sets de tâches de raisonnement incongruentes et congruentes. La détection de conflit sera évaluée à l'aide d'un large éventail de procédures comportementales et neuroscientifiques. Les études prévues peuvent être regroupées en trois phases ou volets de recherche, qui seront chronologiquement construits l'un sur l'autre. Dans la phase initiale ou le premier volet, je vais construire une base de données avec des matériaux appropriés aux différents âges, qui seront utilisés pour mes expériences sur le développement de la détection de conflit. Dans le second volet, je vais me baser sur des procédures de tests comportementaux (i.e., la latence, le rappel, et les mesures de confiance en la réponse) afin de différencier l'efficacité de la détection de conflit entre les différents groupes d'âge. Dans le troisième volet, je vais préciser et valider les résultats à l'aide de procédures de tests neuroscientifiques (i.e. IRMf, EEG, et mesures de conductance de la peau) qui se concentreront directement sur le rôle du Cortex Cingulaire Antérieur.

Prises toutes ensemble, les études prévues donneront lieu à une spécification comportementale et neurale complète du développement des processus de détection de conflit durant les activités de pensée. Ainsi, le projet identifiera directement les changements possibles dans le site de la pensée biaisée tout au long de notre développement et servira de pierre angulaire indispensable pour développer des programmes d'entraînement efficaces dans le but d'aider les enfants à éviter des activités de pensée biaisée.

Collection of published papers presented in the thesis

- **De Neys**, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. ***Cognition**, 106*, 1248-1299.

- **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.

- Franssens, S., & **De Neys**, W. (2009). The effortless nature of conflict detection during thinking. ***Thinking & Reasoning**, 15*, 105-128.

- **De Neys**, W., Moyens, E., & Vansteenwegen, D. (2010). Feeling we're biased: Autonomic arousal and reasoning conflict. ***Cognitive, Affective, & Behavioral Neuroscience**, 10*, 208-216.

- **De Neys**, W., Cromheeke, S., & Osman, M. (2011). Biased but in doubt: Conflict and decision confidence. ***PLoS ONE**, 6*, e15954.

- **De Neys**, W., & Franssens, S. (2009). Belief inhibition during thinking: Not always winning but at least taking part. ***Cognition**, 113*, 45-61.

- **De Neys**, W. (2012). Bias and conflict: A case for logical intuitions. ***Perspectives on Psychological Science**, 7,* 28-38.

*See www.wdeneys.org/publications for electronic pdf versions

Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Cognition 106 (2008) 1248–1299
COGNITION
www.elsevier.com/locate/COGNIT

Conflict monitoring in dual process theories of thinking ☆

Wim De Neys [a,*], Tamara Glumicic [b]

[a] *Lab Experimental Psychology, University of Leuven, Tiensestraat 102, 3000 Leuven, Belgium*
[b] *York University, 4700 Keele Street, Toronto, ON, Canada*

Received 17 October 2006; revised 5 June 2007; accepted 5 June 2007

Abstract

Popular dual process theories have characterized human thinking as an interplay between an intuitive-heuristic and demanding-analytic reasoning process. Although monitoring the output of the two systems for conflict is crucial to avoid decision making errors there are some widely different views on the efficiency of the process. Kahneman [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] and Evans [Evans, J. St. B. T. (1984). Heuristic and analytic processing in reasoning. *British Journal of Psychology, 75*, 451–468], for example, claim that the monitoring of the heuristic system is typically quite lax whereas others such as Sloman [Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*, 3–22] and Epstein [Epstein, S. (1994). Integration of the cognitive and psychodynamic unconscious. *American Psychologists, 49*, 709–724] claim it is flawless and people typically experience a struggle between what they "know" and "feel" in case of a conflict. The present study contrasted these views. Participants solved classic base rate neglect problems while thinking aloud. In these problems a stereotypical description cues a response that conflicts with the response based on the analytic base rate information. Verbal protocols showed no direct evidence for an explicitly experienced conflict. As Kahneman and Evans predicted, participants hardly ever mentioned the base rates and seemed to base their judgment exclusively on heuristic reasoning. However, more implicit measures of conflict detection such as

☆ This manuscript was accepted under the editorship of Jacques Mehler.

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

doi:10.1016/j.cognition.2007.06.002

participants' retrieval of the base rate information in an unannounced recall test, decision making latencies, and the tendency to review the base rates indicated that the base rates had been thoroughly processed. On control problems where base rates and description did not conflict this was not the case. Results suggest that whereas the popular characterization of conflict detection as an actively experienced struggle can be questioned there is nevertheless evidence for Sloman's and Epstein's basic claim about the flawless operation of the monitoring. Whenever the base rates and description disagree people will detect this conflict and consequently redirect attention towards a deeper processing of the base rates. Implications for the dual process framework and the rationality debate are discussed.

Keywords: Reasoning; Decision making; Heuristics and biases; Conflict monitoring; Dual process theories

1. Introduction

In the spring of 2006 racial tensions in Belgium rose to a boiling point after a white, Belgian high school student was violently stabbed to death by two youths thought to be of African decent. A striking aspect of the sad case was how readily many civilians, politicians, and media were willing to blame the African community based on some initial rumors. The violent murder fitted with people's stereotypical (but mistaken) beliefs about Africans' aggressive and criminal nature. What most people disregarded was that, as in most European countries, African immigrants are just a small minority group in Belgium. They are outnumbered by a factor of ten by people with European roots. Logically speaking, in the absence of clear evidence to the contrary it is far more likely that an assailant will come from another ethnic group. However, many people were tempted to neglect this information and readily believed the initial reports about the involvement of the African youths. The ungrounded accusations backlashed when 2 weeks later the actual culprits were identified as being Europeans.

The above case is a regrettable illustration of a common human tendency to base judgments on prior beliefs and intuition rather than on a logical reasoning process. Over the last decades numerous studies have shown that this tendency is biasing performance in many classic reasoning and decision making tasks (Evans, 2002; Tversky & Kahneman, 1974).

Influential dual process theories of thinking have explained people's "rational thinking failure" by positing two different human reasoning systems (e.g., Epstein, 1994; Evans, 1984, in press; Evans & Over, 1996; Goel, 1995; Kahneman, 2002; Kahneman & Frederick, 2005; Sloman, 1996; Stanovich & West, 2000). Dual process theories come in many flavors but generally they assume that a first system (often called the heuristic system) will tend to solve a problem by relying on prior knowledge and beliefs whereas a second system (often called the analytic system) allows reasoning according to logical standards. The heuristic default system is assumed to operate fast and automatically whereas the operations of the analytic system would be slow and heavily demanding of people's computational resources. Dual process theories state that the heuristic and analytic system will often interact in concert. Hence, on these occasions the heuristic default system will provide us with fast, frugal, and correct conclusions. However, the prepotent heuristics can also bias reasoning in situations that require more elaborate, analytic processing. That is, both systems will sometimes conflict and cue different responses. In these cases the analytic system will need to override the belief-based response generated by the heuristic system (Stanovich & West, 2000).

Although the dual process framework has been very influential it has also been criticized. Many researchers have pointed out that the differential processing characteristics of the two systems are not sufficiently specified: Dual process theories nicely describe "what" the two systems do but it is not clear "how" the systems actually operate (Evans, 2007a; Gigerenzer & Regier, 1996; Osman, 2004; Reyna, Lloyd, & Brainerd, 2003; Stanovich & West, 2000). The characterization of the conflict detection process is a crucial case in point. Dual process theories generally state that the analytic system is monitoring the output of the heuristic system. When a conflict with analytic knowledge (e.g., sample size considerations) is detected, the analytic system will attempt to intervene and inhibit the prepotent heuristic response. However, if one looks at the literature it becomes clear that there are some widely different views on the efficiency of the conflict monitoring component during judgement and decision making. This results in a different characterization of the nature of the dominant reasoning error. The classic work of Evans (1984) and Kahneman and colleagues (e.g., Kahneman & Frederick, 2002), for example, claims that the monitoring of the heuristic system is quite lax. It is assumed that by default people will tend to rely on the heuristic route without taking analytic considerations into account. In some cases people can detect the conflict and the analytic system will intervene but typically this will be quite rare. Most of the time people will simply not be aware that their response might be incorrect from a normative point of view. As Kahneman and Frederick (2005, p. 274) put it: "People who make a casual intuitive judgement normally know little about how their judgment come about and know even less about its logical entailments". Thus, in this view people mainly err because they fail to detect a conflict.

In the work of Epstein (1994) and Sloman (1996) one finds a remarkably different view on conflict monitoring and the nature of reasoning errors. These authors assume that in general the heuristic and analytic routes are simultaneously activated and people typically do experience a conflict between two types of reasoning. People would "simultaneously believe two contradictory responses" (Sloman, 1996, p. 11) and therefore "behave against their better judgement" (Denes-Raj & Epstein, 1994, p. 1) when they err. Thus, people would be taking analytic considerations in mind and notice that they conflict with the heuristically cued belief. The problem, however, is that they do not always manage to override the compelling heuristics. In this view there is nothing wrong with the conflict detection process. Errors arise because people fail to inhibit the prepotent heuristic beliefs. Sloman argued that classic reasoning tasks can be thought of as perceptual illusions in this respect. In the Muller–Lyer illusion, for example, perception also tells us that one line is longer than the other while logic tells us that it is not. Even though we can measure the lines and know they are of equal size our perception of them does not change. We simultaneously experience two contradictory beliefs. In order to correctly answer the question about the length of the lines we will need to override the erroneous heuristic perception.

In a recent review, Evans (2007a) has pointed to the inconsistencies in the field. Evans' work indicates that different views on conflict monitoring are not only linked with different views on the nature of reasoning errors (i.e., conflict detection or inhibition failure) but also with a different characterization of the interaction between the analytic and heuristic system (i.e., parallel or serial). Sloman and Epstein assume that whenever people are confronted with a reasoning problem both routes will process it simultaneously. People take analytic considerations into account right from the start and detect possible conflicts with heuristically cued beliefs. Here it is believed that both systems operate in parallel. In Kahneman's framework and Evans' own dual process model, however, only the heuristic route is initially activated. The analytic system is assumed to monitor the output of the heuristic system and might intervene in a later stage when a conflict is detected. As Evans noted, here the interplay between the two systems has a more serial nature.

Based on the available data it is hard to decide between the different models and determine which conflict detection view is correct. Sloman (1996) and Epstein (1994), for example, refer to the outcome of perspective change and instruction experiments in support of their views. It has indeed been shown that simply instructing people to evaluate problems "from the perspective of a statistician" helps boosting their performance. In the same vein Sloman stresses the casual observation that people often have no trouble recognizing their error once it is explained to them. Such observations do suggest that people have readily access to two different modes of reasoning and that they can easily switch between them. However, they do not show that both routes are activated simultaneously. No matter how easily one takes analytic considerations into account when prompted, one cannot conclude that this knowledge was also activated during reasoning in the absence of these prompts.

More compelling evidence for successful conflict detection during decision making comes from a number of intriguing anecdotes and spontaneous reports. Epstein (1994; Denes-Raj & Epstein, 1994; Epstein & Pacini, 1999), for example, repeatedly noted that when picking an erroneous answer his participants spontaneously commented that they did "*know*" that the response was wrong but stated they picked it because it "*felt*" right. Sloman (1996) cites evolutionary biologist Steven Jay Gould who relates experiencing a similar conflict between his logical knowledge and a heuristically cued stereotypical belief when solving Kahneman's and Tversky's infamous "Linda" problem.[1] The problem, however, is that spontaneous self-reports and anecdotes are no hard empirical data. This is perhaps best illustrated by the fact that Kahneman (2002, p. 483) also refers to "casual observation" of his participants to suggest that only in "some fraction of cases, a need to correct the intuitive

[1] Gould (1991) wrote: "I know the [conjunction] is least probable, yet a little homunculus in my head continues to jump up and down, shouting at me – 'but she can't just be a bank teller; read the description'" (p. 469).

judgements and preferences will be acknowledged''. It is clear that in order to conclude something about the efficiency of the conflict detection we need a straightforward empirical test to establish precisely how frequently people experience this conflict. The present study addresses this issue.

Experiment 1 adopted a thinking aloud procedure (e.g., Ericsson & Simon, 1980). The thinking aloud procedure has been designed to gain reliable information about the course of cognitive processes. Participants are simply instructed to continually speak aloud the thoughts that are in their head as they are solving a task. Thinking aloud protocols have been shown to have a superior validity compared to interpretations that are based on retrospective questioning or people's spontaneous remarks (Ericsson & Simon, 1993; Payne, 1994).

Participants were asked to solve problems that were modeled after Kahneman and Tversky's classic (1973) base rate neglect problems. In these problems people first get information about the composition of a sample (e.g., a sample with 995 females and 5 males). People are told that short personality descriptions are made of all the participants and they will get to see one description that was drawn randomly from the sample. Consider the following example:

> In a study 1000 people were tested. Among the participants there were 5 men and 995 women. Jo is a randomly chosen participant of this study.
>
> 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

The normative response based on the group size information is (b). However, people will be tempted to respond (a) on the basis of heuristic beliefs cued by the description.

Given Kahneman and Tversky's (1973) classic findings one can expect that in the majority of cases people will err and pick the heuristically cued response in this task. The crucial question is whether people's verbal protocols indicate that they nevertheless take analytic considerations into account. In this task ''analytic considerations'' can be operationalized as referring to the group size information during the reasoning process (e.g., '' … because Jo's drinking beer and loud I guess Jo'll be a guy, although there *were more women*…''). Such basic sample size reference during the reasoning process can be considered as a minimal indication of successful conflict monitoring. It shows that this information is not simply neglected. If Sloman and Epstein's idea about the parallel operation of the heuristic and analytic route is correct, such references should be found in the majority of cases. If Kahneman and Evans' idea about the lax nature of the conflict monitoring is correct, people will simply not be aware that the base rates are relevant and should hardly ever mention them during decision making.

It should be noted that both camps in the conflict monitoring debate, as the reasoning field at large, have conceptualized the conflict between the analytic and heuristic system as a consciously experienced, verbalizable event. Conflict monitoring is considered as a controlled process arising from the central executive aspect of working memory. Since James (1890) there is indeed a long tradition in psychology to consider such central, controlled (vs. automatic) processing as being consciously experienced (Feldman Barrett, Tugade, & Engle, 2004). However, the available evidence from the cognitive literature suggests that this needs not always be the case (e.g., Pashler, Johnston, & Ruthruff, 2001; Shiffrin, 1988). Although controlled processing can occur with a feeling of conscious deliberation and choice, it needs not (Feldman Barrett et al., 2004).

While it is held that thinking-aloud is an excellent method to tap into the content of conscious thinking it cannot provide us with the information about cognitive processes that do not reach the conscious mind (Crutcher, 1994). Consequently, even if participants do not verbalize their experience of the conflict, one cannot exclude that the conflict monitoring might nevertheless have been successful. To capture such *implicit* detection participants were also presented with an unannounced recall test in our study. After a short break following the thinking-aloud phase participants were asked to answer questions about the group sizes in the previous reasoning task. If people have successfully detected the conflict this implies that the group size has been taken into account and people spent some time processing it. Indeed, the detection of the conflict should trigger analytic system intervention which should result in some further scrutinising of the sample information. In sum, successful conflict detection should be accompanied by a deeper processing of the base rate information which should benefit recall. This recall index does not require that the conflict is consciously experienced and verbalizable.[2]

To validate the recall hypothesis participants were also presented with additional control problems. In the classic base rate problems the description of the person is composed of common stereotypes of the smaller group so that base rates and description disagree. In addition to these classic problems we also presented problems where base rates and description both cued the same response. In these *congruent* problems the description of the person was composed of stereotypes of the *larger* group (e.g., Ferreira, Garcia-Marques, Sherman, & Garrido, 2006). Hence, contrary to the classic (i.e., *incongruent*) problems base rates and description did not conflict and the response could be rightly based on the salient description without further analytic intervention/processing. For a reasoner who neglects the base rates and does not detect the conflict on the classic problems both types of problems will be

[2] Note that we refer to implicit detection to contrast it with the more direct verbal conflict measure. It should be clear that we do not claim that a lack of verbalization necessarily implies that people have no conscious access to the process. The role of verbalization as a prerequisite for conscious processing is a matter of debate (e.g., Moors & de Houwer, 2006). The point is that if conflict detection were successful but not verbalized, the implicit measure still allows us to track it.

completely similar and base rate recall should not differ. However, if one does detect the conflict, the added analytic processing of the base rates should result in a better recall for the classic problems than for the congruent control problems.

In Experiment 2 the conflict monitoring issue is further examined by focusing on participants' problem processing time. A core characteristic of analytic reasoning is that it is slow and time-consuming (e.g., Evans, 2003; Sloman, 1996). While the analytic base rate scrutinizing associated with conflict detection might benefit subsequent recall, it will also take up some additional processing time. Reasoning latencies thereby provide an additional test of the opposing conflict monitoring views. One may assume that people will be fastest to solve the congruent control items since the response can be fully based on mere heuristic reasoning without any further analytic intervention. Correctly solving the classic problems should be slowest since it requires people to detect the conflict and inhibit the heuristic response which are both conceived as time-demanding processes (e.g., De Neys, 2006a). The crucial question concerns the processing time of erroneously solved incongruent problems (i.e., responses on the classic problems based on the description). If people simply fail to detect the conflict and reason purely heuristically, reasoning latencies for incorrectly solved incongruent and correctly solved congruent problems should not differ. If people do detect the conflict, they should take longer to respond to the incongruent problems. Consequently, reasoning latencies for the incorrectly solved incongruent problems should fall somewhere in between those of correctly solved incongruent problems and congruent control problems.

To validate the idea that upon conflict detection people spend specific time processing the base rates Experiment 2 also introduces a rudimentary "moving window" procedure (e.g., Just, Carpenter, & Wooley, 1982). In the experiment the group size information and the description are presented separately. First, the base rates are presented on a computer screen. Next, the description and question are presented and the base rates disappear. Participants have the option of visualizing the base rates afterwards by holding a specific button down. Such base rate reviewing can be used as an additional conflict detection index. One might expect that when people detect that the description conflicts with the previously presented base rates they will spend extra time scrutinizing or "double checking" the base rates. With the present procedure the time spent visualizing the base rates can be used as a measure of this reviewing tendency. Longer overall response latencies after successful conflict detection should thus be accompanied by a stronger tendency to visualize the base rates. If people simply neglect the base rates, there is also no reason to review and visualize them after the initial presentation.

2. Experiment 1

Participants in Experiment 1 solved a set of base rate problems while thinking aloud. In the classic, *incongruent* problems base rates and description conflicted whereas in the *congruent* problems base rates and description were consistent. In addition, participants also received a set of *neutral* problems where the description only mentioned characteristics that were neutral with respect to group membership (e.g., ''the person has black hair and blue eyes''). In these problems the description will not clearly cue a response and will therefore not bias or facilitate decision making. Correct responses will be based on mere base rate scrutinizing. On the congruent and neutral control problems a high number of correct responses is expected. On the incongruent problems one can expect that in the majority of cases people will err and pick the heuristically cued response. The crucial question is to what extent people notice the conflict and refer to the base rates when solving these incongruent problems.

2.1. Methods

2.1.1. Participants

Twelve undergraduate students at York University (Toronto, Canada) participated in return for credit in a psychology course.

2.1.2. Materials

2.1.2.1. Decision making task. Participants solved a total of 18 problems that were modelled after Kahneman and Tversky's (1973) base rate neglect items. Six of these were the crucial *incongruent* problems where 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). There were also six *congruent* control items where the description and the base rates agreed. Finally, we also presented six *neutral* control items where the description only mentioned characteristics that were neutral with respect to group membership while the base rates were indicating which group was larger. The following are examples of the three problem types:

> *Incongruent*
> In a study 1000 people were tested. Among the participants there were 4 men and 996 women. Jo is a randomly chosen participant of this study.
>
> 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
>
> *Congruent*
> In a study 1000 people were tested. Among the participants there were 995 who buy their clothes at high-end retailers and five who buy their clothes at Wal-Mart. Karen is a randomly chosen participant of this study.

Karen is a 33-year-old female. She works in a business office and drives a Porsche. She lives in a fancy penthouse with her boyfriend.

What is most likely?

a. Karen buys her clothes at high end retailers
b. Karen buys her clothes at Wal-Mart

Neutral
In a study 1000 people were tested. Among the participants there were 5 who campaigned for George W. Bush and 995 who campaigned for John Kerry. Jim is a randomly chosen participant of this study.

Jim is 5 ft and 8 in. tall, has black hair, and is the father of two young girls. He drives a yellow van that is completely covered with posters.

What is most likely?

a. Jim campaigned for George W. Bush
b. Jim campaigned for John Kerry

Problems were based on a wide range of stereotypes (e.g., gender, age, race, and job related groups and stereotypical characteristics). Descriptions were selected on the basis of an extensive pilot study where four students constructed a large number of stereotypical and neutral descriptions. Five raters then judged on an 11-point scale how well the generated descriptions fitted each of the two groups referred to in the problems (0 – extremely unlikely that this person belongs to this group, 10 – extremely likely that this person belongs to this group). Selected descriptions for the incongruent and congruent problems moderately but consistently cued one of the two groups whereas selected neutral descriptions had to be as similar as possible. Mean ratings for the descriptions used in the incongruent and congruent problems were 8.1 ($SD = .69$) for the most likely group and 2.6 ($SD = 1.01$) for the least likely one. For the neutral descriptions ratings were 5.5 ($SD = .78$) and 5 ($SD = .7$), respectively.

The different problems were presented with slightly varied base rates. More precisely, for each problem type two problems were presented with a 995/5, two with a 996/4, and two with a 997/3 base rate ratio. While piloting this study some participants reported they simply did not mention the base rates because they were always identical in the different problems. The variation was included to counter such superficial base rate neglect resulting from the repeated testing. Post hoc analyses confirmed that task performance for the three base rate levels did not differ in the present study.

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.

Problems were printed one to a page in a booklet. The first page of the booklet stated the instructions:

> 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.

A complete overview of all 18 problems is presented in the Appendix. The problems were presented in pseudo-random order. Participants always started with an incongruent problem followed by a congruent and neutral problem. The remaining problems were presented in a randomly determined order.

2.1.2.2. Recall task. Participants were asked to write down the base rates for each problem they previously solved. The following is an example of the recall question:

> One of the problems you just solved concerned Jo whose description was drawn at random from a sample of men and women. Try to answer the following questions.
> How many men were there exactly in the study? ____ (write down)
> How many women were there exactly in the study? ____ (write down)

After the base rate question followed two easy filler questions in multiple choice format that referred to the description of the problem. For example:

> Mark which statement is correct:
>
> Jo likes to cruise with friends and drink beer
> Jo loves watching television
> Jo jogs every morning
> Jo is 6.3 ft tall

Performance on these filler problems was uniformly high. Each base rate question together with the two filler questions was printed one to a page in a booklet. Recall questions were presented in the same order as the decision making problems had been solved.

2.1.3. Procedure

Participants were first introduced to the thinking aloud procedure. Participants received the following instructions:

> In this experiment we try to find out how people solve everyday reasoning problems. Therefore, we ask you to "think aloud" when you're solving the

> problems. You start by reading the complete problem aloud. When you're solving the problem you have to say everything that you're thinking about. All inferences you're making, all comments you're thinking of, basically everything that is going through your mind, you have to say aloud. You should be talking almost continuously until you give your final answer. Try to keep on thinking aloud the whole time. Whenever you're not saying anything for a while I'll remind you of this.

Once the participants were clear on the instructions they were presented with the decision making task. After participants had read the instructions for the decision making task the experimenter emphasised the thinking-aloud instructions once more and started the session. The complete session was tape-recorded and later typed out. Coding of the verbal reports simply focused on whether the participants gave the correct answer[3] and whether they referred to the base rate information during decision making. A statement like "…because Jo's drinking beer and likes loud music I guess *Jo'll be a guy*, although there *were more women*" would be coded as an incorrect response since the participant did not pick the response (i.e., women) consistent with the largest sample group and as an instance of base rate mentioning. The following are some straightforward further illustrations of the protocol codings:

> …This guy is an engineer, because he likes computers and science fiction, and he seems like a loner…no wife. (Participant #12, problem #b: incorrect response, base rates not mentioned).

> …It depends how you want to go *if you want to go according to the statistics there is a greater chance he is a lawyer* but because of the things he does, he is introverted, spends his time reading fiction and writing computer games it makes more sense that he is an engineer so…I don't know I will go with that (Participant #1, problem #b: incorrect response, base rates mentioned).

> …ok *5 engineers*…you would think he is an engineer *but cause there were more lawyers* he is a lawyer. (Participant #6, problem #b: correct response, base rates mentioned).

[3] Consistent with previous dual process studies, responses that were in line with the base rates (i.e., selection of the largest group as most likely answer) were labelled as correct answers. It should be noted that especially in the case of the classic, incongruent problems the actual normative status of the 'correct' response is sometimes debated (Gigerenzer, Hell, & Blank, 1988). The present paper is concerned with the empirical question as to what extent people take the base rates into account during decision making whether or not the base rates ultimately turn out to be "normative" or not. Therefore, one can adopt a nominalist stance towards the use of the terms 'correct' and 'error'.

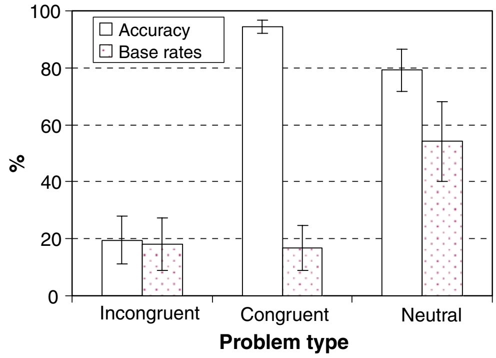

Fig. 1. Mean proportion correct responses and explicit base rate mentioning in verbal protocols. Error bars are standard errors.

In a few cases the participants could not spontaneously decide which answer they considered more likely. In these cases the experimenter asked them to indicate which response they would pick if they were forced to choose. This response was coded as their final answer.

After completing the decision making task, participants had a short break and then were presented with the recall task. The recall task was not announced at the start of the experiment so participants did not know base rate recall would be tested until they had completed the decision making task. Recall performance was scored in terms of whether the direction of the base rates was correctly recalled (i.e., which population group mentioned in the problem was larger and which group was smaller).

2.2. Results and discussion

2.2.1. Decision making task

On each problem we coded whether the participant gave the correct answer (i.e., accuracy) and whether the participant referred to the base rate information during decision making (i.e., base rate mentioning). Fig. 1 present an overview of the mean performance on the different problem types.[4]

As in Tversky's and Kahneman's classic studies, accuracy on the incongruent problems was very low. Participants were clearly biased by the salient description and selected the correct response in fewer than 20% of the cases. As expected, participants had far less difficulties with the neutral and congruent problems where the description was simply neutral or consistent with the base rates. An ANOVA established that the difference in accuracy between the problem types was significant, $F(2,22) = 54.07$, $MSE = .04$, $p < .001$.

[4] Participants solved six items of each problem type. We calculated the average performance for each participant on each of the three problem types. These averages were subjected to ANOVAs. Reported percentages are always based on these averages calculated over participants unless otherwise noted.

The more crucial question, however, is to what extent people take analytic considerations into account when solving these problems and refer to the base rates during decision making. An ANOVA established that the frequency of base rates mentioning depended on the type of problem, $F(2,22) = 9.50$, $MSE = .06$, $p < .005$. As Fig. 1 shows, the verbal protocols indicate that on the majority of the neutral problems (54%) participants are considering the base rate information. However, once these same people are faced with a stereotypical description in the congruent and incongruent problems they seem to be completely discarding the base rates. On the crucial incongruent problems the base rates are mentioned only 18% of the time. People seem to be exclusively referring to the match between their response and the description without much evidence for a consciously experienced conflict.

Table 1 presents some interesting additional data. As Table 1 indicates, the few times that participants did mention the base rates on the incongruent problems ($n = 13$, out of a total number of 72 trials) they also tended to solve the problem correctly ($n = 11$ out of these 13 trials or 85% correct when base rates mentioned). The other way around, whenever participants did manage to give the correct response ($n = 14$) they typically also referred to the base rates ($n = 10$ out of these 14 trials or 71% base rates mentioned when correct). The same pattern was observed for the neutral problems. Indeed, participants' average accuracy and base rate mentioning correlated for the incongruent, $r = .92$, $p < .001$, and neutral problems, $r = .88$, $p < .001$. Not surprisingly, for the congruent problems where the description cues the correct response, accuracy did not depend on base rate mentioning, $r = .22$. In sum, whenever the classic incongruent problems were solved correctly, people successfully detected the conflict between the description and base rates. However, people erred on the vast majority of the problems and there was hardly any evidence for an explicitly experienced conflict in these cases. Indeed, on the 80% of the

Table 1
Overview of additional performance measures

Measure	Problem type					
	Incongruent		Congruent		Neutral	
	%	*n*	%	*n*	%	*n*
% Correct when base rates mentioned	85	11/13	92	11/12	100	39/39
% Base rates mentioned when correct	71	10/14	16	11/68	70	39/57
% Base rates mentioned when incorrect	5	3/58	25	1/4	0	0/17
r (base rate mentioning and accuracy)	.92*	12	.22	12	.88*	12
First problem						
% Correct	0	0/12	92	11/12	83	10/12
% Base rates mentioned	0	0/12	8	1/12	50	6/12
Overall						
% Correct	19	14/72	94	68/72	79	57/72
% Base rates mentioned	18	13/72	17	12/72	54	39/72

* $p < .001$.

incongruent problems that were solved incorrectly participants mentioned the base rates only 5% of the time. Consistent with Kahneman's claim about the lax nature of the conflict monitoring process, most of the time people do not seem to be aware that the base rates are relevant for solving the incongruent problems.

One reason for the lack of base rate mentioning in the present experiment might be the repetitive nature of the problem presentation. Participants had to solve a total of six incongruent problems and they might have stopped verbalizing their processing of the base rates after a while because they became less motivated or because they figured they had already sufficiently clarified their reasoning on the previous trials. Such confound would have decreased the average performance. We therefore examined the data for the first three presented problems separately. The first one of these was always an incongruent problem, the second one a congruent, and the third one a neutral problem. As Table 1 shows, the general pattern was present right from the start. Contrary to the motivation hypothesis, performance on the first, incongruent problem was even worse. None of the participants solved it correctly or mentioned the base rates.

2.2.2. Recall task

Fig. 2 presents an overview of the recall findings. The verbal protocols already indicated that participants were taking base rates into account when solving the neutral problems. Accuracy was high and participants mentioned the base rates on the majority of the trials. As Fig. 2 shows, the processing of the base rates during the neutral problem solving also resulted in a decent recall performance. Although participants did not know they had to memorize the base rates during decision making, on average, they correctly identified which group was the largest 66% of the time for the neutral problems. For the congruent trials, where the description cued the correct response and base rates were hardly explicitly considered, correct base rate recall reached only 36%. The crucial finding, however, concerns the incongruent problems. Although the verbal protocols showed no direct evidence for a consciously experienced conflict

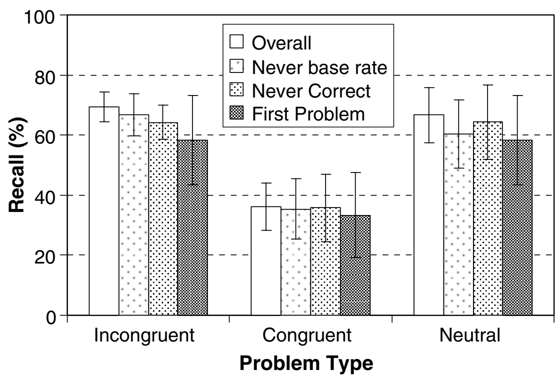

Fig. 2. Mean overall proportion of correct base rate recall. Recall performance for participants who never mentioned the base rates (Never base rate), always erred on the incongruent problems (Never correct), and the first presented problem of every type (First problem) are also presented. Error bars are standard errors.

and participants seemed to be almost completely neglecting the base rates, recall performance did indicate that the base rates had been processed. With an average performance of 69% correct identification recall was at par with the neutral problems and clearly superior to the recall for the congruent problems were there was no conflict to be detected. An ANOVA established that the recall performance significantly differed between the problem types, $F(2,22) = 9.26$, $MSE = .04$, $p < .001$.

The only difference between the congruent and incongruent problems was the conflicting nature of the description and base rates. If people would not be detecting the conflict and would simply neglect the base rate information on the incongruent problems, as the verbal reports suggested, recall performance for congruent and incongruent problems should not have differed.

Fig. 2 also shows the results of a number of additional control analyses. One could argue that the better recall on the incongruent problems might have been inflated because of the few trials where the base rates were explicitly mentioned. A purer measure of implicit conflict detection would concern the recall performance on those trials where the base rates were not explicitly mentioned. Fig. 2 presents the results of an extreme test of this hypothesis. Eight participants never mentioned the base rates on any of the incongruent problems they solved. As Fig. 2 shows, they nevertheless showed a similar recall pattern. Although they never mentioned the base rates on the incongruent problems, recall was still at par with the neutral problems and clearly superior to the congruent problems, $F(2,14) = 4.55$, $MSE = .05$, $p < .05$.

Similarly, one can look at accuracy and restrict the analysis to those participants who did not give a single correct response on any of the incongruent problems. This was the case for seven participants. As the recall findings in Fig. 2 show, even people who always erred showed the superior recall for incongruent problems. The recall effect still reached marginal significance, $F(2,12) = 3.39$, $MSE = .06$, $p < .07$, in this small group.

Finally, one could remark that the recall findings might have resulted from the repeated testing in the present experiment. The within-subject design might have made the conflict especially salient and cued a more profound conflict monitoring. To check this hypothesis we examined the recall data for the first three presented problems separately. The first one of these was always an incongruent problem. As Fig. 2 demonstrates, although correct recall for the first items tended to decrease somewhat the basic recall pattern was present right from the start. Base rates for the first, incongruent problem (58%) are still recalled almost twice as well as for the subsequently presented congruent problem (33%), $F(1,11) = 11.96$, $MSE = .23$, $p < .01$.

A final alternative explanation for the better base rate recall for incongruent and neutral problems vs. congruent problems might be the serial position of the presented problems. It is well established in memory studies that the first and last items on a list are better recalled than items that are presented closer to the middle (e.g., Glanzer & Cunitz, 1966). Although we used an unannounced recall procedure, the findings could have been affected if incongruent and neutral problems were presented more frequently in the beginning or at the end of the experiment. We therefore calculated the average distance of the 18 items from the middle position in the presentation order (i.e., the first problem received rank 1, the eighth and tenth problem rank 8 and so on). Incongruent and congruent problems had the same average distance (i.e., position 4.7) whereas the neutral items were actually presented somewhat closer to the middle (i.e., position 5.7). This shows that the presentation position factor cannot account for the recall pattern findings. Indeed, if the serial position would explain the better recall on the first (incongruent) over the second (congruent) problem, for example, recall on the thirdly presented neutral problem should have been even worse. As Fig. 2 shows, this was clearly not the case.

2.3. Conclusions

Experiment 1 showed that when people solve classic base rate problems there is hardly any evidence for an explicitly experienced conflict between problem solutions that are cued by the analytic and heuristic reasoning system. Only in 18% of the cases participants referred to the base rates and indicated they were taking analytic considerations in mind. However, the recall data showed that the base rates were not merely neglected. We might not be explicitly reporting an active struggle but our cognitive system does seem to be detecting the special status of the incongruent problems. Even when participants never mentioned the base rates and always erred on the incongruent problems they nevertheless managed to correctly identify which group was larger on the vast majority of the problems. For the congruent problems where the descriptions and base rates agreed this was not the case. If people were not detecting the conflict and were simply neglecting the base rate information on the incongruent problems, recall performance for congruent and incongruent problems should not have differed. In sum, while Experiment 1 showed that the anecdotal characterization of conflict detection as an actively experienced struggle is far from prototypical, there is evidence for Sloman and Epstein's basic idea about the efficiency of the conflict monitoring process. Even when we err our reasoning engine seems to be picking up that the description disagrees with the base rates. This suggests that the dominance of heuristic reasoning should not be attributed to a lack of conflict monitoring.

3. Experiment 2

In Experiment 2 the findings of Experiment 1 are further validated. Participants solved similar base rate problems but were no longer requested to think aloud. Experiment 2 focused on participants' problem processing time. While the analytic base rate scrutinizing associated with conflict detection might benefit subsequent recall, it will also take up some additional processing time. Reasoning latencies thereby provide an additional test of the opposing conflict monitoring views. One may assume that people will be fastest to solve the congruent items since the response can be fully based on mere heuristic reasoning without any further analytic intervention. Correctly solving the classic problems should be slowest since it requires people to detect the conflict and inhibit the heuristic response which are both conceived as time-demanding processes (e.g., De Neys, 2006a). The crucial question concerns the processing time of erroneously solved incongruent problems. If people simply fail to detect the conflict and reason purely heuristically, reasoning latencies for incorrectly solved incongruent and correctly solved congruent problems should not differ. If people detect the conflict, they should take longer to respond to the incongruent problems.

To validate the idea that upon conflict detection people spend specific time processing the base rates Experiment 2 also introduces a manipulation inspired by the "moving window" procedure (e.g., Just et al., 1982). The base rate information disappears from the screen once the description and question are presented. Participants have the option of visualizing the base rates afterwards. Such base rate reviewing can be used as a more specific test of the conflict detection claim. It is expected that when people detect that the description conflicts with the previously presented base rates they will spend extra attention to the base rates. With the present procedure the time spent visualizing the base rates can be used as a measure of this reviewing tendency. Longer overall response latencies after successful conflict detection on the incongruent items should thus be accompanied by a stronger tendency to visualize the base rates. If people simply neglect the base rates, there is also no reason to review and visualize them after the initial presentation.

Experiment 1 already showed that when the description was neutral and did not cue a response people were explicitly referring to the base rate information during decision making. Therefore, one can expect that people will also tend to review the base rates when they are faced with the neutral problems in the present experiment. The analytic base rate processing on the neutral problems should also result in somewhat longer decision making times compared to the congruent problems.

The crucial recall findings in Experiment 1 were based on a new task that was presented to a relatively small sample of participants. To validate the findings participants in Experiment 2 were also presented with the unannounced recall task after they finished the decision making task.

As a final control, some participants were simply asked to read the problems. These people saw the base rate information and description of the problems on the screen but the actual question to decide to which group the person most likely belonged was not presented. Hence, participants in the reading group were not engaged in any real decision making. People might visualize the base rates in the present experiment for a variety of reasons that have nothing to do with decision making. For example, people might have a basic tendency to go back to a visual stimulus whenever it disappears or they might simply enjoy playing around with the visualization key. The reading group should give us an idea of this baseline reviewing level. When simply reading, there is no reason to process the incongruent and congruent problems differently. Therefore, it is expected that the superior base rate recall and reviewing for the incongruent and neutral problems will only be observed in the decision making group.

3.1. Methods

3.1.1. Participants

A total of 86 students of the University of Leuven (Belgium), Department of Psychology, participated in return for credit in a psychology course.

3.1.2. Materials

3.1.2.1. Decision making task. Participants were presented with Flemish versions of the base rate problems. Problem content was slightly adapted for the Flemish test context (e.g., we used the well known low-end European retailer *Aldi* instead of *Wall-Mart*). As with the Canadian versions in Experiment 1 the stereotypical descriptions were validated in a pilot rating study. As in Experiment 1, participants solved a total of 18 problems (6 incongruent, 6 congruent, and 6 neutral ones) with slightly varied base rates (e.g., 995/5, 996/4, 997/3).

The lack of explicit base rate mentioning for the Flemish versions was also validated in a short thinking aloud study with 14 Flemish undergraduates (these people did not participate in Experiment 2). Participants solved four incongruent problems aloud. Results replicated the thinking aloud findings with the Canadian students in Experiment 1. Only 21% of the problems were solved correctly (i.e., $n = 12$ correct responses out of a total of 56 trials) and base rates were mentioned in fewer than 20% of the cases (i.e., $n = 11$ out of 56). When the problem was solved erroneously base rates were only mentioned 11% of the time (i.e., $n = 5$ out of 44). This established that, as in Experiment 1, participants were typically not explicitly referring to the base rates with the adapted material.

Experiment 2 was run on a computer. The problem was presented in two parts. First, the information about the sample composition and base rates was presented (i.e., italicized part in the example below). Participants were instructed to read this information and press the enter-key when they were ready. When the enter-key was pressed the remaining problem information (i.e., the underlined part in the example) was presented and the first part disappeared. Participants had the option of visualizing this first part with the crucial base rates afterwards by pressing the arrow-key. As long as they held down the arrow key, the first part remained visible. Once the arrow key was released, the information disappeared again. The second part with the description always remained visible after the initial presentation. The following is an example of the screen lay-out:

> *In a study 1000 people were tested. Among the participants there were 995 who buy their clothes at high-end retailers and 5 who buy their clothes at Wal-Mart. Karen is a randomly chosen participant of this study.*
>
> <u>Karen is a 33-year-old female. She works in a business office and drives a Porsche. She lives in a fancy penthouse with her boyfriend.</u>
>
> <u>What is most likely?</u>
>
> <u>a. Karen buys her clothes at high end retailers</u>
> <u>b. Karen buys her clothes at Wal-Mart</u>
>
> <u>Type down the letter reflecting your decision: __</u>

Participants started the experiment by reading the following general instructions:

> In this experiment you will have to solve a number of decision making problems. Each item will be presented in two parts. Once you've finished reading the first part you'll have to press the ENTER-key. The first part will disappear and the second part will be presented. If you want to, you can always review the first part of the item by holding the arrow-key (number key '8') down. If you release the key, the information will disappear again.
>
> You can take as much time as you want to think about the problem. Once you've made up your mind you must enter your answer ('a' or 'b') immediately and then the next problem will be presented.

Participants were given a congruent practice problem to familiarize themselves with the task format. Afterwards they received the same task specific instructions as in Experiment 1 and started the experiment.

To avoid any systematic primacy and recency bias on the recall measure, the 18 problems were always presented in a completely random order.[5]

Three latency measures were calculated. The time that elapsed between presentation of the first part of the problem and participants' ENTER-key pressing (that indicated they finished reading the information) will be referred to as *initial base rate reading time*. The total time between the enter-key press and the final response ('a' or 'b') entering will be referred to as *decision making time*. The specific amount of time a participant held down the arrow-key and visualized the base rates will be referred to as *base rate reviewing time*. The labels "reading" time and "decision" time and the precise splitting point are of course somewhat arbitrary. The rationale was that the crucial conflict in the decision making process can only start being detected once the second part with the description and answer-alternatives is presented.

3.1.2.2. Reading task. In the reading group participants were told that they were participating in a pilot study in which we wanted to determine the average time people needed to read some new material we were developing. Participants in the reading group received the same general instructions about the serial nature of the item presentation but all references to 'problem solving' or 'decision making' were avoided:

> In this pilot study you will have to read a number of items. Each item will be presented in two parts. Once you've finished reading the first part you'll have to press the ENTER-key. The first part will disappear and the second part will be presented. If you want to, you can always review the first part of the item by holding the arrow-key (number key '8') down. If you release the key, the information will disappear again.

[5] With hindsight, one downside to the random presentation was that it was hard to examine the impact of presentation order on the decision making performance. However, Experiment 1 already showed that the average effects did not differ from the pattern that was observed on the first problems.

> You don't have to rush, just read all the information in a natural pace. Once you've completely processed the information we ask you to press the 'a' key immediately and then the next problem will be presented.

With this goal in mind the question and response alternatives (e.g., 'What is most likely? (a) Karen buys her clothes at high end retailers. (b) Karen buys her clothes at Wall-Mart') of the problem were not presented. Thus, in the reading group the second part of the problem only contained the description and people were not encouraged to engage in any problem solving. The first part of the problems was completely identical in both groups. As in the decision making group, participants received a practice item so they could familiarize themselves with the reviewing procedure. The 18 items were also presented in random order.

3.1.2.3. Recall task. The recall task used the same format as in Experiment 1. Since Experiment 1 showed that recall was stable over the different items it was decided to restrict the recall test to four problems of each item type (e.g., in Experiment 1 recall of all 18 items was probed). We selected the 12 problems with the most diverse content. Despite the decent recall performance, a number of participants in Experiment 1 remarked that the task was quite lengthy and repetitive. It was hoped that the shorter and more diverse version would result in a more engaging task and possibly a more optimal measurement of the recall performance. The questions were printed one to a page in a booklet. Recall questions were presented in one of eight randomly determined orders. Except for the phrasing of the first sentence (e.g., 'One of the problems you just solved/read...') booklets for the reading and decision making group were completely similar. Of course, as in Experiment 1, recall was not announced before the reading or decision making task was completed.

3.1.3. Procedure

Participants were tested in small groups of 11 to 20 participants. Participants were randomly assigned to the decision making ($n = 44$) or reading group ($n = 42$). After completing the decision making or reading task, participants had a short break and then were presented with the recall task. Recall data of four participants was discarded because the booklet was not or not completely filled in.

3.2. Results

3.2.1. Decision making task accuracy

Participants' accuracy on the base rate problems very closely replicated the findings of Experiment 1. On average, only 22% of the incongruent problems were solved correctly but participants had far less trouble in selecting the correct response on the congruent (97%) and neutral (80%) problems, $F(2, 86) = 184.19$, $MSE = 1.33$, $p < .0001$. This establishes that the task adaptations did not fundamentally change the nature of the task.

3.2.2. Decision making task latencies

More crucial is the time participants needed to solve the different problems. For each participant we calculated the mean time needed to correctly solve the incongruent, congruent, and neutral problems. Latencies for erroneously solved incongruent problems were also entered in the analyses. As Fig. 3 shows, the decision making time for the four types of problems clearly differed, $F(3,57) = 7.98$, $MSE = 45.77$, $p < .001$. As expected, the congruent problems were solved fastest. People needed more time to solve the neutral problems where the heuristic system does not cue a response and correct responding requires analytic base rate reviewing. As one would predict people needed even more time to override the erroneous heuristic response and select the correct answer on the incongruent problems. The crucial finding is that even when an incongruent problem was solved incorrectly, people spent more time processing it than when solving the congruent problems. Newman–Keuls tests established that the decision making time of incorrectly solved incongruent problems fell in between that of correctly solved incongruent and congruent problems.

Fig. 3 also shows the initial base rate reading time (i.e., the time people initially spent reading the first part of the problem) for the four types of decisions. As one would expect, the latencies indicate that before the description is presented the base rate information is not processed any differently in the four cases, $F(3,57) < 1$.

Note that because of the within-subject nature of the decision making time analysis, the findings in Fig. 3 concern only those participants who solved at least one incongruent problem correctly ($n = 20$). However, about half of the participants erred on all the incongruent problems. One might argue that those people who at least sometimes manage to give the correct response are more cognitively gifted (e.g., Stanovich & West, 2000) and successful conflict detection in case of an error would only occur for this limited subgroup. Such a confound would restrict the

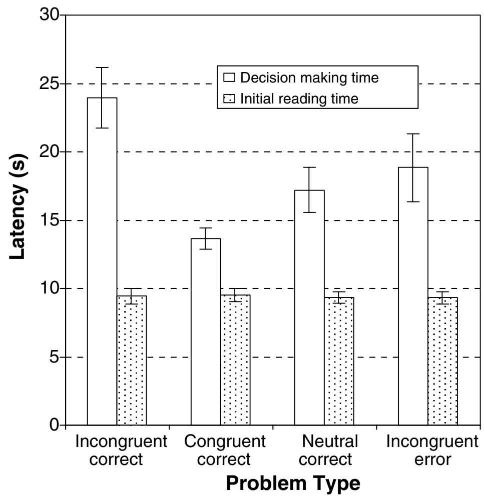

Fig. 3. Mean time (s) needed to make four crucial types of decisions. The time needed to read the preambles (initial reading time) in each of the four cases is also presented. Error bars are standard errors.

implication of the findings. We therefore compared the decision making latencies for two subgroups based on a median split of the accuracy on the incongruent problems. In the low score group ($n = 22$) participants solved all incongruent problems incorrectly. In the high score group ($n = 22$) participants solved at least one problem correctly (average accuracy was 44%). Fig. 4 presents the results. The incongruent latencies concern both correctly and incorrectly solved trials, congruent and neutral latencies concern correctly solved problems.

There was no main effect of score group, $F(1,41) = 1.22$, $MSE = 56.97$, but the factor did interact with problem type, $F(2,82) = 7.27$, $MSE = 12.56$, $p < .005$. As Fig. 4 shows, participants in the high score group took more time to solve the incongruent problems (i.e., more problems were solved correctly of course), $F(1,41) = 6.61$, $MSE = 29.66$, $p < .015$. The two groups' decision time on the congruent and neutral problems did not differ, $F(1,41) < 1$. The crucial finding was that even in the low score group the trend towards longer decision making latencies on the incongruent vs. congruent problems was readily clear, $F(1,41) = 10.65$, $MSE = 9.20$, $p < .01$. Thus, even those people who always err on the incongruent problems take more time to solve them compared to the congruent problems. These data underline the generality of the findings. Everyone seems to be spending more time to process the incongruent problems. As argued, the only difference between the incongruent and congruent problems is the presence of a conflict between the base rates and description on the incongruent problems. If people would simply neglect the base rates and fail to detect this conflict, decision making latencies should not differ.

3.2.3. *Base rate reviewing*

It was hypothesised that the longer decision making time on incongruent problems would be associated with a specific tendency to review the base rates in response to conflict detection. Half of the participants were simply instructed to read the problems and were not engaged in decision making. Base rate reviewing was expected to be less pronounced in this control group.

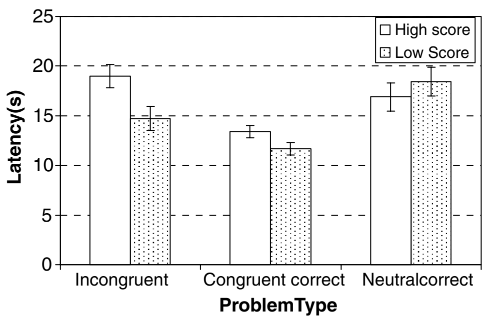

Fig. 4. Decision making time (s) as a function of the accuracy on the incongruent problems. The low score group are those people who failed to solve any of the incongruent problems correctly. Error bars are standard errors.

For every problem we coded whether or not a participant reviewed the first part of the problem with the crucial base rate information after the initial presentation. These averages were entered in a 2 (group, decision making or reading) × 3 (problem type) ANOVA. Fig. 5 shows the results. The main effects of group, $F(1,84) = 3.44$, $MSE = 7.44$, and problem type, $F(2,168) = 2.98$, $MSE = 1.15$, were not significant but, as expected, both factors interacted, $F(2,168) = 7.83$, $MSE = 1.15$, $p < .001$. As Fig. 5 indicates, people's base rate reviewing did not differ for the three problem types when merely reading, $F(2,82) = 1.56$, $MSE = 1.39$, but the effect did reach significance during decision making, $F(2,86) = 11.25$, $MSE = .93$, $p < .0001$. On the congruent problems the base rate reviewing frequency did not exceed the baseline level of the reading group, $F(1,84) < 1$. However, on the incongruent and neutral problems people reviewed significantly more during decision making than during reading, $F(1,84) = 8.62$, $MSE = 4.94$, $p < .005$. Thus, as expected, the base rates were specifically reviewed during decision making whenever the description was conflicting or simply uninformative.

In addition to the frequency of reviewing we also analysed the time people spent reviewing the base rates (i.e., how long the base rate information was visualized). As Fig. 5 illustrates, results were in line with the review frequency findings. Base rates were reviewed longer when solving incongruent and neutral problems than when solving congruent problems, $F(2,86) = 9.91$, $MSE = 31.7$, $p < .001$, but review time did not differ when the problems were merely read, $F(2,82) = 1.19$, $MSE = 12.11$ (Problem Type × Group interaction, $F(2,168) = 9.3$, $MSE = 22.14$, $p < .005$). The main effects of Problem Type, $F(1,168) = 5.23$, $MSE = 22.14$, $p < .01$, and Group, $F(1,84) = 7.42$, $MSE = 126.6$, $p < .01$, were also significant in the review time analysis. As in the frequency analysis, the longer review time on the incongruent and neutral problems during decision making exceeded the base line level of the reading group, $F(1,84) = 10.99$, $MSE = 120.51$, $p < .005$, whereas review time on the congruent problems did not differ during decision making or mere reading, $F(1,84) < 1$.

For the above comparisons of the reviewing tendencies in the reading and decision making groups the data was analysed independent of whether participants had solved the decision making problem correctly or incorrectly. We also wanted

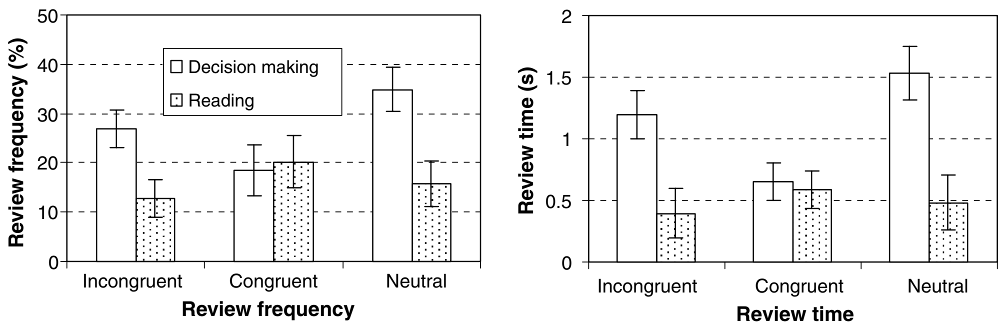

Fig. 5. Mean proportion of base rate reviewing and base rate reviewing time (s) for the three different problem types. Error bars are standard errors.

to verify whether the review results during decision making differed in terms of the accuracy on the incongruent problems. As with the latency findings, we compared the performance of the group of people who erred on all incongruent problems with the group who solved at least one problem correctly. As Fig. 6 shows, the high score group showed overall a more pronounced base rate reviewing than the low score group. People who always erred reviewed less frequently, $F(1,42) = 5.05$, $MSE = 6.93$, $p < .05$, and not as long as the higher scoring group, although the difference in terms of review time did not reach significance, $F(1,42) = 2.46$, $MSE = 169.38$. Hence, overall the tendency to review the base rates was linked to a better reasoning performance. The main effect of problem type was also significant, both in terms of review frequency, $F(2,84) = 11.09$, $MSE = .94$, $p < .0001$, and review time, $F(2,84) = 9.94$, $MSE = 31.6$, $p < .005$. The crucial finding is that the two factors did not interact, neither in terms of review frequency, $F(2,84) < 1$, nor review time, $F(2,84) = 1.15$, $MSE = 31.6$. As Fig. 6 shows, the trend towards more and longer base rate reviewing on the incongruent and neutral problems was clear in both capacity groups. Even though less gifted reasoners may be generally less inclined to review the base rates, they still review more on incongruent than on congruent problems.

3.2.4. Recall task

As with the review data, we first compared the recall performance in the decision making and reading group. Fig. 7 shows the results. As expected, the recall pattern for the three problem types tended to differ in both groups, $F(2,160) = 2.85$, $MSE = .05$, $p < .07$. There was also a main effect of problem type, $F(2,160) = 3.67$, $MSE = .05$, $p < .05$, whereas the main effect of task group was not significant, $F(1,80) < 1$. In the decision making group the findings of Experiment 1 were replicated. Base rates of incongruent and neutral problems were better recalled than the base rates of congruent problems, $F(2,82) = 6.80$, $MSE = .05$, $p < .005$. In the reading only group base rate recall did not differ on the three problems, $F(2,78) < 1$. Simple effect tests established that recall on the congruent problems was not better after simple reading than after decision making,

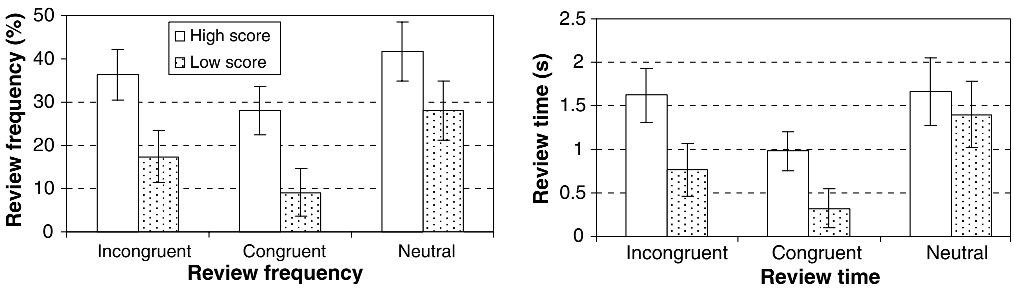

Fig. 6. Mean proportion of base rate reviewing and review time (s) for the three different problem types as a function of performance on the incongruent problems. The low score group are those people who erred on all incongruent problems. Error bars are standard errors.

$F(1,80) = 2.49$, $MSE = .05$. The increased recall performance for incongruent and neutral problems in the decision making group did exceed the performance of the reading group, $F(1,80) = 4.67$, $MSE = .05$, $p < .05$.

We also verified whether the recall results during decision making differed in terms of the accuracy on the incongruent problems. As with the review findings, we compared the performance of the high and low scoring group. Results are presented in Fig. 8. Overall, recall performance of people in the high score group was better than that of the people who always solved the incongruent problems incorrectly, $F(1,40) = 24.41$, $MSE = .05$, $p < .0001$. There was also a main effect of problem type, $F(2,80) = 6.97$, $MSE = .05$, $p < .005$, but as in the base rate review analysis, both factors did not interact, $F(2,80) = 2.07$, $MSE = .05$. The two score groups showed the same basic recall trend. Even people who solved all incongruent problems incorrectly managed to correctly recall the direction of the base rates on more than 75% of the incongruent trials. As Fig. 8 suggests, if anything the superior recall on incongruent and neutral vs. congruent problems even tended to be somewhat more pronounced for the low score group.

3.3. Conclusions

Experiment 2 validated and extended the findings of Experiment 1. People showed a superior recall when the description of the problems conflicted with the base rates or was simply neutral. The better recall was accompanied by longer decision making times and a specific tendency to review the base rate information. Since the description does not cue a response on the neutral problems it is not very surprising that participants go back to the base rates after reading the uninformative description and spend additional time reviewing them. However, on the incongruent and congruent problems the description does cue a stereotypical response. The only difference between the two problems is that on the incongruent problems this stereotypical response disagrees with the base rates whereas there is no such conflict on the congruent problems. The present data suggest that participants detect the conflict on the incongruent problems and consequently redirect attention to an additional processing of the base rate information. This base rate reviewing is

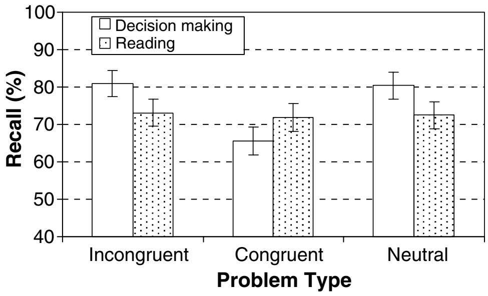

Fig. 7. Mean proportion correct base rate recall after decision making and mere reading. Error bars are standard errors.

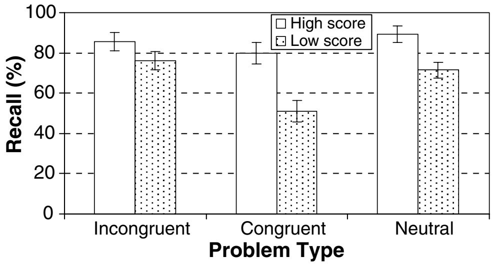

Fig. 8. Mean proportion correct base rate recall as of function of the accuracy on the incongruent problems. Error bars are standard errors.

resulting in longer decision making times and better memorization of the base rate information.

The evidence for conflict detection was evident whether the incongruent problem was solved correctly or not. Even people who erred on every single incongruent problem needed more time to solve them and showed more extensive reviewing and better recall on the incongruent than on the non-conflicting congruent problems. Thus, even the accuracy-wise most ungifted reasoners were detecting the special status of the incongruent problem. Although this did not suffice to override the response cued by the tempting stereotypical description, it does show that the dominance of heuristic responding during decision making should not be attributed to a lack of conflict detection. Indeed, the present data clearly suggest that successful conflict detection is omnipresent during decision making.

Of course, the evidence for successful conflict detection only concerned people's performance during decision making. When participants were merely reading the material, incongruent and congruent problems were not processed any differently. Latencies, review tendencies, and recall were completely similar. Thus, during reading people did not care about the special status of the incongruent problems. This points to the goal-directed nature of the conflict monitoring process and analytic system intervention. People do not spoil resources monitoring for a possible conflict between different problem solutions when they are not engaged in decision making. Analytic system intervention after conflict detection will only be recruited when we intend to make a decision.

4. General discussion

The present study contrasted opposite views on conflict monitoring in dual process theories of reasoning and decision making. According to Kahneman and colleagues (e.g., Kahneman, 2002; Kahneman & Frederick, 2005) and the classic work of Evans (1984) conflict monitoring is typically quite lax. It is assumed that most of the time people rely exclusively on the heuristic route while making decisions without taking analytic considerations into account. In this view, people are typically biased during decision making because they fail to detect a conflict. Authors like

Epstein (1994) and Sloman (1996) on the other hand, claim that the heuristic and analytic route are simultaneously activated and people experience a struggle whenever the two systems cue different responses. In this view, people always take analytic considerations in mind and detect that they conflict with the heuristically cued belief. Hence, according to these authors there is nothing wrong with the conflict monitoring during decision making.

For the development of the dual process framework it is crucial to determine which conflict detection view is correct. The present study pointed towards some clear conclusions. People's verbal reports in Experiment 1 indicated that they were not experiencing a conflict between the description and base rate information. Whenever there was a stereotypical description available the base rate information was hardly ever explicitly referred to. However, Experiment 1 also showed that even when participants never mentioned the base rates and always erred on the incongruent problems they nevertheless managed to correctly identify which group was the largest on the vast majority of the problems. For the congruent problems where the descriptions and base rates agreed this was not the case. Experiment 2 replicated the recall findings and showed that the better recall for erroneously solved incongruent problems was associated with longer decision making times and more extensive reviewing of the base rate information. Taken together results indicate that whereas the popular characterisation of conflict detection as an actively experienced struggle can be questioned there is nevertheless evidence for Sloman's and Epstein's basic idea about the flawless operation of the conflict monitoring process. The differential processing of the congruent and incongruent problems supports the claim that whenever the base rates and description disagree people will detect this conflict and consequently redirect attention towards a deeper processing of the base rate information.

The nature of conflict monitoring has interesting implications for the way reasoning errors and the interaction between the two reasoning systems are characterized. These and related implications of the present findings are elaborated on in the following sections. We start by commenting on the status of the conflict detection experience.

4.1. Implicit vs. explicit detection?

Given the present findings one may wonder to what extent people have conscious access to the conflict that is being detected. We labeled the detection experience as implicit to contrast it with the verbal protocol findings. The traditional measure of explicit awareness, peoples' verbalizations, did not show any evidence for an actively experienced conflict. However, the more indirect measures that did not rely on explicit verbalization consistently indicated that the conflict had been detected. Hence, our data show that people are not verbalizing the conflict they are detecting. This is interesting because it indicates that the anecdotal sketch of the detection process as a dramatic struggle where people report to be torn between two alternatives is far from prototypical. However, as we noted in the introduction it should be clear that the lack of verbalization does not necessarily imply that the detection process is unconscious. Whether or not verbalization needs to be considered as the key prerequisite for conscious, explicit processing is the focus of a long standing and open debate (e.g., Moors & de Houwer, 2006). Depending on one's position in the debate one will put a different label on the detection experience. In our view, such a label discussion is not very informative. What matters is that the present findings clarify that people are detecting and processing the conflict between analytic and heuristic problem solutions whatever the exact level of conscious "conflict feeling" it may precisely involve. Moors and de Houwer already advised cognitive scientists to refrain from dichotomous implicit–explicit claims and favored a more gradual approach. With this in mind one could argue that a lack of verbalization suggests that the conflict experience might be less explicit than traditionally assumed but any stronger claims should be avoided.

4.2. *Conflict monitoring and heuristic errors*

The evidence for successful conflict monitoring was clear even when the incongruent problems were solved erroneously. Hence, the dominance of heuristic responses should not be attributed to a lax monitoring process. This implies that errors are not arising because a reasoner has simply not acquired the relevant normative principles, fails to retrieve them, or considers the principles irrelevant. If people were not taking analytic considerations (e.g., the role of group size) into account, the base rates would not be attended to and people would simply not detect that there is a conflict. The elimination of these claims lends credence to the alternative explanation that the dominance of heuristic responses should be attributed to an inhibition failure. People will not always manage to discard the compelling heuristics. This finding is consistent with recent claims about the role of inhibitory processing capacity in reasoning (e.g., De Neys, Schaeken, & d'Ydewalle, 2005; Handley, Capon, Beveridge, Dennis, & Evans, 2004; Markovits & Doyon, 2004; Simoneau & Markovits, 2003). Furthermore, it can help explaining why it has sometimes been observed that extensive tutoring in logic and probability theory has only a small impact on people's performance. Indeed, even expert populations of ace mathematicians and statisticians have been shown to fail to solve straightforward classic reasoning tasks (e.g., Burns & Wieth, 2004; Kahneman, Slovic, & Tversky, 1982). This seems hard to grasp and has been interpreted as a severe blow to the rationality of the human species.

Interpreted in the light of the present findings the counter-intuitive results concerning the impact of tutoring are making good sense, however. Our data show that untrained participants are already taking base rates into account and detect the conflict with the heuristically cued response. Thus, people know all too well that the base rate information is relevant when making a decision. Hence, it is not surprising that additional clarifications of the role of base rates in tutoring sessions will not sort a lot of effect. People's problem is not a lack of statistical sophistication. What they seem to struggle with is overriding the tempting heuristics. One can find some interesting support for this view in the work of Houdé and Moutier (1996). Houdé and Moutier asked people to solve the Wason selection task, a classic deductive reasoning task where intuitive, heuristic responses conflict with the logically correct response.

One group of participants received an extensive logical training between the pre-test and post-test. A second group did not receive any logical training but received a practical inhibition training that strengthened their ability to discard intuitively cued responses. Consistent with the above claim, Houdé and Moutier observed that the inhibition training resulted in a spectacular performance boost whereas the reasoning performance did not improve after the logical training (see Moutier & Houdé, 2003, for similar findings with the conjunction fallacy task). This pattern is precisely what one would expect if people's problem is a lack of inhibitory capacity rather than a lack of conflict detection.

4.3. Parallel or serial interaction?

We noted that the different views on the efficiency of the conflict monitoring process in the literature are related to somewhat different conceptualisations of the interaction between the analytic and heuristic system. As Evans (2007a) pointed out, Sloman (1996) and Epstein (1994) are proposing a more parallel interaction where both routes are supposed to be simultaneously computing a problem solution from the start. In Kahneman and Frederick's (2002) framework and Evans' (1984) own model one can find a more serial characterisation where a reasoner initially starts with heuristic reasoning and the analytic system only intervenes in a later stage. Evans (2007a) has labelled these parallel and default-interventionist models, respectively. It should be clear that these are only general labels. At present all theories lack a clear processing specification and it is not clear how extreme the parallel and serial operation is conceived. The present conflict monitoring data can be especially helpful to clarify the nature of the interaction. In this section we will first argue that the suggestion of a purely serial or parallel mechanism cannot be maintained and propose an alternative that centres around the idea of a *shallow analytic monitoring* process.

The simultaneous heuristic and analytic processing idea in a parallel model naturally captures the finding that people are ace conflict detectors. If people always engage analytic processing together with the heuristic activations, it makes sense that they will face little difficulties noticing that the two systems cue different responses. The parallel processing architecture would be pretty advantageous in those cases where the heuristic route cues a conflicting response. However, a fully parallel model is quite disadvantageous when both routes cue the same response. Indeed, the serial model where people reason purely heuristically at the start presents a major computational advantage here. In those cases where the heuristic system cues a correct response the serial system will take advantage of the fast and undemanding heuristic route. In the parallel model the analytic route is blindly engaged right from the start. People always work through the time-consuming and demanding analytic computations. The parallel model thus throws away the benefits of the heuristic route. When there is an easy and correct heuristic problem solution at hand, it is redundant to complete the demanding analytic operations.

A purely serial model, however, is problematic when the heuristic system cues a conflicting response. The default-interventionist serial model states that the analytic system will be engaged in case of conflict detection. However, one can only detect a conflict if one is at least taking some minimal analytic considerations into account. Indeed, successful conflict detection requires that one monitors for a conflict and monitoring for a conflict requires that one knows what to monitor and look for. There has to be some minimal analytic operation right from the start otherwise it is not possible to determine whether the heuristic response can be sanctioned and further analytic scrutinizing is required. For example, in the case of making decisions about base rate problems people have to be at least aware that group size is relevant for the solution of the problem and therefore needs to be attended to. Otherwise, our reasoning engine would simply not be able to characterize the response triggered by the description as conflicting. By definition, detecting a conflict requires that one compares at least two different pieces of information. If one is only accessing one route there would simply never arise a conflict. The default-interventionist idea that analytic thinking only kicks in once a conflict is detected is tempting but begs the question of how that conflict was detected in the first place.

In sum, postulating a purely parallel or serial reasoning architecture does not work for dual process theories. On one hand, a purely serial dual process model is paradoxical. If one wants to avoid relying on a little conflict detecting homunculus the heuristic route needs to be monitored and this requires some minimal analytic thinking. A purely parallel model on the other hand violates the principle of cognitive economy. People would always work through the demanding analytic computations even when the undemanding heuristic route cues exactly the same decision. In this view, the heuristic route would be nothing but an evolutionary artefact that has no longer any purpose and only serves to bias our thinking. Such a dual process view would present a very bleak picture of the human reasoning engine in which the power of heuristic thinking is completely neglected. Moreover, any fully parallel model would not be able to account for the present data. In a fully parallel model the presence or absence of a conflict would not affect the actual base rate processing. People are supposed to complete the analytic process in all circumstances. Whether or not the two responses agree or disagree should not affect the actual analytic processing. Hence, people should spend the same amount of attention processing the base rates on all problems. The present findings clearly showed that people process the base rates differently on congruent and incongruent problems. Moreover, the differential processing did not start right away. When people initially read the base rate information, reading times for the different problem types did not differ. Base rate were re-evaluated once the description had been processed. This indicates that initially, before the conflict was detected, the base rate information was not yet fully processed. Indeed, even on neutral problems, where the description did not cue a response, reasoners did not tend to give the correct response right away but also needed to go back to the sample information for some additional scrutinizing. This suggests a two-stage analytic reasoning process. Initially, the base rate information needs to be processed and maintained in working memory. This allows the system to compare the base rates with the heuristically cued response. When the description conflicts with the stored base rate information or when the description does not cue a decision, additional, deeper analytic processing will be recruited.

We noted that dual process theorists are not very explicit about the exact nature of the architecture they propose. Evans (2007a) rightly stressed that the writings of different authors point towards both parallel and serial conceptualizations. However, the question is how extreme the parallel and serial claims need to be interpreted. We argued above that a purely serial or parallel model does not seem to be making sense. Here we suggest a less extreme, alternative view that hinges on the idea that decision making is characterized by a *shallow analytic monitoring* process. Rather than being fully parallel or serial this is a hybrid two-stage model that captures the basic ideas behind the more extreme models but avoids the conceptual pitfalls. On one hand, it shares the idea with the parallel model that all heuristic thinking is always accompanied by a simultaneous analytic monitoring process. On the other hand, it shares the idea with the serial model that this monitoring is not full-fletched analytic thinking. The initial monitoring would be shallow in the sense that it only recruits and keeps activated some general analytic principles while taking up but a minimal amount of cognitive resources. The shallow analytic monitoring allows the reasoner to determine whether or not the heuristically cued response can be sanctioned but does not suffice to make a decision in case of a conflict. This will require additional analytic processing where the analytic and heuristic responses are further weighted against each other. Hence, the crucial difference with the parallel models is that the analytic process is not blindly engaged. People will not continue computing an analytic response when the heuristic response is not labelled as conflicting during the initial monitoring.

It will be clear that the postulation of a two-stage analytic reasoning process consisting of an initial shallow monitoring and optional deeper processing stage will need to be further tested. Interestingly, the basic idea does seem to be getting some support from findings in related fields. For example, Ball, Phillips, Wade, and Quayle (2006) analysed eye-movements when reasoners were solving deductive syllogisms. In these problems the logical validity of an argument structure will sometimes conflict with the believability of its conclusion (e.g., a valid syllogism with an unbelievable conclusion, for example, ‘All mammals can walk. Whales are mammals. Thus, Whales can walk’). As in the classic base rate problems, solving such problems calls for an analytic intervention. Although Ball et al. where addressing a different question, their data does indicate that these syllogistic conflict problems were longer inspected than similar problems were believability and the logical status of the problem were in line. Moreover, as in the present study, initial inspection times of the incongruent and congruent syllogisms did not differ. It was only after participants encountered a conflicting conclusion that they went back to the premises for additional scrutinizing. Such observations fit well with the suggestion of a two-stage analytic reasoning process and point to the possible generality of the present findings.

Lastly, with respect to the further refinement of the present framework we want to signal the relevance of the large body of work in the cognitive neuroscience literature on conflict monitoring and cognitive control (e.g., Botvinick, Braver, Barch, Carter, & Cohen, 2001; Botvinick, Cohen, & Carter, 2004; MacDonald, Cohen, Stenger, & Carter, 2000; Ridderinkhof, Ullsperger, Crone, & Nieuwenhuis, 2004). These studies suggest that the detection of conflict is among the functions of a specific brain region of the human frontal lobe, the anterior cingulate cortex (ACC). It is assumed that this conflict signal triggers activation in more lateral frontal regions (LPFC), resulting in adjustments in cognitive control. One function of the LPFC would be to inhibit one of the conflicting responses so that the conflict is resolved and the ACC activation will decrease. For example, the ACC typically responds to tasks such as the Stroop (e.g., naming the ink color when the word WHITE is written in black ink) that involve a conflict in the form of competition between the correct response and the one that needs to be overridden. While the ACC signals the detection, correct responding and actually overriding the erroneous, prepotent response has been shown to depend on the LPFC recruitment (e.g., MacDonald et al., 2000).

Linking this general research on cognitive control might be especially fruitful to further examine the conflict monitoring process during decision making. One suggestion that might help to clarify the nature of reasoning errors would be to examine the ACC and LPFC activations during base rate problem solving. Correctly solving the base rate problems requires that the conflict between the two reasoning systems is detected and the heuristic response inhibited. Based on the cognitive control findings we could thus predict to see both ACC and LPFC activation when incongruent trials are solved correctly. For erroneously solved problems we should not see LPFC activation since the heuristic response was not successfully inhibited. The crucial question concerns the activation of the ACC when people err on the incongruent problems. If we assume that the ACC indeed plays the role of conflict detector the present data suggest that we would also find ACC activation for the erroneously solved problems. If the Kahneman and Evans view about the lax nature of the conflict monitoring is right, people will not detect a conflict, and we would not expect to see ACC activation. Such predictions remain speculative of course but they demonstrate the potential of binding the two fields more closely together.

4.4. Implications for the rationality debate

The evidence for the efficiency of the conflict monitoring during decision making has some important implications for the debate on human rationality (e.g., Stanovich & West, 2000; Stein, 1996). This rife debate centres around the question whether the traditional norms (such as standard logic and probability theory) against which the rationality of people's decisions are measured are valid. It has been questioned for example why preferring base rates over beliefs would be more rational or "correct" than pure belief-based reasoning (e.g., Oaksford & Chater, 1998; Todd & Gigerenzer, 2000). One reason for criticizing the norm has been the consistent very low number of correct responses that has been observed on the classic reasoning and decision making tasks. If over 80% of well-educated, young adults fail to solve a simple decision making task, this might indicate that there is something wrong with the task scoring norm rather than with the participants. However, the debate, as the vast majority of dual process research, has often been characterized by an exclusive focus on people's response output (i.e., whether or not people manage to give the correct response) and not on the underlying cognitive processes (De Neys, 2006b; Gigerenzer et al., 1988; Hertwig & Gigerenzer, 1999; Hoffrage, 2000; Reyna et al., 2003). The present data clarify that giving an erroneous belief-based response does not imply mere belief-based reasoning where people completely disregard the traditional norm. Results indicate that even people who consistently err detect the conflict between base rates and the description and allocate additional resources to a deeper base rate processing. If people did not believe that the group size information matters during problem solving, they would not waste time processing it. People might not always manage to adhere to the norm but they 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. Clearly, people are more normative than their answers show.

Interestingly, past studies pointing to the pervasive impact of Heuristics and biases (e.g., Tversky & Kahneman, 1974) have progressively deemphasized the importance of normative standards in human thinking. Researchers became increasingly convinced that reasoning was in essence a purely automatic, heuristic process with little or no role for traditional standards of rationality (for a review see Evans, 2002). One could say that the present work helps the pendulum swing back in the other direction. The evidence for successful conflict monitoring reverses the claim and suggests that there is actually no such thing as pure heuristic thinking.[6] At least in case of the classic base rate neglect phenomenon, heuristic thinking seems to be always accompanied by successful analytic monitoring.

4.5. Caveats and conclusions

The present findings concern a sample of highly educated participants (i.e., university students) who were asked to reason in a quite formal setting (i.e., sitting behind a computer or next to an experimenter while participating in an experiment in return for course credit). As always, it cannot be excluded that in the population at large or in more daily life settings conflict monitoring might be far less successful and decision making nothing more than an automatic, heuristic process. Nevertheless, it is this same group of young, educated adults whose reasoning performance has been the subject of dual process theorizing and the rationality debate. The specific decision making task we selected is also one of the most intensely studied tasks in the field and the very same one that inspired Kahneman's view about the lax nature of the monitoring process (e.g., Kahneman, 2002). Hence, one cannot argue that the present sample and task selection would not be justified to validate the claims. Of course, it will still be necessary to extend the present approach to other decision

[6] In a way, dual process theorists have always acknowledged the idea that heuristic thinking is accompanied by some analytic processing. However, the analytic processing in this sense typically refers to some controlled aspect of the task that is not directly related to the reasoning process. Kahneman (2002) and Evans (2007b), for example, have stated that when people give a heuristic response they will also need to read the problem, construct a mental representation of it, select one of the possible responses and write it down. Indeed, even a heuristically cued response will need to be overtly expressed and this expression itself might require some controlled or analytic processing. The point here is that in Kahneman's view the origin of the response is still considered to be cued purely heuristically without deliberate reasoning.

making and reasoning tasks. Procedures such as the moving window and recall manipulations that were introduced in the present paper might be adjusted to work with other paradigms and could prove to be very useful in this respect.

With these stipulations in mind the present study did allow to conclude that the conflict monitoring process is far from lax. People typically detect the conflict between salient heuristic beliefs and analytic knowledge such as sample size considerations. With respect to the opening example this suggests that while people might not be able to resist the urge to blame small but visible minority groups, they at least seem to notice that their judgement is not fully justified. Although this does not pardon the unfounded judgment it does hold some promise. People are no pure heuristic thinkers who are not sensitive to normative considerations. In general, we seem to be less ignorant about the implications of our judgements than the actual judgements show.

Acknowledgments

Wim De Neys is a Post Doctoral Fellow of the Flemish Fund for Scientific Research (Post doctoraal Onderzoeker FWO – Vlaanderen). Experiment 1 was conducted during a stay at Vinod Goel's lab at York University, Toronto (Canada).

Appendix A

The 18 problems used in Experiment 1.

A.1. Incongruent problems

(a) In a study 1000 people were tested. Among the participants there were 4 men and 996 women. Jo is a randomly chosen participant of this study.

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

(b) In a study 1000 people were tested. Among the participants there were 5 engineers and 995 lawyers. Jack is a randomly chosen participant of this study.

Jack is 36 years old. He is not married and is somewhat introverted. He likes to spend his free time reading science fiction and writing computer programs.

What is most likely?

a. Jack is an engineer
b. Jack is a lawyer

(c) In a study 1000 people were tested. Among the participants there were three who live in a condo and 997 who live in a farmhouse. Kurt is a randomly chosen participant of this study.

Kurt works on Wall Street and is single. He works long hours and wears Armani suits to work. He likes wearing shades.

What is most likely?

a. Kurt lives in a condo
b. Kurt lives in a farmhouse

(d) In a study 1000 people were tested. Among the participants there were 997 nurses and 3 doctors. Paul is a randomly chosen participant of this study.

Paul is 34 years old. He lives in a beautiful home in a posh suburb. He is well spoken and very interested in politics. He invests a lot of time in his career.

What is most likely?

a. Paul is a nurse
b. Paul is a doctor

(e) In a study 1000 people were tested. Among the participants there were four whose favorite series is Star Trek and 996 whose favorite series is Days of Our Lives. Jeremy is a randomly chosen participant of this study.

Jeremy is 26 and is doing graduate studies in physics. He stays at home most of the time and likes to play video-games.

What is most likely?

a. Jeremey's favorite series is Star Trek
b. Jeremey's favorite series is Days of Our Lives

(f) In a study 1000 people were tested. Among the participants there were 5 sixteen-year olds and 995 fifty-year olds. Ellen is a randomly chosen participant of this study.

Ellen likes to listen to hip hop and rap music. She enjoys wearing tight shirts and jeans. She's fond of dancing and has a small nose piercing.

What is most likely?

a. Ellen is sixteen
b. Ellen is fifty

A.2. Congruent problems

(a) In a study 1000 people were tested. Among the participants there were 995 who buy their clothes at high-end retailers and five who buy their clothes at Wal-Mart. Karen is a randomly chosen participant of this study.

Karen is a 33-year-old female. She works in a business office and drives a Porsche. She lives in a fancy penthouse with her boyfriend.

What is most likely?

a. Karen buys her clothes at high end retailers
b. Karen buys her clothes at Wal-Mart

(b) In a study 1000 people were tested. Among the participants there were 997 girls and 3 boys. Erin is a randomly chosen participant of this study.

Erin is 13 years old. Erin's favourite subject is art. Erin's favourite things to do are shopping and having sleepovers with friends to gossip about other kids at school.

What is most likely?

a. Erin is a girl
b. Erin is a boy

(c) In a study 1000 people were tested. Among the participants there were 997 who have a tattoo and three without tattoo. Jay is a randomly chosen participant of this study.

Jay is a 29-year-old male. He has served a short time in prison. He has been living on his own for 2 years now. He has an older car and listens to punk music.

What is most likely?

a. Jay has a tattoo
b. Jay has no tattoo

(d) In a study 1000 people were tested. Among the participants there were 996 kindergarten teachers and 4 executive managers. Lilly is a randomly chosen participant of this study.

Lilly is 37 years old. She is married and has 3 kids. Her husband is a veterinarian. She is committed to her family and always watches the daily cartoon shows with her kids.

What is most likely?

a. Lilly is a kindergarten teacher
b. Lilly is an executive manager

(e) In a study 1000 people were tested. Among the participants there were 4 Bruce Springsteen fans and 996 Britney Spears fans. Tara is a randomly chosen participant of this study.

Tara is 15. She loves to go shopping at the mall and to talk with her friends about their crushes at school.

What is most likely?

a. Tara is a Bruce Springsteen fan
b. Tara is a Britney Spears fan

(f) In a study 1000 people were tested. Among the participants there were 5 Americans and 995 French people. Martine is a randomly chosen participant of this study.

Martine is 26 years old. She is bilingual and reads a lot in her spare time. She is a very fashionable dresser and a great cook.

What is most likely?

a. Martine is American
b. Martine is French

A.3. Neutral problems

(a) In a study 1000 people were tested. Among the participants there were five who campaigned for George W. Bush and 995 who campaigned for John Kerry. Jim is a randomly chosen participant of this study.

Jim is 5 ft and 8 in. tall, has black hair, and is the father of two young girls. He drives a yellow van that is completely covered with posters.

What is most likely?

a. Jim campaigned for George W. Bush
b. Jim campaigned for John Kerry

(b) In a study 1000 people were tested. Among the participants there were 996 men and 4 women. Casey is a randomly chosen participant of this study.

Casey is a 36-year-old writer. Casey has two brothers and one sister. Casey likes running and watching a good movie.

What is most likely?

a. Casey is a man
b. Casey is a woman

(c) In a study 1000 people were tested. Among the participants there were 997 who play the drums and three who play the saxophone. Tom is a randomly chosen participant of this study.

Tom is 20 years old. He is studying in Washington and has no steady girlfriend. He just bought a second-hand car with his savings.

What is most likely?

a. Tom plays the drums
b. Tom plays the saxophone

(d) In a study 1000 people were tested. Among the participants there were 997 pool players and 3 basketball players. Jason is a randomly chosen participant of this study.

Jason is 29 years old and has lived his whole live in New York. He has green colored eyes and black hair. He drives a light-gray colored car.

What is most likely?

a. Jason is a pool player
b. Jason is a basketball player

(e) In a study 1000 people were tested. Among the participants there were four who live in New York and 996 who live in Los Angeles. Christopher is a randomly chosen participant of this study.

Christopher is 28 years old. He has a girlfriend and shares an apartment with a friend. He likes watching basketball.

What is most likely?

a. Christopher lives in New York
b. Christopher lives in Los Angeles

(f) In a study 1000 people were tested. Among the participants there were 5 computer science majors and 995 English majors. Matt is a randomly chosen participant of this study

Matt is 20 years old and lives in downtown Toronto. Matt's favourite food is pasta with meatballs. His parents are living in Vancouver.

What is most likely?

a. Matt is a Computer Science major
b. Matt is an English major

Appendix B

The verbal protocols of the 12 participants in Experiment 1.

B.1. Incongruent problems

B.1.1. Incongruent problem (a)

1. So I would assume that Jo is a man just because. . .I don't know when I think of engineering sometimes I think of men more quickly. Also he goes out cruising with friends while listening to loud music, which can really be for both man or a woman but I automatically think of a man, I am not really sure why.

2. I am gonna guess Jo is a man because he likes to go out cruising with friends and drink beer.
3. (a) Jo is a man because he drinks beer, he studies engineering and...that's it.
4. Jo is a man because he likes to go cruising with friends and drinks beer, and that is a characteristic...and a stereotype of men.
5. Man, he likes to drink beer, loud music and what else, driving drunk yeah.
6. Ok so *996 women only 4 men*...but his name is Jo he has to be a man, I don't know anymore.
7. He listens to loud music and drinks beer so he is a man.
8. Jo is a man because it says that he likes to drink beer and cars and loud music, so he is a guy.
9. Jo is a man because...he is an engineer and that sounds more like a man and because he...he likes cruising with friends, listening to loud music and drinks beer.
10. Ok so Jo is a man
11. ...Well...I think Jo is a man because...he likes drinking beer and cruising with friends...and that is like the typical stuff that guys do.
12. Ok...well he likes to cruise on Friday nights....so I would say he is a guy... plus he likes drinking beer.

B.1.2. Incongruent problem (b)

1. ...*It depends how you want to go if you want to go according to the statistics there is a greater chance he is a lawyer but because of the things he does*...he is introverted, spends his time reading fiction and writing computer games, it makes more sense that he is an engineer so...I don't know I will go with that.
2. So I am gonna guess he is an engineer because he likes writing computer programs.
3. Jack is most likely an engineer is the answer, because he writes science programs and reads science fiction novels.
4. Jack is an engineer because he likes science fiction and writing computer programs.
5. He is an engineer because he likes writing computer programs.
6. ...ok *5 engineers*... you would think he is an engineer but cause *there were more lawyers* he is a lawyer.
7. He reads science fiction and writes computer programs so he would be an engineer.
8. Jack is a...engineer because they are good with computers, and he is introverted, for a lawyer you have to be active.
9. Jack is an engineer because he likes reading science fiction and writing computer programs.
10. Jack would most likely be...an engineer.
11. I would say Jack is an engineer...because he likes to write computer programs...and science fiction...and engineering is a science thing I guess.

12. This guy is an engineer, because he likes computers and science fiction, and he seems like a loner…no wife.

B.1.3. Incongruent problem (c)

1. So… this can go either way if you want to go according to…Even though there is more…there is *997 people who live in a farmhouse so it is more likely because of the things that describe to me…that he works long hours which seems more like a city type a job*, Armani suits for work, in a farmhouse area I just think of more you know ripped jeans and a T-shirt…*So it seems that he lives in a condo, but again there were only three who live in a condo so according to statistics it is more likely that he actually lives in a farmhouse*.
2. …I am gonna guess he lives in a condo because he works on Wall Street.
3. Kurt lives in a condo, he is single so he can't run a farm on his own, he wears Armani suits to work so it is not work at a farm.
4. I would say he lives in a condo because he works on Wall Street and there is probably no farms in that area.
5. Condo because if he works a lot his job should be near to where he lives, and he likes Versace and other expensive stuff. So he probably lives in a condo because he can afford it, and condo is more expansive than a farm house.
6. In a condo…or no…*997 lived in a farmhouse so it could be a farmhouse*…I am undecided.
7. Armani suits to work and shades that would be…or he lives in a condo, he can't wear those at a farm.
8. Kurt lives in a condo there is no farmhouses close to the Wall Street.
9. Kurt lives in a condo because he works on Wall Street and…he dresses well.
10. Kurt lives in a condo he is rich.
11. Kurt lives in a condo, because he works on Wall Street, and… I don't think that he would be too close to farmhouses…and he wears Armani suits.
12. Seems like a cool guy…shades, Armani…he lives in a Condo.

B.1.4. Incongruent problem (d)

1. He is most likely a nurse just because being well spoken and interested in politics and having lots of time has nothing to do with being a doctor, and he can be a very good nurse, and *there are more nurses than there are doctors*.
2. …*997 nurses well I don't know I guess he would be a doctor* sounds more like that.
3. He is a doctor because he is a male and not a lot of nurses are male, and he is well off and invests a lot of time in his career.
4. I would say Paul is a doctor because he invests a lot of time in his career and that probably takes more time than being a nurse.
5. Paul is…a doctor because he has a beautiful house doctors make more money than nurses.
6. *997 nurses…but he sounds like a doctor I guess…Ok he is a nurse*.

7. He invests a lot of time in his career, and that's why I would say he is a doctor, because I know doctors work like 24/7.
8. Paul is a doctor, not a lot of males are nurses.
9. Paul is a doctor because… he has a beautiful home so he is wealthy, and he invests a lot of time in his career and that is kind of a characteristic of doctors.
10. Paul is a doctor.
11. …Paul is a doctor…because he invests time in his career, and doctors have to do that…and he lives in a posh suburb…and doctors make more money, so I guess he can afford it.
12. Again like the other one, *there were more nurses*… Even though nurses are usually women… This Paul is probably a Nurse.

B.1.5. Incongruent problem (e)

1. So even though he likes physics and video games, so *without the statistics I would say Star Trek, but because of the statistics I will say he is most likely to watch Days of Our Lives*.
2. …He likes to play video games so most likely his favorite show is Star Trek.
3. What a nerd he watches Star Trek for sure, because number one he is a guy and they don't like watching soup operas, and he likes physics which kind of goes hand in hand with Star Trek.
4. I would say his favourite series is Star Trek because he likes physics.
5. Star Trek because Star Treks uses a lot of physics to create whatever things they want.
6. *4…Star Trek…996 Days of Our Lives…*so (b) Jeremy's favourite series is Days of Our Lives.
7. He is studying physics, plays video games…just sounds more like someone who watches Star Trek.
8. Star Trek because he plays video games and watches Star Trek so he stays at home, so he must be a nerd, so… he watches Star Trek.
9. Jeremy's favourite series is Star Trek because he is doing graduate studies in physics and likes to play video games…so it sounds like he would watch something like Star Trek.
10. So my answer is… Jeremy's favourite series is Star Trek.
11. Jeremy's favourite series is Star Trek, because he likes to stay at home and play video games.
12. He stays at home and plays video games…so obviously he likes Star Trek…he seems like a nerd haha.

B.1.6. Incongruent problem (f)

1. Even though what is described to me says that she is a sixteen year old it doesn't really make a difference, she could still be *fifty and according to the statistics she is more likely to be fifty*.

2. … I guess she is younger so I am gonna answer 16.
3. I hope she is a 16-year-old because it would be horrible if she was a fifty year old who liked to wear tight cloths and had a nose piercing. So I think she is a sixteen year old.
4. I would say she is 16 because I don't thing a fifty year old would have a nose ring or would wear tight shirts and jeans.
5. I think Ellen is 16 because that is the time girls that age are mostly fond of things on TV.
6. Sixteen because it sounds like a 16-year-old. I mean tight shirts and a nose piercing…yeah she must be 16.
7. She likes to listen to hip-hop and rap and has a nose piercing so she is sixteen.
8. Ellen is 16 because old people do not listen to hip-hop and rap.
9. Ellen is 16…because she listens to hip-hop and rap, and…wears tight cloths so it sounds like someone younger.
10. I don't really have to think about this I can just say she is sixteen. Do I have to say why? Ok then I didn't really think anything I just know she is 16.
11. I say that Ellen is 16, because I don't see a 50-year-old wearing tight clothes and listening to rap… Yeah and having a nose ring.
12. …Even though I don't want to see a 50-year-old in tight jeans and small shirt rapping to hip- hop…*there were more 50 year olds*…so maybe she is a fifty year old…yuk…haha.

B.2. Congruent problems

B.2.1. Congruent problem (a)

1. So I would assume that she buys her cloths at high-end retailers just because it seems she is very wealthy person, which doesn't mean she doesn't buy her cloths at Wal-Mart, she is just more likely to buy her cloths at more expansive place.
2. So I am gonna guess she buys her clothes at high-end retailers because she drives a Porsche.
3. (a) Karen shops at high-end retails because she drives a Porsche and she lives in a fancy penthouse.
4. Karen buys her cloths at high-end retailers because of the life style she is accustomed to, she drives a Porsche and it is very expensive to drive a Porsche.
5. High-end retailers because she drives a Porsche, that means she can afford a lot of stuff.
6. High-end retailers because…she is rich, she drives a Porsche and lives in a fancy house….or maybe her boyfriend is rich.. still the answer is (a).
7. She drives a Porsche, and lives in a nice house so she buys expensive cloths… I mean at high end-retailers.
8. I think she buys her cloths at high end-retailers because she has lots of money and drives a Porsche.

9. Karen buys her cloths at high-end retailers because she drives a Porsche and lives in a fancy penthouse, so she is well off...and would buy more expansive cloths.
10. Karen buys her cloths at high-end retailers
11. I think Karen probably buys her clothes at the high-end stores...because she drives a Porsche...sounds like she has money...So why would she shop at Wal-Mart?
12. High-end store...because she works in an office, so she has to dress nice...and she's got the money to spend...I mean she drives a Porsche.

B.2.2. Congruent problem (b)

1. So Erin is a girl, not only because she does all these things but there were *997 girls and only 3 boys so she is defiantly more likely to be a girl.*
2. I am gonna guess she is a girl because she likes to gossip and go shopping.
3. Erin is most likely a girl (a) is my answer because she likes to shop, and gossip with her friends and Erin with an E is more of a girl's name whereas Aaron with a double A is more boy's name.
4. I would say Erin is a girl because she likes to go shopping and gossiping.
5. A girl because she likes to gossip.
6. So she is a girl because shopping, sleepovers, gossip and yeah.
7. She loves shopping and gossiping, and...art I guess, she is a girl.
8. I think Erin is a girl because she likes shopping and sleepovers and she likes to gossip...guys don't do that so... she is a girl.
9. Erin is a girl because her favourite subject is art, and...she likes shopping and gossiping about other kids at school.
10. Erin is a girl
11. I would say Erin is a girl...because boys don't like to shop and have sleepovers...that is a girly thing.
12. Obviously Erin is a girl...13-year-old boys don't gossip about friends, and have sleepovers. I hope they don't haha.

B.2.3. Congruent problem (c)

1. ... Because *there were 997 who have a tattoo I am gonna say that Jay has a tattoo*.
2. ...He served a short time in jail and he listens to punk music so I guess he has a tattoo.
3. Jay has a tattoo (a) is my answer because *there were more participants with a tattoo* than without, he also was in prison so he probably got a tattoo there.
4. He has a tattoo, he was in prison, listens to punk music and he just has that kind of personality.
5. He has a tattoo because he listens to punk and was in prison, obvious.
6. *997...and hee...he has a tattoo*.
7. He was in prison, and listens to punk music, so he would have a tattoo.

8. Jay has a tattoo because he listens to punk and was in prison so he…he has one.
9. Jay has a tattoo because he served a short time in prison, and…listens to punk music.
10. Jay has a tattoo.
11. Jay has a tattoo. I mean he was in prison, and he listens to punk music…so yeah…he definitely has a tattoo.
12. This guy definitely has a tattoo. Prison, punk music…probably covered in tattoos haha.

B.2.4. Congruent problem (d)

1. …She is most likely to be a kindergarten teacher not only because she…*the statistics show she is most likely kindergarten teacher*, but also because of things she does…Well it doesn't really make a difference an executive manager can be committed to her family and watch daily cartoons with her kinds so.
2. Ok so *996 kindergarten teachers I guess she is most likely to be one of the kindergarten teachers*
3. (a) Lily is a kindergarten teacher because *there were more teachers in the study* and she is committed to her family, which means she does not spend a lot of time at her job which is needed for an executive manager.
4. I would say she is an executive manager because she watches daily cartoons with her kids, if she was a kindergarten teacher she would probably be in school at that time so she would not watch it.
5. She has kids, she likes spending time with them so I am guessing kindergarten teacher.
6. *996*…family…Lilly is a kindergarten teacher.
7. She has a lot of kids and watches cartoon shows with her kids so it sounds like a kindergarten teacher.
8. Lily is a kindergarten teacher because she watches cartoons and if she were a manager she would not have time for that.
9. Lily is a kindergarten teacher because she has 3 kids and…she likes spending time with them…so she just kind of sounds like one.
10. It is most likely that Lilly is a kindergarten teacher.
11. I say Lilly is a kindergarten teacher, because she likes watching the cartoons with her kids.
12. I say she is an executive…because kindergarten teacher have to be at work when the cartoons are on…but executives can work whenever they want.

B.2.5. Congruent problem (e)

1. So not only does she do very girlish things, *the statistics also show that it would make more sense that she is a Britney Spear's fan*.
2. …She is someone young so now days I guess she would be Britney Spear's fan

3. Tara is most likely a Brittney Spear's fan for sure because she is 15, I don't think she even knows who Bruce Springsteen is cause I don't, and there were more Britney Spear's fans and she is a girl and yeah
4. I would say Tara is a Britney Spears fan because she is 15.
5. Tara is a Britney Spears fan because she likes shopping, and she is 15.
6. *4…996 Britney so (b)* Tara is a Britney Spears fan.
7. Loves shopping, talking to her friends about crushes…so she is young and she is a Britney Spears fan.
8. Tara is a Britney Spears fan because she is 15 and it just makes more sense that she would be a Britney Spears fan.
9. Tara is a Britney Spears fan because she loves to go shopping at the mall and…yeah because she is 15 she is more likely to be a Britney Spears fan.
10. Definitely a Britney Spears fan.
11. I guess Tara is a Britney fan, because she is a 15-year-old girl.
12. Tara is a Britney fan, first she is a girl, and she is only 15…she probably doesn't even know who Springsteen is.

B.2.6. Congruent problem (f)

1. So what she does really has nothing to do with weather she is American or French and *because the statistics say that there were more French people she is most likely to be French*.
2. *…1000 people were tested, 5 Americans…* Martine…just thinking about what she is doing…I guess because she is fashionable dresser she is American.
3. She is more likely to be French because she speaks two languages I guess, and she has good fashion sense.
4. I would say Martine is French because her name is Martine.
5. French because French people are known for cooking, they are well dressed and she speaks French.
6. Ok so because she is bilingual and we are in Canada I would say Martine is French.
7. I guess all this stuff fashionable dresser, great cook, reading…would be something a French person would do.
8. Martine is French because she knows how to cook and she is a fashionable dresser… Americans are not good cooks they only eat fast food.
9. Martine is French because she is bilingual so…that makes me think that she speaks both English and French.
10. Martine, so she is French
11. Martine is French, because she is fashionable and bilingual…most likely French haha.
12. Well she is probably French, because she is bilingual.

B.3. Neutral problems

B.3.1. Neutral problem (a)

1. So…I am just going to automatically go with John Kerry because *in the study there were 995 people who went for John Kerry out of a 1000, so it's just more likely he is a participant there and there is nothing more describing him that would lead me to believe that he is out of the only five who campaigned for George Bush*.
2. …I am gonna guess because *I don't know there were more people who campaigned for John Kerry in this study* so I am gonna guess he campaigned for John Kerry.
3. Hm Jim campaigns for John Kerry… *because there were more participants that campaigned for John Kerry in total*.
4. …I have absolutely no idea but I am gonna say he campaigns fro John Kerry.
5. …Wow… black hair, 5 ft…short fellow, father of two girls, drives a yellow van…he is not conservative cause of his car, so I am guessing John Kerry.
6. …Ok so John Kerry because *995 campaigned for John Kerry and only five For Bush*.
7. …Because of the van that is covered with posters I would say John Kerry.
8. …Jim campaigned for John Kerry because his van is covered with posters and…John Kerry's campaign was advertised with stickers and…all that stuff.
9. *5 for George W. Bush, ok so because there were more participants who campaigned for John Kerry*, I'm gonna say Jim campaigned for John Kerry.
10. …So I am just guessing again he campaigns for George W. Bush?
11. A…well….I guess since this guy is a family man…he would vote for Bush because he is a family man too…and because he drives a van? I don't know…yeah Bush I guess.
12. Well… Ok….*I guess because there were more people who wanted Kerry*, most likely this guy wanted Kerry too.

B.3.2. Neutral problem (b)

1. So again even though…the things that describe.. Actually, it doesn't really make a difference I would say that Casey is a man because she .. or sorry Casey doesn't do anything that is very typical for a woman running and watching a good movie can be for both men and women and *according to the statistics there were much more men than women so it is definitely more likely that Casey is a man.*
2. …I am guessing that sounds like a girl's name, so I am guessing she is a woman.
3. Hm…Casey is most likely a man because *there were more participants who are men than women*.
4. I would say Casey is a woman because she likes to watch a movie and…running.
5. A woman, a writer.

6. …*996 men* so Casey is a man.
7. …Casey is a girl's name isn't it? Ok well she is a woman.
8. …It can be both…this does not make sense…Ok I am going to say that Casey is a man. Do I have to say why? Ok because it says that there were a 1000 people tested and *there were 996 men and 4 women*… so it is a greater chance that Casey is a man.
9. More men so I would say she is or Casey is…it sounds like a girls name…Ok but Casey is a man because there were more participants who were men.
10. …I don't know guessing because the name it's a woman
11. Well…a writer…I don't know…I guess Casey is a woman…because she is a writer and likes running…I guess?
12. Ok…Casey has two brothers… running…*well there were 996 men studied*, so most likely Casey is a man.

B.3.3. Neutral problem (c)

1. … So I am assuming that because *there were so many more who play drums he plays a drum.* He is most likely to play a drum even though what describes to me…it doesn't really make a difference I would say he plays a drum.
2. …Most likely to play drums *because there were more people who play drums.*
3. Tom most likely plays the drum *because there were more participants who played a drum than a saxophone*.
4. I would say he plays a saxophone because it is cheaper than the drums and he just bought a second hand car so he probably does not have money.
5. He plays a saxophone because he is a "playa" (has no steady girlfriend).
6. *Again 997 who play the drums* so Tom plays the drums.
7. …Because he is 20 so he is younger, I would say he would be more into playing drums.
8. I think Tom plays the drums because *there were more people who play the drums*.
9. *997 who play the drums, so Tom plays the drums*.
10. …Again I am just guessing drums… I don't know why
11. I guess Tom play the Sax…because…he just bought a second hand car…doesn't really have money…so…a saxophone is cheaper than drums? I don't know haha.
12. So he is a musician…well…*there were only three studied who played saxophone*…so I guess he is most likely a drummer.

B.3.4. Neutral problem (d)

1. … It depends *again if I didn't have statistics I would say he was a basketball player but because there were more pool players I am gonna say a pool player*.
2. … Well there *were 997 pool players so I guess he is a pool player*.

3. *There were 997 pool players so it is most likely that he is a pool player because it is a higher probability*.
4. He is a pool player…because he drives a light grey car.. I don't know a basketball player would drive a different car.
5. … O wow… green eyes yes… he is 29, lives in New York, light-grey coloured car…Pool player because to be a pool player you have to be calm and he has a grey colour and grey is a calm colour.
6. Ok…so *997 pool players… Jason is a pool player*.
7. …He is 29 and I think that is too old for a basketball player. I don't know maybe not, but I'll say he is a pool player.
8. …He is a pool player because *there were more pool players*…There is only 0.3% chance that he is a basketball player.
9. *997 pool players, so Jason is a pool player because there were more participants*.
10. …Pool Player…just my gut feeling haha.
11. Haha…well I say he likes to play pool…because he lives…in New York, and I guess there are more pool halls in New York.
12. He lived his whole life in New York…. but *more people here played pool*…so I guess he probably plays pool.

B.3.5. Neutral problem (e)

1. So I am gonna go with that he lives in Los Angeles because the *statistics say that that is more likely*.
2. …Ok so if he likes watching basketball and there were *996 people from Los Angeles I guess he lives in Los Angeles*.
3. *So there were 996 from Los Angeles and only four from New York, so he is most likely from Los Angeles because there were more people tested from Los Angeles*.
4. I would say Christopher lives in New York because he shares an apartment with a friend not his girlfriend
5. …Can it be both? Christopher lives in New York why? Yes he is old he lives with a friend cause apartments in New York are expensive.
6. *4 who live in New York…and 996 in Los Angeles, so he lives in L.A*.
7. …Hm just because he watches basketball or no..yeah I'll say he lives in Los Angeles.
8. He lives in Los Angeles because basketball is more popular there than in New York.
9. 4 who live in New…Ok so Christopher lives in Los Angeles because there were more participants who lived in Los Angles.
10. …I have to guess again, so I'll say New York.
11. I say he lives in New York, because he shares an apartment with his friend…and lives in New York…yeah.
12. Probably lives in LA, because there were only 4 people surveyed who live in New York.

B.3.6. Neutral problem (f)

1. . . .*Ok since there is more English majors* I am going to guess English major.
2. He is more of an *English major just because there were more tested*
3. . . .I would say he is an English major because his favourite food is meat balls with pasta and that is more of an Italian food and English is more. . .tough subject than computer science.
4. . . .English because he is 20 years old, and down town area, and mostly in down town area are people who are artsy.
5. . . .Ok I don't know. . . oh *5 computer science and 995 English*. . .more likely so it is more likely he is an English major, I didn't look at these before can I go back? Ok whatever.
6. . . .Hmm. . .well. . .because he lives in Toronto I would say he is an English major.
7. English major because *there were more English majors*.
8. . . .*Ok because there were only 5 computer majors and 995 English majors I would say that Matt is an English major* because it is . . .what do you call that?. . .well it is more likely that he is.
9. . . .Ok for this one I do not know. . .I'll say a computer science major. I know more people who are majoring in computer science so I'll just pick this one.
10. . . .I guess because he likes pasta, and lives in downtown. . . I guess he is an English major. . .I don't know.
11. English. . .studies English because there are *only 5 Computer people here*. . .plus he likes pasta? Haha Ok English.

References

Ball, L. J., Phillips, P., Wade, C. N., & Quayle, J. D. (2006). Effects of belief and logic on syllogistic reasoning: Eye-movement evidence for selective processing models. *Experimental Psychology, 53*, 77–86.

Burns, B. D., & Wieth, M. (2004). The collider principle in causal reasoning: Why the monty hall dilemma is so hard. *Journal of Experimental Psychology: General, 133*, 434–449.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*, 624–652.

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

Crutcher, R. J. (1994). Telling what we know. *Psychological Science, 5*, 241–252.

Denes-Raj, V., & Epstein, S. (1994). Conflict between intuitive and rational processing: When people behave against their better judgement. *Journal of Personality and Social Psychology, 66*, 819–829.

De Neys, W. (2006a). Automatic-heuristic and executive-analytic processing in reasoning: Chronometric and dual task considerations. *Quarterly Journal of Experimental Psychology, 59*, 1070–1100.

De Neys, W. (2006b). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science, 17*, 428–433.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2005). Working memory and everyday conditional reasoning: Retrieval and inhibition of stored counterexamples. *Thinking & Reasoning, 11*, 349–381.

Epstein, S. (1994). Integration of the cognitive and psychodynamic unconscious. *American Psychologists, 49*, 709–724.

Epstein, S., & Pacini, R. (1999). Some basic issues regarding dual-process theories from the perspective of cognitive–experiential self-theory. In S. Chaiken & Y. Trope (Eds.), *Dual Process Theories in Social Psychology* (pp. 462–482). New York: Guilford Press.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*, 215–251.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

Evans, J. St. B. T. (1984). Heuristic and analytic processing in reasoning. *British Journal of Psychology, 75*, 451–468.

Evans, J. St. B. T. (2002). Logic and human reasoning: An assessment of the deduction paradigm. *Psychological Bulletin, 128*, 978–996.

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. (2007a). On the resolution of conflict in dual process theories of reasoning. *Thinking and Reasoning,* doi:10.1080/13546780601008825.

Evans, J. St. B. T. (2007b). The heuristic-analytic theory of reasoning: Extension and Evaluation. *Psychonomic Bulletin and Review, 13*, 378–395.

Evans, J. St. B. T., & Over, D. E. (1996). *Rationality and reasoning*. Hove, UK: Psychology Press.

Feldman Barrett, L., Tugade, M. M., & Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin, 130*, 553–573.

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.

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.

Gigerenzer, G., & Regier, T. (1996). How do we tell an association from a rule?: Comment on Sloman (1996). *Psychological Bulletin, 119*, 23–26.

Glanzer, M., & Cunitz, A. R. (1966). Two storage systems in free recall. *Journal of Verbal Learning and Verbal Behaviour, 5*, 351–360.

Goel, V. (1995). *Sketches of thought*. Cambridge, MA: MIT Press.

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

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., & Moutier, S. (1996). Deductive reasoning and experimental inhibition training: The case of the matching bias. *Current Psychology of Cognition, 15*, 409–434.

James, W. (1890). *The principles of psychology*. Oxford, England: Holt.

Just, M. A., Carpenter, P. A., & Wooley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 111*, 228–238.

Kahneman, D. (2002, December). 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.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgement. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgement* (pp. 49–81).

Kahneman, D., & Frederick, S. (2005). A model of heuristic judgement. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 267–293). Cambridge, MA: Cambridge University Press.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgement under uncertainty: Heuristics and Biases*. Cambridge, MA: Cambridge University Press.

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

MacDonald, A. W., Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science, 288*, 1835–1838.

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

Moors, A., & de Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin, 132*, 297–326.

Moutier, S., & Houdé, O. (2003). Judgements under uncertainty and conjunction fallacy inhibition training. *Thinking and Reasoning, 9*, 185–201.

Oaksford, M., & Chater, N. (1998). *Rationality in an uncertain world: Essays on the cognitive science of human reasoning*. Hove, UK: Psychology Press.

Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin & Review, 11*, 988–1010.

Pashler, H., Johnston, J., & Ruthruff, E. (2001). Attention and performance. *Annual Review of Psychology, 52*, 629–651.

Payne, J. W. (1994). Thinking aloud: Insights into information processing. *Psychological Science, 5*, 241–248.

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.

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science, 306*, 443–447.

Shiffrin, R. M. (1988). Attention. In R. C. Atkinson, R. J. Herrnstein, G. Lindsay, & R. D. Luce (Eds.). *Steven's handbook of experiment psychology* (Vol. 2, pp. 739–811). Oxford, England: Wiley.

Simoneau, M., & Markovits, H. (2003). Reasoning with premises that are not empirically true: Evidence for the role of inhibition and retrieval. *Developmental Psychology, 39*, 964–975.

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

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

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

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

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science, 185*, 1124–1131.

Research Article

Smarter Than We Think

When Our Brains Detect That We Are Biased

Wim De Neys,[1] Oshin Vartanian,[2] and Vinod Goel[2]

[1]University of Leuven and [2]York University

ABSTRACT—*Human reasoning is often biased by stereotypical intuitions. The nature of such bias is not clear. Some authors claim that people are mere heuristic thinkers and are not aware that cued stereotypes might be inappropriate. Other authors claim that people always detect the conflict between their stereotypical thinking and normative reasoning, but simply fail to inhibit stereotypical thinking. Hence, it is unclear whether heuristic bias should be attributed to a lack of conflict detection or a failure of inhibition. We introduce a neuroscientific approach that bears on this issue. Participants answered a classic decision-making problem (the "lawyer-engineer" problem) while the activation of brain regions believed to be involved in conflict detection (anterior cingulate) and response inhibition (lateral prefrontal cortex) was monitored. Results showed that although the inhibition area was specifically activated when stereotypical responses were avoided, the conflict-detection area was activated even when people reasoned stereotypically. The findings suggest that people detect their bias when they give intuitive responses.*

Half a century of reasoning and decision-making research has sketched a bleak picture of human rationality. Hundreds of studies have shown that when making decisions, people seem to overrely on intuitions and stereotypical beliefs, instead of basing their decisions on more demanding, deliberative reasoning. Although this intuitive, or so-called heuristic, thinking might sometimes be useful, it will often cue responses that are not warranted from a normative point of view. For example, jurors' decisions to sentence a Black defendant to death may be based more on negative stereotypical beliefs about Black people's criminal nature than on objective criteria, such as the number of suspects in the case or previous convictions of the defendant (e.g., Eberhardt, Davies, Purdie-Vaughns, & Johnson, 2006). Likewise, people's risk assessment tends to be based on the operation of simple heuristic associations, rather than on a consideration of the relevant statistics. Despite numerous health-education programs, for example, teenagers tend to ignore the warnings about the dangers of smoking; basing their thinking on the stereotypical idea that only old people get lung cancer (e.g., Peters, McCaul, Stefanek, & Nelson, 2006; Slovic, 2000), they erroneously conclude that smoking is less harmful for younger people (Slovic, 2000).

A classic demonstration of the pervasive impact of intuitive operations on people's decision making is found in Kahneman and Tversky's (1973) studies of base-rate neglect. In these studies, people responded to problems in which a stereotypical description cued a salient but inappropriate response. The problems first provided information about the composition of a sample (e.g., a sample with 995 lawyers and 5 engineers), and then people were told that they would see a short personality description of a randomly selected individual from the sample. The task was to indicate which group the individual most likely belonged to. Statistically speaking, it was likely that a randomly drawn individual would be from the larger, rather than the smaller, group. However, people might be tempted to respond on the basis of stereotypical beliefs cued by the personality description. Indeed, Kahneman and Tversky observed that the vast majority of well-educated university students failed to answer the problem correctly. Even university professors were not immune to the heuristic bias, seeming to neglect the crucial base-rate information.

Although it is clear that people are often biased, the nature of this bias is poorly understood. Some authors claim that people reason heuristically by default and that most of the time they are simply not aware that their intuitions might be wrong. The dominance of intuitive thinking is attributed to a failure to monitor the output of the heuristic reasoning process. In this view, because of lax monitoring, people fail to detect that an intuitive response conflicts with the response favored by probability. The problem is that people do not know that their judgment is biased. This view has been popularized by the work of authors such as Kahneman (2002) and Evans (1984, 2003).

Address correspondence to Wim De Neys, Experimental Psychology Lab, University of Leuven, Tiensestraat 102, 3000 Leuven, Belgium, e-mail: wim.deneys@psy.kuleuven.be.

However, other authors, such as Epstein (1994; Epstein & Pacini, 1999) and Sloman (1996), argue that people always engage in probabilistic thinking and detect when their intuitive response is inappropriate. According to this view, heuristic and probabilistic thinking operate in parallel: People simultaneously engage in both intuitive and more deliberate probabilistic thinking. Consequently, people readily detect a conflict between their stereotypical intuition and the appropriate response. Hence, in this view, there is nothing wrong with the conflict-monitoring process. People know that their intuitive responses are not valid. The problem is that despite this knowledge, they do not always manage to inhibit tempting intuitive beliefs. Thus, people "behave against their better judgment" (Denes-Raj & Epstein, 1994, p. 819) when they give a stereotypical response: They detect that they are biased but fail to block the biased response. In sum, in this view, biased decisions are attributed to an inhibition failure, rather than to a conflict-detection failure per se.

Clarifying the exact nature of the heuristic bias is important for the development of reasoning and decision-making theories. The issue also has far-reaching implications for views of human rationality (e.g., see De Neys, 2006; Stanovich & West, 2000). However, it is hard to decide between the alternative views on the basis of traditional reasoning data (Evans, 2007). The problem is at least in part due to the fact that reasoning and decision-making studies tend to focus on the accuracy of the output (i.e., whether or not people give the correct response), and not on the underlying cognitive processes (e.g., Hoffrage, 2000). Although recently there have been some initial attempts to break the stalemate by developing behavioral processing measures of conflict detection during reasoning (e.g., De Neys & Glumicic, 2008), the rival views persist. The present study addresses this issue by focusing on the neural basis of conflict detection and response inhibition.

In the past decade, numerous imaging studies have established that conflict detection and actual response inhibition are mediated by two distinct regions in the brain. Influential work on cognitive control (e.g., Botvinick, Cohen, & Carter, 2004; Ridderinkhof, Ullsperger, Crone, & Nieuwenhuis, 2004; van Veen & Carter, 2006) has shown that detection of an elementary conflict between competing responses is among the functions of the medial part of the frontal lobes, more specifically, the anterior cingulate cortex (ACC). Whereas the ACC signals the detection of conflict, responding correctly (i.e., overriding the erroneous, prepotent response) depends on the recruitment of the more lateral part of the frontal lobes. Indeed, there is abundant evidence indicating that the right lateral prefrontal cortex (RLPFC), in particular, plays a key role in response inhibition (e.g., for a review, see Aron, Robbins, & Poldrack, 2004). Recent imaging work in the reasoning and decision-making field also suggests that these same two brain structures, the ACC and RLPFC, mediate the detection of conflict between intuition and probability and the subsequent inhibition of the intuitive response in classic reasoning tasks (e.g., De Martino, Kumaran, Seymour, & Dolan, 2006; Goel & Dolan, 2003; Prado & Noveck, 2007; Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003).

This background suggests how examining the brain might help resolve the dispute about the nature of heuristic bias. Solving a classic decision-making problem that cues a salient but inappropriate intuitive response requires that reasoners first detect that the intuitive response conflicts with the probabilistic response and then successfully inhibit the intuitive response. If the ACC and RLPFC mediate these conflict-detection and inhibition processes, respectively, then correct probabilistic reasoning should be associated with increased activation in both areas (De Martino et al., 2006). It should therefore be possible to clarify the nature of the intuitive bias by contrasting ACC and RLPFC activations observed when participants give probabilistic and stereotypical responses. The bias-as-inhibition-failure and bias-as-detection-failure views make different predictions with respect to the activation of the conflict-detection region. If the former view is right, and people detect that the intuitive response conflicts with more normative probabilistic considerations, the ACC should be activated whether or not people reason stereotypically. However, if the latter view is right, and biased decisions arise because people fail to detect that the intuitive response is inappropriate, people do not experience a conflict when they give a stereotypical response and the ACC should not be activated under these conditions.

To test these predictions, we conducted a functional magnetic resonance imaging (fMRI) study, focusing on participants' ACC and RLPFC activations while they were responding to problems that were modeled after Kahneman and Tversky's (1973) classic base-rate problems, which instigated much of the debate on heuristics and human rationality (Barbey & Sloman, 2007). We also included a number of control problems in which there was no conflict between the cued intuitive response and the probabilistic response. If ACC activation signals the detection of a conflict between probabilistic thinking (cued by consideration of the base rates) and stereotype-based intuition, the ACC would not be expected to be activated in this control condition.

EXPERIMENT

Method

Participants

Thirteen participants (mean age = 27.9 years, $SD = 3.7$; mean education level = 16.1 years, $SD = 1.1$) gave informed consent to participate in the study in return for a monetary reimbursement.

Stimuli

We constructed four types of base-rate problems to test our hypotheses: incongruent, congruent control, neutral control, and heuristic control items. In the crucial incongruent items, the stereotype-based response cued by the description conflicted with the response cued by the base rates, as in the classic, standard problems.[1] In the three kinds of control problems, responses cued by base rates and responses cued by stereotypical thinking did not conflict: In each congruent control item, the description described a typical member of the larger group, so that stereotypical beliefs and base rates cued the same response. In the neutral control items, the descriptions were completely neutral (e.g., "Jack has brown hair and green eyes"); hence, these items did not trigger stereotypical, heuristic responses, and participants were expected to respond by relying on the base rates. Finally, in the heuristic control items, the base rates were neutral (e.g., a sample with 500 lawyers and 500 engineers) and did not cue a response; consequently, responses depended on stereotypical thinking about the descriptions. Table 1 presents examples of the four kinds of items.

Participants answered 24 problems of each type (96 problems in all). The problems were based on a wide range of stereotypes (e.g., involving gender, age, and race) and were selected on the basis of an extensive pilot study.

Instructions

Before going into the scanner, participants were familiarized with the task format. The problems did not explicitly repeat the classic lines about the total sample size and random sampling (e.g., "A total of 1,000 people were tested . . . . The description was drawn at random from the sample . . . ."), in order to avoid repetition and limit the amount of text presented. However, this information was clearly emphasized in the instructions. To make sure that participants grasped the concept of random sampling, we included a training problem in which we demonstrated how 1 description was drawn from an urn containing 10 descriptions (e.g., Gigerenzer, Hell, & Blank, 1988). We also clarified that participants needed to think as statisticians when answering the problems (e.g., Schwartz, Strack, Hilton, & Naderer, 1991). These simple manipulations have been shown to minimize misinterpretation of the task.

Stimulus Presentation

The items were presented in one of two random orders. The beginning of a trial was signaled by a fixation cross that was presented for 500 ms. Next, the problem was presented in three parts. First, the line with the base-rate information was presented for 4,000 ms. Second, the description was presented for 5,000 ms (the base rates remained on the screen). Finally, the question and two response alternatives appeared. Once the question appeared, the entire problem remained on the screen for another 8,500 ms. Hence, each trial lasted exactly 18,000 ms. Participants responded by pressing one of two buttons on a key pad.

fMRI Scanning Technique

Participants were scanned in a 4-T magnet at the Robarts Institute in London, Ontario (Canada). Twenty-three T2*-weighted interleaved multishot, contiguous, echo-planar images, 5 mm thick, were acquired; the voxel size was uniformly $3.44 \times 3.44 \times 5.0$ mm. The images were axially positioned to cover the whole brain. A total of 624 volume images was acquired over two sessions (312 volumes per session); the repetition time (TR) was 3 s/volume. The first 6 volumes in each session were discarded (leaving 306 volumes per session). Each session lasted 15.6 min. The scanner was synchronized with the presentation of each trial.

fMRI Data Analysis

Data were analyzed using SPM2 (Friston et al., 1995). Each volume was realigned to the first image of the session. Head movement was less than 2 mm in all cases. The images were smoothed with an isotropic Gaussian kernel with full width at half maximum equal to 12 mm.

Condition effects at each voxel were estimated using a general linear model (GLM), and regionally specific effects were compared using linear contrasts in the GLM. Each contrast produced a parametric map of the t statistic, which was subsequently transformed to a normal Z distribution at each voxel. The blood-oxygenation-level-dependent (BOLD) signal was modeled as a hemodynamic response function during the interval between the presentation of the description and the motor response, on a trial-by-trial, subject-by-subject basis. The presentation of the base rates and the motor response were incorporated into the design but modeled out of the analysis by assigning null weights to their corresponding regressors.

The exact locations of our ACC and RLPFC regions of interest (ROIs) were based on previous work by Klein et al. (2007) and Goel and Dolan (2003), respectively. The ROIs were spheres (12-mm radius) centered on the voxels that showed peak activation in those studies: a right inferior lateral prefrontal ROI (coordinates of the center voxel = 51, 21, 12)[2] and a more medial frontal ACC ROI (coordinates of the center voxel = 1, 15, 43). Figure 1 illustrates where these regions are located in the brain. Reported activations in the ROIs were significant at a voxel-level intensity threshold of $p < .01$ (uncorrected), using a random-effects model.

[1]We assumed that our incongruent problems would elicit the same kind of biases as the classic problems did. Responses in line with the base rates are referred to as "correct." Strictly speaking, however, the stereotype-based responses do not necessarily represent normative violations. In our problems, both categories of responses can be technically consistent with probability theory. Our point is that responses in line with base rates are much more likely to reflect probabilistic consideration of base rates than are responses in line with stereotypes.

[2]SPM2 uses a standard brain from the Montreal Neurological Institute (MNI) as its reference brain. Therefore, all coordinates reported in this article are in standard MNI space.

TABLE 1
Examples of the Four Kinds of Item Types

Incongruent

Study with 5 engineers and 995 lawyers.

Jack is 45 and has four children. He shows no interest in political and social issues and is generally conservative. He likes sailing and mathematical puzzles.

What is most likely?

a. Jack is an engineer^{+}
b. Jack is a lawyer*

Congruent control

Study with 5 Swedish people and 995 Italians.

Marco is 16. He loves to play soccer with his friends, after which they all go out for pizza or to someone's house for homemade pasta.

What is most likely?

a. Marco is Swedish
b. Marco is Italian*$^{+}$

Neutral control

Study with 5 people who campaigned for Bush and 995 who campaigned for Kerry.

Jim is 5 ft. and 8 in. tall, has black hair, and is the father of two young girls. He drives a yellow van that is completely covered with posters.

What is most likely?

a. Jim campaigned for Bush
b. Jim campaigned for Kerry*

Heuristic control

Study with 500 forty-year-olds and 500 seventeen-year-olds.

Rylan lives in Buffalo. He hangs out with his buddies every day and likes watching MTV. He is a big Korn fan and is saving to buy his own car.

What is most likely?

a. Rylan is forty
b. Rylan is seventeen^{+}

Note. For each item, the table presents the information given to participants (sample composition, individual description), along with the question to be answered and response options. Symbols have been added to identify responses cued by the base-rate information (*) and by stereotypes (+).

Results and Discussion

Behavioral Results

Behavioral scores were in keeping with expectations. As Table 2 shows, participants answered nearly all the control problems correctly. On average, more than 90% of these items were answered correctly.[3] However, participants were much less accurate in responding to the incongruent problems, $F(1, 12) = 30.69$, $p_{\text{rep}} = .99$, $\eta_p^2 = .72$. As Kahneman and Tversky (1973) found, participants were biased by their stereotypical beliefs on the majority of the incongruent trails. The base-rate response (i.e., response cued by the base rates) was selected in only 45% of these trials. These findings were mirrored in the response latencies. Overall, control problems were answered more quickly than incongruent problems, $F(1, 12) = 13.93$, $p_{\text{rep}} = .97$, $\eta_p^2 = .54$. Stereotype-based responses to the incongruent problems tended to be given more quickly than base-rate responses, $F(1, 12) = 4.6$, $p_{\text{rep}} = .87$, $\eta_p^2 = .28$.

fMRI Results

We started by contrasting ACC and RLPFC activations for base-rate and stereotype-based responses to the incongruent problems (i.e., base-rate responses – stereotype-based responses). As expected, RLPFC activation increased when people refrained from stereotypical thinking and selected the base-rate response (coordinates of peak activation: 56, 24, 18; $Z = 2.37$). This finding is consistent with the general idea that this area is typically involved in inhibitory control (e.g., Aron et al., 2004).

[3]The few control problems that were not answered correctly were discarded from the remaining analyses.

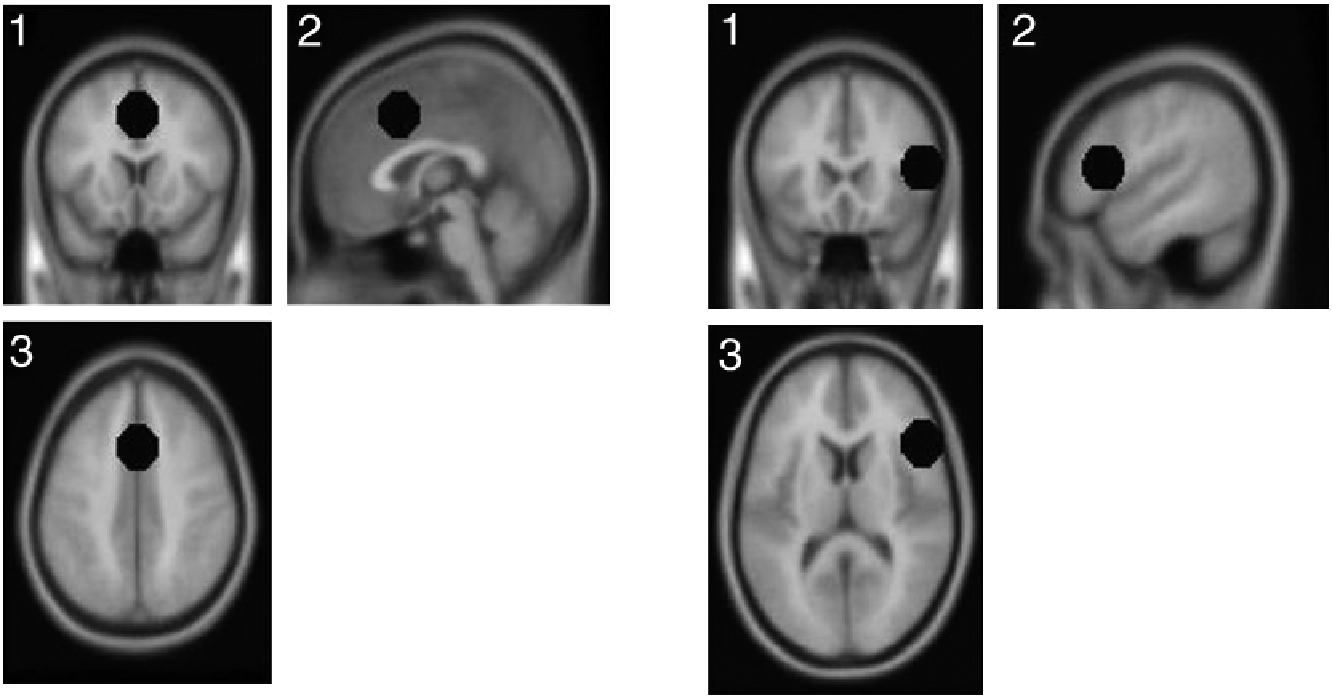

Fig. 1. Brain images showing the regions of interest (ROIs) in anterior cingulate cortex (left panel) and right lateral prefrontal cortex (right panel). The location of each ROI is superimposed on coronal (1), sagittal (2), and transverse (3) sections of a magnetic resonance image, which is in standard space.

With respect to the ACC, the direct contrast between base-rate and stereotype-based responses on the incongruent trials did not show any differential activation, even though we used a very liberal activation threshold and ROI definition. This finding is consistent with the claim that conflict detection is successful even when participants fail to select the appropriate, base-rate response.

However, the lack of differential ACC activation does not suffice to validate this claim. Alternative accounts can be put forward. For example, the ACC might mediate a more general function (e.g., directing attention) that is unrelated to conflict detection but is always engaged when solving decision problems. Alternatively, the ACC might not be involved in decision making and therefore might never be activated in this task. To rule out such explanations of the lack of differential ACC activation for base-rate and stereotype-based responses, we had to establish that the ACC specifically signals the detection of a conflict between base rates and stereotypical thinking. This is where the control problems came into play.

In contrast with the incongruent problems, the control problems did not present a conflict between stereotype-based and base-rate responses: Either probabilistic and stereotypical thinking cued the same response (congruent problems), or the problems cued only a base-rate (neutral control) or a stereotype-based (heuristic control) response. If the ACC signals successful detection of the conflict between the cued responses for incongruent problems (whatever the final response may be), ACC activation should differ between incongruent and control trials. This prediction was confirmed. We observed significant ACC activation in all three contrasts of control trials with incongruent trials: incongruent minus congruent control (−2, 24, 42; $Z =$

TABLE 2

Performance as a Function of Trial Type

Trial type	Score (percentage correct)	Reaction time (ms)
Congruent control	93 (0.11)	2,806 (1,304)
Heuristic control	93 (0.08)	2,894 (1,431)
Neutral control	88 (0.18)	3,056 (1,222)
Incongruent, base-rate response	45 (0.32)	4,044 (1,857)
Incongruent, stereotype-based response	55 (0.32)	3,501 (1,483)

Note. Standard deviations are given in parentheses. On congruent control trials, base rates and stereotypes cued the same response. On heuristic control trials, only the descriptions cued a response. On neutral control trials, only the base rates cued a response. On incongruent trials, base rates and descriptions cued conflicting responses.

2.76), incongruent minus neutral control (0, 26, 44; $Z = 2.4$), and incongruent minus heuristic control (0, 26, 44; $Z = 2.91$). If the ACC mediated a general process that is always engaged in decision making, or if the ACC were simply not involved in decision making, these contrasts should not have yielded significant activations. Furthermore, we never observed activation in the ACC region when we contrasted the activations for different kinds of control problems (i.e., congruent control – neutral control, congruent control – heuristic control, neutral control – heuristic control).[4] These findings establish that the ACC specifically responds to the conflict between the cued responses in the classic, incongruent base-rate problems.

In sum, the crucial finding is that stereotype-based and probabilistic responses to the classic base-rate problems differed only in RLPFC recruitment. Responding to incongruent problems did engage the ACC region, but the activation did not differ between base-rate and stereotype-based responses.

GENERAL DISCUSSION

In the present study, we tried to disentangle two rival views on the nature of the heuristic reasoning bias. Participants solved classic base-rate problems while we monitored the activation of two frontal brain regions believed to be involved in conflict detection (i.e., the ACC) and response inhibition (i.e., the RLPFC). Results showed that although the inhibition area was activated only when people avoided tempting stereotype-based responses, the conflict-detection area was activated even when people reasoned stereotypically. On control problems in which the cued base rates and stereotype did not conflict, the ACC was not engaged. The RLPFC and ACC activation patterns lend credence to the view that biased decision making results from a failure to override intuitive heuristics, and not from a failure to detect the conflict between these heuristics and normative information. If people were mere heuristic thinkers and neglected probabilistic sample-size considerations, our participants should have failed to detect that their intuitive responses conflicted with the base rates, and the ACC should not have been activated.

We noted that there have been some initial attempts to develop behavioral processing measures of conflict detection during reasoning. For example, De Neys and Glumicic (2008) presented participants with an unannounced recall test after they had solved a set of base-rate problems. The authors reasoned that successful conflict detection would result in deeper processing of the base-rate information, and consequently better memorization of that information. Results indicated that participants had no trouble recalling the base-rate information of the incongruent problems they had previously answered (even when they had not answered correctly). Base-rate information of congruent control problems, in which the base rates did not conflict with the intuitive response, was not remembered as well. Hence, this behavioral study is consistent with the present imaging findings in indicating that successful conflict detection is omnipresent, regardless of whether participants answer problems correctly or incorrectly.

Our findings indicate that heuristic bias should be attributed to an inhibition failure. We characterize inhibition as a basic cognitive mechanism whereby participants actively try to withhold a salient, but inappropriate, default response. A failure to inhibit an intuitively cued stereotype-based response after successful conflict detection thus implies that the heuristic response was not overridden. One might wonder whether the inhibition failure also has an affective component (e.g., do people "regret" their stereotype-based response after an inhibition failure?). Our data do not speak to this issue, but as one reviewer noted, possible affective reactions might be linked to cases of "weakness of will." For example, people who are addicted to nicotine might know they are damaging their health and regret this, but because of weakness of will continue to smoke. This example suggests that inhibition failure during decision making and behavior associated with weakness of will (e.g., smoking or other addictions) are related. Although it may be premature to emphasize this similarity at this time, the issue underscores the point that the decision-making field will benefit in the future from a more detailed characterization of the inhibition process per se.

People's probabilistic-thinking failures have been demonstrated in a wide variety of reasoning and decision-making tasks. We focused on base-rate-neglect problems because of the central role they play in the discussions on human rationality. Although our findings will need to be extended and generalized to different decision-making settings in future studies, we want to point to some practical and theoretical implications of our results. At the practical level, one might note that educational programs intended to improve students' decision making in risky situations (e.g., reckless driving, binge drinking, unprotected sex) have been largely ineffective (e.g., Reyna & Farley, 2006; Steinberg, 2007). Likewise, experimental studies in which people received extensive tutoring in logic and probability theory showed only a minimal impact on their performance (e.g., Kahneman, Slovic, & Tversky, 1982). In light of the present study, these results are not surprising. Intervention studies have typically been designed to alter or optimize people's knowledge. Our data indicate, however, that the problem is not a lack of statistical sophistication. People know all too well that base-rate information is relevant to their decisions. Rather, what people seem to struggle with is overriding the temptation of heuristic thinking. This suggests that interventions might be more successful if they were more specifically targeted at improving students' inhibitory capacities (e.g., Houdé, 2007).

At a more theoretical level, the evidence for successful conflict detection helps to sketch a less bleak picture of human rationality. Our findings indicate that people's thinking is more

[4]Likewise, as one might expect given that inhibition was not required for the control problems, the RLPFC did not show significant activation in these control contrasts either. Note that this finding is evidence against the claim that the RLPFC is activated by mere effort per se. Neutral control trials required more effort than congruent control trials (e.g., latencies were slightly longer, and the error rate was higher), but did not require heuristic inhibition. The absence of significant RLPFC activation in the control contrasts indicates that the RLPFC is specifically recruited for inhibitory purposes.

normative than the infamous failure to solve classic decision-making tasks suggests. If people did not know or care about the implications of sample-size considerations, for example, they would not detect conflicts between their intuitive responses and base rates. Although people might not always manage to override the temptation of heuristic thinking, they do seem to recognize when their intuitive answers are not fully warranted. Base rates are not simply neglected, and people are not merely intuitive thinkers. Our findings are in line with Sloman's (1996) and Epstein's (1994) original claims, suggesting that people go against their better judgment when they give heuristic responses. Heuristic bias points to a lack of inhibitory processing. It does not imply that people are irrational beings who lack probabilistic sophistication. In this sense, people are truly smarter than one might think.

Acknowledgments—W.D.N. is a postdoctoral fellow of the Fonds Wetenschappelijk Onderzoek-Vlaanderen. V.G. is funded by the Natural Sciences and Engineering Research Council of Canada.

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.

Barbey, A.K., & Sloman, S.A. (2007). Base-rate respect: From ecological rationality to dual processes [Target article and commentaries]. *Brain and Behavioral Sciences*, *30*, 241–298.

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

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. (2006). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science*, *17*, 428–433.

De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition*, *106*, 1245–1299.

Denes-Raj, V., & Epstein, S. (1994). Conflict between intuitive and rational processing: When people behave against their better judgement. *Journal of Personality and Social Psychology*, *66*, 819–829.

Eberhardt, J.L., Davies, P.G., Purdie-Vaughns, V.J., & Johnson, S.L. (2006). Looking deathworthy: Perceived stereotypicality of Black defendants predicts capital-sentencing outcomes. *Psychological Science*, *17*, 383–386.

Epstein, S. (1994). Integration of the cognitive and psychodynamic unconscious. *American Psychologist*, *49*, 709–724.

Epstein, S., & Pacini, R. (1999). Some basic issues regarding dual-process theories from the perspective of cognitive-experiential self-theory. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 462–482). New York: Guilford Press.

Evans, J.St.B.T. (1984). Heuristic and analytic processing in reasoning. *British Journal of Psychology*, *75*, 451–468.

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*, 321–339.

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

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.

Hoffrage, U. (2000). Why the analyses of cognitive processes matter. *Behavioral and Brain Sciences*, *23*, 679–680.

Houdé, O. (2007). First insights on "neuropedagogy of reasoning." *Thinking and Reasoning*, *13*, 81–89.

Kahneman, D. (2002). *Maps of bounded rationality: A perspective on intuitive judgement and choice*. Retrieved January 11, 2006, from http://nobelprize.org/nobel_prizes/economics/laureates/2002/kahnemann-lecture.pdf

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgement under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

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

Klein, T.A., Endrass, T., Kathman, N., Neumann, J., von Cramon, D.Y., & Ullsperger, M. (2007). Neural correlates of error awareness. *NeuroImage*, *34*, 1174–1781.

Peters, E., McCaul, K.D., Stefanek, M., & Nelson, W. (2006). A heuristic approach to understanding cancer risk perception: Contributions from judgement and decision-making research. *Annals of Behavioral Medicine*, *31*, 45–52.

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., & Farley, F. (2006). Risk and rationality in adolescent decision making: Implications for theory, practice, and public policy. *Psychological Science in the Public Interest*, *7*(1).

Ridderinkhof, K.R., Ullsperger, M., Crone, E.A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science*, *306*, 443–447.

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.

Schwartz, N., Strack, F., Hilton, D., & Naderer, G. (1991). Base-rates, representativeness, and the logic of conversation: The contextual relevance of "irrelevant" information. *Social Cognition*, *9*, 67–84.

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

Slovic, P. (2000). What does it mean to know a cumulative risk?: Adolescents' perceptions of short-term and long-term consequences of smoking. *Journal of Behavioral Decision Making*, *13*, 259–266.

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

Steinberg, L. (2007). Risk taking in adolescence: New perspectives from brain and behavior science. *Current Directions in Psychological Science*, *16*, 55–59.

van Veen, V., & Carter, C.S. (2006). Conflict and cognitive control in the brain. *Current Directions in Psychological Science*, *15*, 237–240.

(RECEIVED 6/7/07; REVISION ACCEPTED 10/23/07)

THINKING & REASONING, 2009, 15 (2), 105–128

Psychology Press
Taylor & Francis Group

The effortless nature of conflict detection during thinking

Samuel Franssens and Wim De Neys
University of Leuven, Belgium

Dual process theories conceive human thinking as an interplay between heuristic processes that operate automatically and analytic processes that demand cognitive effort. The interaction between these two types of processes is poorly understood. De Neys and Glumicic (2008) recently found that most of the time heuristic processes are successfully monitored. This monitoring, however, would not demand as many cognitive resources as the analytic thinking that is needed to solve reasoning problems. In the present study we tested the crucial assumption about the effortless nature of the monitoring process directly. Participants solved base-rate neglect problems in which heuristic and analytic processes cued a conflicting response or not. Half of the participants reasoned under a secondary task load. A surprise recall task was used as an implicit measure of whether the participants detected the conflict in the problems. Results showed that, even under load, base-rate recall performance was better for conflict problems than for no-conflict problems. Although participants made more reasoning errors under load, recall of the conflict problems was not affected by the working memory load. These findings support the claim about the successful and undemanding nature of the conflict detection process during thinking.

Keywords: Reasoning; Decision-making; Conflict monitoring; Cognitive control; Dual-task.

Human judgements are often based on intuition and prior beliefs rather than on a logical reasoning process. Sometimes this leads to bad decisions. A well-known reasoning error is the base-rate fallacy. The following

Correspondence should be addressed to Samuel Franssens, Lab Experimental Psychology, University of Leuven, Tiensestraat 102, 3000 Leuven, Belgium.
E-mail: samuel.franssens@gmail.com

http://www.psypress.com/tar DOI: 10.1080/13546780802711185

example is an adaptation of the classic base-rate neglect problem (Kahneman & Tversky, 1973):

In a study 1000 people were tested. Among the participants there were 5 sixteen-year-olds and 995 forty-year-olds. Els is a randomly chosen participant of the study.

Els likes to listen to techno and electronic music. She often wears tight sweaters and jeans. She loves to dance and has a small nose piercing.
What is most likely?
1. Els is 16 years old.
2. Els is 40 years old.

In this problem people are presented with two types of knowledge concerning a hypothetical person. First they get information about the number of members of two social groups in a sample. Then they read a description that clearly matches a stereotypical member of the smallest group. When asked which social group the person most likely belongs to, people tend to choose the group that is cued by the description. However, because the person was chosen randomly, the normative response should be the biggest group in the sample. People underuse or ignore prior probabilities when making their judgement. This reasoning flaw is called the base-rate fallacy and has been demonstrated in numerous studies (Tversky & Kahneman, 1974).

Dual process theories of thinking explain this tendency to rely on descriptive information by assuming two kinds of reasoning processes (Epstein, 1994; Evans, 2006; Kahneman, 2002; Sloman, 1996; Stanovich, 1999). A general distinction is made between *Type 1* or heuristic processes that are fast and automatic, and *Type 2* or analytic processes that are slow and controlled. Type 1 processes have high capacity, do not require much effort, and can operate in parallel. Type 2 processes, however, have limited capacity, require much cognitive effort, and operate sequentially (Evans, in press). Type 2 processes are related to individual differences in working memory and general intelligence, but Type 1 processes are not (Kokis, MacPherson, Toplak, West, & Stanovich, 2002). Faced with certain problems, the two types of processes will cue the same response. Heuristic processes then quickly deliver a correct answer, and laborious analytic reasoning becomes superfluous. However, in cases like the above example, the heuristic and the analytic response will be different: Heuristic processes will claim Els is 16 years old, but analytic processes will declare she is 40 years old. Choosing the normative answer then requires analytic processes to override the heuristic response (Stanovich, in press; Stanovich & West, 2000).

Clearly, sound reasoning and successfully overriding the heuristic response require that the output of the two processes is monitored continuously. Conflict monitoring is a key component of any dual process theory of thinking, but the process is poorly understood. De Neys and Glumicic (2008) recently started examining the process and suggested that the monitoring is quite flawless. They argued that people would have little difficulty in detecting that heuristic responses conflict with analytic considerations. In the De Neys and Glumicic study participants solved a set of classic base-rate neglect problems (Kahneman & Tversky, 1973) while thinking aloud. Verbal reports indicated that the participants almost never referred to the base-rates in the tasks. Although this finding seemed to suggest that they did not detect a conflict between base-rates and descriptions, De Neys and Glumicic noted that it is possible the detection of a conflict is not easily verbally expressed. They reasoned that, even though participants almost never report experiencing a struggle between the two options, it is possible that conflict monitoring is successful on a more implicit level. To capture such implicit detection, participants were also presented with a surprise recall task that tested the recollection of the base-rates after they had finished solving all the base-rate problems. There were more correct recall answers on questions about classic base-rate neglect problems (the description refers to a member of the smallest group) compared with questions about control problems in which there was no conflict (a member of the biggest group is described). This suggests that the participants somehow noticed the conflicting nature of these items, which prompted them to further scrutinise them, yielding a better memory for the group sizes. Even the participants who failed to correctly solve any of the classic base-rate neglect problems displayed this enhanced memory for base-rates of classic problems. Other implicit measures such as problem-processing times and visual inspection times also indicated that these items were inspected more thoroughly.

De Neys and Glumicic (2008) therefore characterised conflict monitoring as a successful implicit process. In the present study we further explore the characteristics of the monitoring process by focusing on its processing requirements. Given the apparent omnipresence of successful monitoring, De Neys and Glumicic assumed that the process would be undemanding. This assumption helped to explain why even the least gifted reasoners showed almost flawless performance on the conflict-monitoring measure: People's limited cognitive resources might not always suffice to complete the reasoning process, but they would be sufficient for the less demanding monitoring. However, the crucial hypothesis that the monitoring process does not require much cognitive resources is only a post-hoc argumentation. The present research attempts to test this hypothesis directly. Dual process theories state that analytic thinking is slow and resource demanding. A cognitive load that burdens central processing resources should therefore lead to a performance drop on reasoning problems that require analytic thinking. However, if the monitoring process itself indeed demands only a minimal amount of resources, conflict detection should be largely unaffected by such a load.

In the present study participants solved three types of base-rate neglect problems. In the *incongruent* problems the hypothetical person was portrayed as a stereotypical member of the smallest social group in the sample (i.e., base-rates and description conflicted). In the *congruent* problems the base-rates and the description pointed towards the same answer. These congruent problems could be answered correctly by taking the heuristic route. In other problems the description was not stereotypical (*neutral* problems). Because there were no heuristic cues in these items, we expected people to reason analytically and focus on the base-rates. A high number of correct responses was expected for the congruent and the neutral items. For the incongruent problems, however, the prediction from the literature (Kahneman & Tversky, 1973) was that people would follow their intuition and err in the majority of cases.

Half of the participants solved the base-rate neglect problems under a secondary task load. Dual process theories assume that correct answers on incongruent and neutral items depend on the activation of resource-demanding analytic processes. Congruent items, however, can be solved by heuristic processes, which operate automatically. Therefore the expectation was that the secondary task would lead to a drop in reasoning performance for incongruent and neutral items, but not for congruent items.

The second part of the experiment was a surprise recall task. The participants had to answer questions about each of the problems they had just solved. There were easy questions about the descriptions and a crucial question about the base-rates in each problem. As in De Neys and Glumicic's (2008) study, successful conflict detection in the judgement task should result in a better memory for base-rates of items in which there was a conflict (incongruent items), compared with items in which there was no conflict (congruent items), because of a deeper processing of these items.

The aim of the load manipulation was to limit the cognitive resources the participants could allocate to processing information from the reasoning problems. Therefore an overall poorer recall was expected for participants who performed the judgement task under load. However, if the hypothesis about the effortless nature of the conflict-monitoring process is correct, participants who made their judgements under load should still have managed to detect the conflicts in the reasoning problems. This should have resulted in a better processing of base-rates for items in which there was a conflict. Consequently, recall should be better for incongruent items than for congruent items, whether or not the participants reasoned under load.

METHOD

Participants

A total of 74 second-year psychology students at the University of Leuven (Belgium) participated in return for course credit.

Pilot study

We created 53 stereotypical and neutral personality descriptions. Eight second-year psychology students (who did not participate in the main study) then judged how likely it was that the described person belonged to each of two social groups. The raters used an 11-point scale ranging from 0, extremely unlikely, to 10, extremely likely. After they had finished rating all 53 items, the students had to answer two questions about the descriptions in these items. These were multiple-choice questions with four options. From this we selected two sets of four stereotypical items (as explained below, both sets functioned as congruent items as well as incongruent items in the judgement task) and one set of four neutral items. The descriptions in the stereotypical items moderately but consistently cued one of two groups; the descriptions in the neutral items were rated as equally likely to refer to each group. Mean ratings for the descriptions in the two sets of stereotypical items were 7.96 ($SD = 0.64$) and 8.03 ($SD = 0.41$) for the most likely group and 3 ($SD = 0.61$) and 3.21 ($SD = 1.06$) for the least likely group. For the neutral items the mean ratings were 6.71 ($SD = 0.85$) for the most likely group and 6.06 ($SD = 0.96$) for the least likely group. Performance on the recall task was the same for both sets of stereotypical items, with an average of 6.83 out of 8 questions correct for each set (2 questions for each of 4 items). Recall was a bit lower for the neutral items (5.83 out of 8 questions correct). This result was expected because the descriptions in the neutral items by definition did not fit any stereotype, which made them less salient and harder to encode.

Materials

In the judgement task four stereotypical items served as incongruent problems. The other four served as the congruent items. However, a second version of the task was created, in which the base-rates of all items were switched. The incongruent items in the first version now became the congruent items, and vice versa. In the actual experiment half of the participants solved the first version and the other half solved the second version. Results for each item type were averaged over the two versions. Together with the pilot study, this extra control ruled out the possibility that differences in performance between incongruent and congruent items could be explained in terms of differences in item stereotypicality.

The different problems were presented with slightly varied base-rates. More precisely, for each item type two problems were presented with a 995/5, one with a 994/6, and one with a 996/4 base-rate ratio. This variation was included to make the subsequent recall task more engaging. De Neys and Glumicic (2008) have already shown that there is no difference in performance for the three base-rate levels.

Note that we specifically opted to use base-rates that were extreme and descriptions that were moderate. This was done to ensure that the items would evoke a conflict between heuristic and analytic processes. Items with moderate base-rates or extreme descriptions would leave room for discussion about which answer is correct, if the participants adopt a Bayesian approach to solve the items (Gigerenzer, Hell, & Blank, 1988). In order to calculate the probability that the person described belongs to one or the other social group, information in the description will then be combined with the information from the base-rates. This leads to difficulties when the base-rates are too moderate (e.g., when a sample consists of 30 men and 70 women, and the person described is a football fan. Saying that the person is a man cannot unequivocally be considered as wrong) or the descriptions are too extreme (e.g., when a sample consists of 995 women and 5 men, and the person described is the leader of the Catholic Church, i.e., the Pope. No matter what the base-rates are, the person can never be a woman). The combination of extreme base-rates with moderately stereotypical descriptions guaranteed that the response cued by the base-rates was the normatively correct one. Note that we do not wish to associate heuristic processes with wrong answers and analytic processes with right answers. Evans (2006, 2007a, 2007b) has recently argued that analytic processes can be responsible for cognitive biases as much as heuristic processes. We acknowledge that, to elicit a conflict between heuristic and analytic processes, it was not strictly necessary to include a correct answer in the items. However, without the extreme base-rates and the moderate descriptions, it was possible that both types of processes pointed towards the same answer if the participants made their decisions in the aforementioned Bayesian manner. In that case there would not be a conflict between heuristic and analytic processes. Previous work with similar items (De Neys & Glumicic, 2008; De Neys, Vartanian, & Goel, 2008) showed that the extreme base-rates do not boost reasoning performance. As in the classic study of Kahneman and Tversky (1973), people select the heuristic response on the vast majority of the conflict problems.

The order of the two response options (1 and 2) was counterbalanced. For half of the problems the correct response (i.e., the response consistent with the base-rates) was option 1, for the other half the second option

(2) was the correct one. The problems were based on a variety of stereotypes related to gender, race, age, and lifestyle. The following are examples of the three problem types (an overview of all items can be found in Appendix 1):

Incongruent:
In a study 1000 people were tested. Among the participants there were 994 Swedes and 6 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 womaniser. His favourite dish is the spaghetti his mother makes.
What is most likely?
1. Mario is a Swede.
2. Mario is an Italian.

Congruent:
In a study 1000 people were tested. Among the participants there were 5 women and 995 men. Dominique is a randomly chosen participant of the study.

Dominique is 32 years old and is a 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?
1. Dominique is a woman.
2. Dominique is a man.

Neutral:
In a study 1000 people were tested. Among the participants there were 994 people from Antwerp and 6 people from Amsterdam. Bart is a randomly chosen participant of the study.

Bart is sixteen years old and still goes to school. He weighs 80 kilos and has a little sister of 14 years old and an older brother who has already been attending university for two years.
What is most likely?
1. Bart is from Antwerp.
2. Bart is from Amsterdam.

Procedure

The experiment was run on a computer and started with a welcoming screen. The participants received information about the content of the reasoning problems, but any hints on how to solve the problems were avoided:

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.*

Judgement task. The 12 base-rate neglect problems were presented in random order. In each problem the participants first saw a screen with the following introductory information: *In a study 1000 people were tested. Els* [or any of the 12 hypothetical persons] *is a randomly chosen participant of the study.*

After reading this they had to press the space bar to go to the next screen, on which the first base-rate neglect problem appeared:

In the study there were 5 sixteen-year-olds and 995 forty-year olds.

Els likes to listen to techno and electronical music. She often wears tight sweaters and jeans. She loves to dance and has a small nose piercing.
What is most likely?
1. Els is 16 years old.
2. Els is 40 years old.

Judgements had to be made by pressing the "1" or "2" key. The participants then proceeded with the next problem. After they had finished solving all 12 problems the recall task was introduced.

In the *load condition* the participants had to make their judgements while performing a visuospatial secondary task (De Neys, 2006a; 2006b; De Neys & Verschueren, 2006; Verschueren, Schaeken & d'Ydewalle, 2004). After the screen with the introductory information about the base-rate neglect problem, a pattern of four dots in a three-by-three grid (see Figure 1 for an example) was presented for 900 milliseconds. The participants were told that they had to reproduce this pattern after solving the base-rate neglect problem that followed. Upon completion of the base-rate neglect problem, they saw an empty grid that they could fill up with dots by clicking in the grid. It was emphasised that it was important to get the pattern right every time. Previous research has established that this dot memory task efficiently taps executive processing resources (Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001).

Recall task. The questions were presented in a fixed order, with questions about the same problem grouped together. Two questions about the descriptions were followed by two questions concerning the sizes of both groups. The questions about the descriptions were included to make the task less repetitive and more engaging; however, our main interest was participants' recollection of the base-rates. Details about the problem at hand (name of the person, social groups involved) remained visible on the screen: *Among the problems you just solved there was one about Els whose*

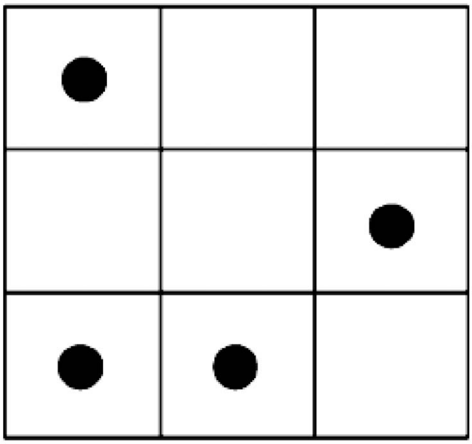

Figure 1. Example of a dot pattern in the visuo-spatial task.

description was randomly chosen from a study in which there were sixteen-year-olds and forty-year-olds.

When questioned about the description, the participants saw four statements about the person and were asked to indicate which one was correct (an overview of all questions can be found in Appendix 2):

Which of the following statements is correct?
1. *Els likes to listen to classical music.*
2. *Els likes to listen to Studio Brussel.* [a popular radio station in Belgium]
3. *Els likes to listen to techno music.*
4. *Els drinks two litres of water every day.*

On the next screen, the participants had to type in the sizes of the two social groups.

Exactly how many sixteen-year-olds were there in the study?
Answer:
Exactly how many forty-year-olds were there in the study?
Answer:

RESULTS AND DISCUSSION

Load task

On average the participants filled in 3.48 ($SD = 0.38$) dots out of 4 correctly, which indicates that the secondary task was properly performed. Scores on the load task did not correlate with the judgement task ($r = .103$, $p = .58$) nor with the recall task ($r = -.186$, $p = .31$), which means there was no trade-off between the load task and judgement or recall.

Judgement task: Accuracy

As in the classic studies of Tversky and Kahneman (1974), the participants selected the heuristic answer on the majority of the incongruent items. However, accuracy was high on congruent and neutral problems. The difference in accuracy between the problem types was significant, $F(2, 84) = 58.90$, $p < .001$, $\eta_p^2 = .58$. As Figure 2 shows, the load manipulation only affected performance on items that required analytic reasoning (incongruent and neutral items), $F(1, 72) = 4.18$, $p = .04$, $\eta_p^2 = .05$. When heuristic reasoning sufficed (congruent items), there was no effect of load, $F(1, 72) < 1$. This result supports the crucial assumption from dual process theories that analytic processes demand cognitive resources, but heuristic processes operate automatically.

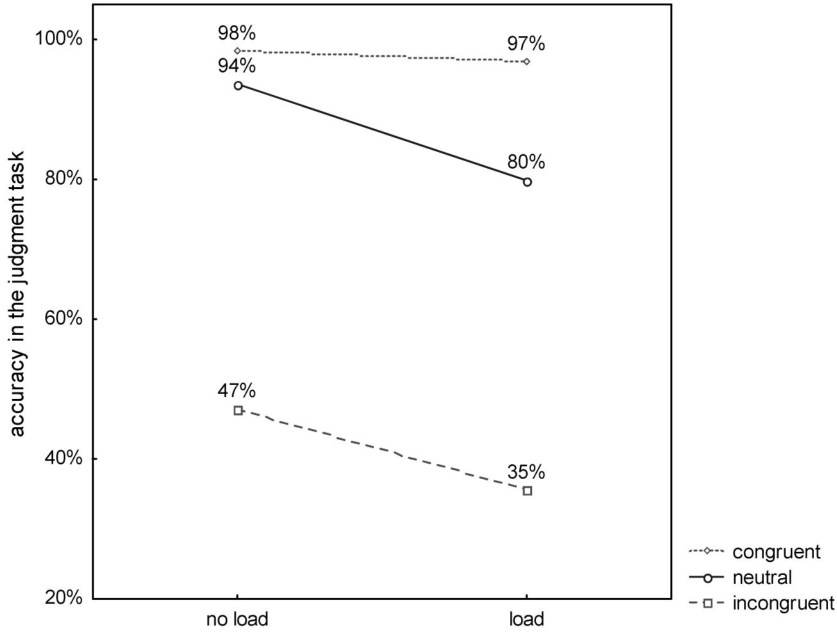

Figure 2. Mean proportion of correctly solved problems in the judgement task.

Judgement task: Response times

There was a main effect of item type on response time, $F(2, 84) = 11.98$, $p < .001$, $\eta_p^2 = .22$ (see Table 1). More specifically, items that required analytic thinking (incongruent and neutral items), took longer to solve than items for which heuristic thinking (congruent items) sufficed, $F(1, 42) = 30.77$, $p < .001$, $\eta_p^2 = .42$. This result is consistent with the dual process theory assumption that analytic processing is more time consuming than heuristic processing. There was no significant effect of load on incongruent and neutral items nor on congruent items, all $Fs(1, 72) < 1$.

Recall task

Accuracy on the two questions about the descriptions was aggregated. Answers on questions about the base-rates were coded as correct when the participants recalled which one of the two groups was the largest (i.e., when the order of magnitude of the base-rates was correctly recalled).[1] Accuracy on the two questions about the base-rates was also aggregated.

Figure 3 presents an overview of the results. First we focus on the data from the participants in the no-load condition. The effect of item type on the recall of the descriptions was significant, $F(2, 84) = 11.22$, $p < .001$, $\eta_p^2 = .21$. Recall of descriptions was lower for neutral items than for incongruent and congruent items, $F(1, 42) = 18.66$, $p < .001$, $\eta_p^2 = .30$. This was expected, because the descriptions in these neutral items were less salient than the descriptions in the other items. There was no difference in recall performance for incongruent and congruent items, $F(1, 42) < 1$.

Our main interest was the performance on questions about the base-rates. The effect of item type was significant, $F(2, 84) = 5.29$, $p = .006$, $\eta_p^2 = .11$.

TABLE 1
Mean response times in seconds (*SD*)

	No-load	*Load*
Congruent	13.8 (3.5)	15.2 (5.3)
Neutral	16.5 (4.2)	17.1 (4.2)
Incongruent	16.5 (4.8)	17.4 (6.5)

[1]If a number higher than 500 was given for the largest population group in the problem and a number lower than 500 for the smallest group, the answer was coded as correct. On the vast majority of trials (95% +) people answered with the base-rates actually presented in the task (i.e., 995/5, 994/6, 996/4). Errors consisted of switching the base-rates around.

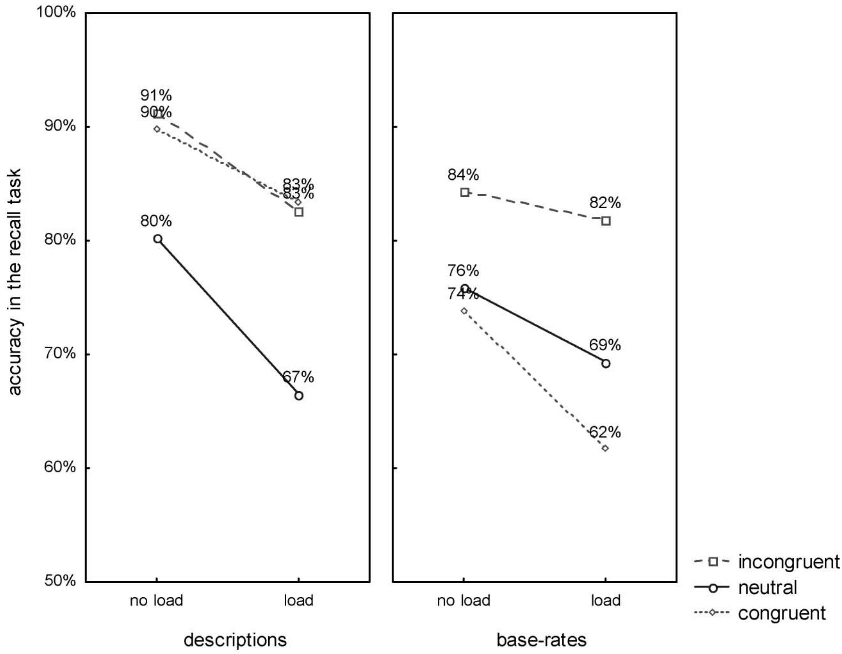

Figure 3. Mean proportion of correctly answered questions in the recall task.

As Figure 3 shows, recall performance was higher for incongruent items than for congruent items, while neutral items fell in between. The crucial difference between the incongruent and congruent items was significant, $F(1, 42) = 7.72$, $p = .008$, $n_p^2 = .15$. This means that incongruent items were processed better than congruent items. The only difference between incongruent items and congruent items was the presence of a conflict between base-rates and description. Therefore, the participants must have detected this conflict, which urged them to further inspect these items, and which in turn resulted in a better recall of these items. This finding is in line with the results from De Neys and Glumicic (2008).

The data from the load condition show that the overall recall of descriptions was lower than in the no-load condition, $F(1, 72) = 9.6$, $p = .002$, $\eta_p^2 = .11$, but we find the same pattern of results for the different item types. Again, recall was lower for neutral items than for incongruent and congruent items, $F(1, 30) = 28.65$, $p < .001$, $\eta_p^2 = .48$, and there was no difference between incongruent and congruent items, $F(1, 30) < 1$.

The crucial question in this experiment was whether the load manipulation could prevent the participants from successfully detecting the conflicts in the judgement task. If conflict detection requires many cognitive resources it should no longer be successful under load, and base-rate recall performance should no longer differ for items in which there was a conflict and items in which there was no conflict. However, if conflict detection operates automatically then it should also be successful under load, and recall of conflict items should be better than recall of no-conflict items. As Figure 3 shows, even under load, base-rates of incongruent items were better recalled than base-rates of congruent items, $F(1, 30) = 9.94$, $p = .003$, $n_p^2 = .24$. Recall of neutral items fell in between that of incongruent and congruent items. Furthermore, recall of base-rates of incongruent items did not decrease under load, $F(1, 72) < 1$. However, recall performance for congruent items was lower in the load condition than in the no-load condition, $F(1, 72) = 4.19$, $p = .04$, $n_p^2 = .05$. Recall for neutral items was not significantly affected $F(1, 72) = 1.20$, $p = .27$, $n_p^2 = .01$. The load manipulation thus resulted in a lower performance on items in which there was no conflict, but when there was a conflict (or when the description did not cue a response), recall performance was unaffected. This finding is further evidence for the hypothesis that conflict detection is an effortless process. Although analytic reasoning performance deteriorated because of the load manipulation, the conflict-monitoring process itself was unaffected by this load.

We argued that the absence of a load effect on our measure of conflict detection, base-rate recall performance, indicated that the monitoring process operates rather automatically. In theory, this lack of a load effect on recall performance could, however, also be attributed to a trade-off between processing time and conflict detection accuracy. Indeed, even if conflict detection were demanding, people might still be successful at it under load by taking additional time to complete the monitoring process. Because people would take more time to make a judgement, recall might also benefit from this extra processing time and remain unaffected. Hence, successful detection and better recall would be bought at the cost of extra processing time. However, such a trade-off implies, that an increase in response times under load should be substantially larger for incongruent problems than for congruent items. The latencies (see Table 1) clearly indicate that this was not the case. The increase in response times was not significant under load and even tended to be somewhat more pronounced for congruent than for incongruent items. This alternative interpretation in terms of a trade-off between secondary task and processing times can therefore be discarded.

GENERAL DISCUSSION

The present study tested the hypothesis that the conflict-monitoring process uses only a minimal amount of cognitive resources and operates quite automatically. In our study half of the participants had to reason under a load that burdened their cognitive resources. Afterwards they were questioned about the base-rates in these problems. Results from the judgement task showed that the load task affected performance on problems that required analytic reasoning, but it did not affect performance on problems for which heuristic reasoning sufficed. This supports the key assumption from dual process theories that analytic thinking is resource demanding, whereas heuristic thinking is not. Results from the recall task showed that participants' memory was better for base-rates of problems in which there was a conflict, than for problems in which there was no conflict. This result replicates the finding from De Neys and Glumicic (2008) concerning the successful nature of the conflict-monitoring process. Our main interest, however, was the recall performance from the participants who had to make their judgements under load. If the conflict-monitoring process does indeed require only a minimal amount of resources, it should not be affected by a cognitive load. This is exactly what we observed. In the load condition base-rate recall was still better for conflict problems than for no-conflict problems. Furthermore, recall performance on conflict items was equally high in the load condition as in the no-load condition. Even though the load task impeded the analytic thinking that was needed to solve the reasoning problems in the judgement task, it did not prevent the participants from detecting the conflict in these reasoning problems. This supports the hypothesis that conflict monitoring during judgement is an effortless process.

The finding of conflict detection as a successful and effortless process contributes to a much-needed specification of the processing characteristics of dual process theories. De Neys and Glumicic (2008) noted that conflict detection as a successful process implies that, whenever people are confronted with a reasoning problem, heuristic processes are accompanied by at least some minimal analytic activation. Other authors defend this same position (Evans, 2006, 2007b; Kahneman & Frederick, 2002; Stanovich, in press). The idea of an effortless monitoring process can also be found in other researchers' work. Thompson (in press) recently suggested that heuristic responses to reasoning problems are accompanied by a certain *feeling of rightness*, an intuition that the answer is correct. The strength of this feeling of rightness determines whether or not analytic processes rethink the heuristic response. A low feeling of rightness triggers analytic intervention. This process might be linked to conflict monitoring. Thompson's characterisation of this process also seems to imply a non-demanding, automatic nature. Evans (in press) recently made an interesting distinction between two kinds of Type 1 processes. *Preattentive* processes, on the one hand, provide working memory with content and thus determine what information analytic processes are working with. *Autonomous* processes, on the other hand, bypass working memory and "can control behaviour directly without need for any kind of controlled attention". Evans argued that the conflict between autonomous Type 1 processes and Type 2 processes is settled by preconscious *Type 3* processes. Analytic processes cannot both monitor the output of heuristic processes and at the same time also engage in demanding computations to work out their own response. In this sense dual process theories have traditionally overburdened analytic processes with the responsibility of monitoring for conflict. The recent views of De Neys and Glumicic, Evans, and Thompson seem to share the basic assumption that processes that monitor for the need of analytic intervention operate in a rather automatic way. The crucial contribution of the present paper is that it supports this assumption with empirical findings.

In this study we presented some of the first direct evidence for the automatic nature of the monitoring process. Caution is needed, however, when interpreting the present results. For example, the results come from a single experiment with one specific judgement task (base-rate neglect problems with extreme base-rates and moderate descriptions). It is clear that the findings need replication with other tasks. One could also question whether the findings are also valid for the population at large. For younger or less-gifted populations (e.g., children or people with specific brain damage), monitoring might be far from effortless. Furthermore, it should be clear that the present results do not necessarily imply that the conflict-monitoring process operates in a completely automatic manner (i.e., that it does not require *any* cognitive resources at all). The point is that the monitoring process was unaffected by a secondary task that did have an effect on analytic reasoning. It is therefore safe to say that the monitoring process is far less demanding than the analytic reasoning process that is needed to complete the reasoning problem.

If we are right about the successful and effortless nature of the conflict-monitoring process, one could wonder why people still make reasoning errors. De Neys and Glumicic (2008) have already suggested that reasoning errors need to be attributed to a failure to inhibit heuristic responses rather than to a monitoring failure (see also Evans, 2007a; Houdé & Moutier, 1996; Stanovich, in press). Although people might detect that the heuristic response conflicts with the analytically appropriate response, they will not always manage to inhibit the tempting heuristic. A recent neuroimaging study by De Neys et al. (2008) presented evidence for this claim. In their study, participants solved base-rate neglect problems while the activation of brain regions believed to be involved in conflict detection (anterior cingulate cortex) and response inhibition (lateral prefrontal cortex) was monitored. Scanning results showed that the conflict detection area was always activated when the participants solved incongruent problems, but this activation was unrelated to the performance on these problems. This finding confirms the hypothesis that conflict detection is successful most of the time.

The inhibition area, however, was only activated when conflict problems were solved correctly. Taken together, this line of research indicates that reasoning errors are the result of a failure to inhibit heuristic processes rather than the result of a failure to detect the erroneous nature of the heuristic response in itself. The present study clarifies that this monitoring for heuristic–analytic conflict during judgement is a successful process that operates rather automatically.

Manuscript received 4 October 2007
Revised manuscript received 26 November 2008
First published online 21 February 2009

REFERENCES

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., & Glumicic, T. (2008). Conflict monitoring in dual process theories of reasoning. *Cognition*, *106*, 1248–1299.

De Neys, W., Vartanian, O., & Goel, V. (2008). Smarter than we think: When our brain detects 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.

Epstein, S. (1994). Integration of the cognitive and psychodynamic unconscious. *American Psychologist*, *49*, 709–724.

Evans, J. St. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin and Review*, *13*, 378–395.

Evans, J. St. B. T. (2007a). On the resolution of conflict in dual process theories of reasoning. *Thinking & Reasoning*, *13*, 321–329.

Evans, J. St. B. T. (2007b). *Hypothetical thinking: Dual processes in reasoning and judgement*. Hove, UK: Psychology Press.

Evans, J. St. B. T. (in press). How many dual process theories do we need: One, two or many? In J. St. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond*. Oxford, UK: Oxford University Press.

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.

Houdé, O., & Moutier, S. (1996). Deductive reasoning and experimental inhibition training: The case of the matching bias. *Current Psychology of Cognition*, *15*, 409–434.

Kahneman, D. (2002, December). *Maps of bounded rationality: A perspective on intuitive judgement and choice*. *Nobel Prize Lecture*. Retrieved 30 April 2007 from: http://nobelprize.org/nobel_prizes/economics/laureates/2002/kahnemann-lecture.pdf

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgement. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics & biases: The psychology of intuitive judgment* (pp. 49–81). New York: Cambridge University Press.

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

Kokis, J. V., MacPherson, R., Toplak, M. E., West, R. F., & Stanovich, K. E. (2002). Heuristic and analytic processing: Age trends and associations with cognitive ability and cognitive styles. *Journal of Experimental Child Psychology*, *83*, 26–52.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, *130*, 621–640.

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

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Stanovich, K. E. (in press). Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory? In J. St. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond*. Oxford, UK: Oxford University Press.

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

Thompson, V. (in press). Dual process theories: A metacognitive perspective. In J. St. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond*. Oxford, UK: Oxford University Press.

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.

Verschueren, N., Schaeken, W., & d'Ydewalle, G. (2004). Everyday conditional reasoning with working memory preload. In *Proceedings of the Twenty-Sixth Annual Meeting of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

APPENDIX 1

The 12 base-rate neglect problems in the first version of the task are presented below (translated from Dutch). In the second version all base-rates were switched. Incongruent problems in version one became congruent problems in version two, and vice versa. For example, the reasoning problem featuring Els became the following in version two:

In a study 1000 people were tested. Among the participants there were 995 sixteen-year-olds and 5 forty-year-olds. Els is a randomly chosen participant of the study.

Els likes to listen to techno and electronic music. She often wears tight sweaters and jeans. She loves to dance and has a small nose piercing.
What is most likely?
1. Els is 16 years old.
2. Els is 40 years old.

Base-rates and descriptions no longer conflict, this is now a congruent problem.

Incongruent problems

Els:

In a study 1000 people were tested. Among the participants there were 5 sixteen-year-olds and 995 forty-year-olds. Els is a randomly chosen participant of the study.

Els likes to listen to techno and electronic music. She often wears tight sweaters and jeans. She loves to dance and has a small nose piercing.
What is most likely?
1. Els is 16 years old.
2. Els is 40 years old.

Etienne:
In a study 1000 people were tested. Among the participants there were 4 people who drive a second-hand Nissan and 996 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 on the outskirts of Charleroi. His wife has left him.
What is most likely?
1. Etienne drives a BMW.
2. Etienne drives a second-hand Nissan.

Mario:
In a study 1000 people were tested. Among the participants there were 994 Swedes and 6 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 womaniser. His favourite dish is the spaghetti his mother makes.
What is most likely?
1. Mario is a Swede.
2. Mario is an Italian.

Sarah:
In a study 1000 people were tested. Among the participants there were 995 Muslims and 5 Buddhists. Sarah is a randomly chosen participant of the study.

Sarah is 19 years old. She likes to philosophise and has an aversion to materialism. She wears second-hand clothes and would love to go to India one day.
What is most likely?
1. Sarah is a Buddhist.
2. Sarah is a Muslim.

Congruent problems

Aline:

In a study 1000 people were tested. Among the participants there were 996 people who like to watch Canvas and 4 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?
> 1. Aline likes to watch Canvas.
> 2. Aline likes to watch VTM.

Debby:

In a study 1000 people were tested. Among the participants there were 994 people who live in the country and 6 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?
> 1. Debby lives in the country.
> 2. Debby lives in the city.

Jeanine:

In a study 1000 people were tested. Among the participants there were 5 people who vote for Groen! and 995 people who vote for Vlaams Belang. Jeanine is a randomly chosen participant of the study.

> Jeanine is 67 years old. She's always worked on the assembly line. She attaches great importance to traditional values and lives in a residential area where there's a lot of crime.
> What is most likely?
> 1. Jeanine votes for Vlaams Belang.
> 2. Jeanine votes for Groen!

Dominique:

In a study 1000 people were tested. Among the participants there were 5 women and 995 men. Dominique is a randomly chosen participant of the study.

> Dominique is 32 years old and is a 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?
> 1. Dominique is a woman.
> 2. Dominique is a man.

Neutral problems

Hugo:

In a study 1000 people were tested. Among the participants there were 995 fifty-year-olds and 5 sixty-year-olds. Hugo is a randomly chosen participant of the study.

> Hugo is very curious about new cultures. He likes to try food from other countries. He just returned from a vacation in Hungary.
> What is most likely?
> 1. Hugo is sixty years old.
> 2. Hugo is fifty years old.

Jan:

In a study 1000 people were tested. Among the participants there were 4 people who play the saxophone and 996 people who play drums. Jan is a randomly chosen participant of the study.

> Jan is 19 years old. He studies in Brussels and doesn't have a girlfriend. He's just bought an old, second-hand car with the money he saved.
> What is most likely?
> 1. Jan plays the saxophone.
> 2. Jan plays drums.

Pablo:

In a study 1000 people were tested. Among the participants there were 5 Club Brugge fans and 995 Anderlecht fans. Pablo is a randomly chosen participant of the study.

> Pablo is 39 years old. He's a dedicated football fan. His week starts off badly when his team loses. He takes his son to watch every home game.
> What is most likely?
> 1. Pablo is an Anderlecht fan.
> 2. Pablo is a Club Brugge fan.

Bart:

In a study 1000 people were tested. Among the participants there were 994 people from Antwerp and 6 people from Amsterdam. Bart is a randomly chosen participant of the study.

> Bart is sixteen years old and still goes to school. He weighs 80 kilos and has a little sister of 14 years old and an older brother who has already been attending university for two years.
> What is most likely?
> 1. Bart is from Antwerp.
> 2. Bart is from Amsterdam.

APPENDIX 2

There were two questions about the descriptions from each problem. Details about the problem at hand (name of the person, social groups involved) remained visible on the screen. For example: *Among the problems you just solved there was one about Els whose description was randomly chosen from a study in which there were sixteen-year-olds and forty-year-olds.*

The participants saw four statements about the person, and were asked to indicate which one was correct ("*Which of the following statements is correct?*"). The statements were numbered, and the participants had to press the number that corresponded with their answer. The response options for every question are presented below.

Els:
Question 1
1. Els likes to listen to classical music
2. Els likes to listen to Studio Brussels
3. Els likes to listen to techno
4. Els drinks two litres of water every day

Question 2
1. Els has a birth mark on her belly
2. Els has a nose piercing
3. Els has a tattoo
4. Els wears a diamond necklace

Etienne:
Question 1
1. Etienne lives in Liège
2. Etienne lives in Charleroi
3. Etienne lives in Hasselt
4. Etienne lives in Knokke

Question 2
1. Etienne works at a lawyer's office
2. Etienne works at a steel plant
3. Etienne works for the government
4. Etienne works for a building industry

Mario:
Question 1
1. Mario is a womaniser
2. Mario has been in a relationship for two years
3. Mario is very punctual

4. Mario gives a lot of compliments

Question 2

1. Mario likes pizza
2. Mario likes spaghetti
3. Mario likes fish
4. Mario likes readings novels

Sarah:

Question 1

1. Sarah likes to philosophise
2. Sarah lives with her parents
3. Sarah is non-violent
4. Sarah follows the rules of her religion

Question 2

1. Sarah wants to go to Thailand
2. Sarah wants to go to Saudi Arabia
3. Sarah wants to go to Spain
4. Sarah wants to go to India

Aline:

Question 1

1. Aline is a cook
2. Aline writes reviews
3. Aline is a cleaning lady
4. Aline is a doctor

Question 2

1. Aline loves her husband very much
2. Aline likes to paint
3. Aline gossips a lot
4. Aline speaks French

Debby:

Question 1

1. Debby often plays squash
2. Debby rides a horse
3. Debby wants to become a vet
4. Debby is a fan of the local football team

Question 2

1. Debby knows how to milk a cow
2. Debby loves babies
3. Debby has animals at home
4. Debby doesn't have room for animals

Jeanine:

Question 1

1. Jeanine is a housewife
2. Jeanine works on the assembly line
3. Jeanine is a psychologist
4. Jeanine is a cashier

Question 2

1. Jeanine lives in a home for the elderly
2. Jeanine shares an apartment with friends
3. Jeanine lives in a residential area
4. Jeanine lives in Oostende

Dominique:

Question 1

1. Dominique is emotional
2. Dominique is aggressive
3. Dominique is self-assured
4. Dominique is careless

Question 2

1. Dominique is not overweight
2. Dominique is well muscled
3. Dominique recently had an accident
4. Dominique is worried about his or her weight

Hugo:

Question 1

1. Hugo likes to try food from other countries
2. Hugo is a chef in a Chinese restaurant
3. Hugo wears traditional African clothes
4. Hugo is about to move to Kenya

Question 2

1. Hugo is unreliable
2. Hugo is a miser
3. Hugo is very funny
4. Hugo is very curious about new things

Jan:

Question 1

1. Jan is married
2. Jan is divorced
3. Jan doesn’t have a girlfriend
4. Jan lives with his girlfriend

Question 2
1. Jan doesn't have a car
2. Jan drives a second-hand car
3. Jan drives a mini-van
4. Jan drives a grey car

Pablo:
Question 1
1. Pablo goes to watch every home game of his team
2. Pablo only goes to watch his team when they play on location
3. Pablo only watches football on TV
4. Pablo watches every game of his team

Question 2
1. Pablo takes his cousin to watch football games
2. Pablo takes his dad to watch football games
3. Pablo watches football games with his friends
4. Pablo takes his son to watch football games

Bart:
Question 1
1. Bart works in a video store
2. Bart studied law
3. Bart is unemployed
4. Bart still goes to school

Question 2
1. Bart has a brother and a sister
2. Bart is an only child
3. Bart has two brothers
4. Bart has one sister

Cognitive, Affective, & Behavioral Neuroscience
2010, 10 (2), 208-216
doi:10.3758/CABN.10.2.208

Feeling we're biased: Autonomic arousal and reasoning conflict

Wim De Neys
CNRS and University of Toulouse, Toulouse, France

and

Elke Moyens and Debora Vansteenwegen
University of Leuven, Leuven, Belgium

Human reasoning is often biased by intuitive beliefs. A key question is whether the bias results from a failure to detect that the intuitions conflict with logical considerations or from a failure to discard these tempting intuitions. The present study addressed this unresolved debate by focusing on conflict-related autonomic nervous system modulation during biased reasoning. Participants' skin conductance responses (SCRs) were monitored while they solved classic syllogisms in which a cued intuitive response could be inconsistent or consistent with the logical correct response. Results indicated that all reasoners showed increased SCRs when solving the inconsistent conflict problems. Experiment 2 validated that this autonomic arousal boost was absent when people were not engaged in an active reasoning task. The presence of a clear autonomic conflict response during reasoning lends credence to the idea that reasoners have a "gut" feeling that signals that their intuitive response is not logically warranted. Supplemental materials for this article may be downloaded from http://cabn.psychonomic-journals.org/content/supplemental.

In the spring of 2009, fears of the H1N1 virus swept the world. The media commonly referred to the new virus as "swine" or "Mexican" flu although it was no longer harbored in swine and had already spread over the world at the time of the outbreak; hence, eating pork or having dinner at your local Mexican restaurant did not pose any clear health risks. The World Health Organization tried hard to inform the public, but the mere intuitive association with the name of the virus seemed to have an irresistible pull on people's behavior: A lot of us stopped eating at Mexican restaurants, Haitian officials rejected an aid ship with Mexican food aid, pork belly futures collapsed on Wall Street, and the Egyptian government even ordered their farmers to kill all of their pigs (Alexander, 2009; Ballantyne, 2009). From a logical point of view, none of these measures was effective to stop the spread of the virus or avoid contamination, but, intuitively, people nevertheless felt they were better off by simply avoiding contact with Mexicans or pork.

People's overreaction to the swine flu threat is a dramatic illustration of a general human tendency to base our judgment on fast intuitive impressions rather than on more demanding, deliberative reasoning. This tendency is biasing people's performance in a wide range of classic logical and probabilistic reasoning tasks (Evans, 2003; Kahneman, 2002). One of the most famous and studied examples is the belief bias phenomenon in syllogistic reasoning. Belief bias refers to the intuitive tendency to judge the validity of a syllogism by evaluating the believability of the conclusion (Oakhill, Johnson-Laird, & Garnham, 1989). Often this is problematic, because the believability of the conclusion conflicts with its logical status. Consider the following example: "All birds have wings. Crows have wings. Therefore, crows are birds." 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 people abandon this mere intuitive, or so-called "heuristic," thinking, and engage in more deliberate, analytic thinking. Unfortunately, this turns out to be quite hard for most people; just as in the swine flu case, many reasoners end up being biased by their intuition.

Although it is a well-established fact that people are often biased, the nature of this bias is unclear. The crucial issue boils down to whether or not people detect that they are biased. Sound reasoning requires that people monitor their intuitions for conflict with more logical considerations. According to one view, people would be very bad at this monitoring (e.g., Kahneman & Frederick, 2005). Because of lax monitoring, people would simply not detect that their intuitions are invalid. However, others have argued that there is nothing wrong with the detection process (e.g., Epstein, 1994; Houdé, 2007; Sloman, 1996). They claim that people have little trouble detecting that

W. De Neys, wim.deneys@univ-tlse2.fr

their intuitions are not fully warranted; the problem, according to this view, is that these intuitions are so tempting that people fail to discard them.

Clarifying the efficiency of the conflict detection process and the resulting nature of the heuristic bias is crucial for the study of human thinking. Recently, De Neys, Vartanian, and Goel (2008) tried to decide between the alternative views by monitoring the activation of the anterior cingulate cortex (ACC), a neural region associated with conflict detection, during reasoning. They observed that the neural conflict region was activated when people gave biased responses. This finding provided some preliminary support for the idea that people detect that they are biased. However, settling the debate requires further validation and characterization of the detection process. The present study addresses this issue by focusing on autonomic nervous system modulation during biased reasoning.

The inspiration for this study came from basic cognitive control studies (e.g., Botvinick, Cohen, & Carter, 2004; Ridderinkhof, Ullsperger, Crone, & Nieuwenhuis, 2004). In these studies people are typically presented with elementary conflict tasks in which they need to withhold an inappropriate but dominant response. Previous work in this field showed that the ACC is especially sensitive to the presence of conflict between competing responses (e.g., van Veen & Carter, 2006). The initial study of De Neys et al. (2008) established that this same cortical conflict region was activated when people gave biased responses in a classic reasoning task. However, more recently it has been shown in the cognitive control field that, besides ACC activation, the elementary conflicts also elicit global autonomic arousal (Kobayashi, Yoshino, Takahashi, & Nomura, 2007). In other words, at least in the elementary control tasks, the presence of conflict seems to be accompanied by visceral arousal, as reflected, for example, in increased skin conductance (Hajcak, McDonald, & Simons, 2003). This suggests that basic measures of electrodermal activation can be used as a biological index of conflict detection in the reasoning field. On the basis of the cognitive control findings, one may expect that if conflict detection during thinking is indeed flawless, solving reasoning tasks in which intuitions conflict with logic will elicit increased skin conductance response (SCR).

In the present study, we tested this idea by monitoring participants' SCR while they were solving the infamous syllogistic reasoning problems. For half of the presented problems, referred to as *conflict problems*, the believability of the conclusion conflicted with its logical status, such that, just as in the introductory example, mere intuitive and logical thinking cued an inconsistent response. The other half of the problems were control or no-conflict problems, in which the believability of the conclusion was consistent with its logical status (e.g., a valid syllogism with a believable conclusion). Consider the following example: "All birds have wings. Crows are birds. Therefore, crows have wings." Both a priori beliefs and logical considerations will tell participants to accept the conclusion. In this case there is no conflict and no need to discard the intuitive beliefs.

Contrasting the SCR while people are solving conflict and no-conflict problems will allow us to settle the bias debate. If the bias-as-inhibition-failure view is right, and people realize that their intuition conflicts with the logical appropriate response, the resulting conflict should elicit autonomic arousal, which should be reflected in increased SCRs for the conflict (vs. no-conflict) problems. However, if the conflict monitoring is lax, and people do not detect the inherent intuition–logic clash on the conflict problems, autonomic arousal levels should not differ for the conflict and no-conflict problems.

We clarified that the present study will provide a much needed test of the initial conflict detection findings. However, examining a possible autonomic conflict response also has important conceptual implications. Note that it has long been known that people's online verbalizations during thinking, and their retrospective response justifications, typically do not suggest that they are taking any logical considerations into account (e.g., Evans & Over, 1996; Wason & Evans, 1975); however, although there might be some initial empirical findings suggesting that people detect the presence of an intuition–logic conflict, it is also pretty clear that people do not express this explicitly (e.g., De Neys & Glumicic, 2008). Therefore, it has been hypothesized that conflict detection during thinking is a quite implicit process. Franssens and De Neys (2009; see also Thompson, 2009) suggested that it should be conceived as a "gut feeling": People would experience some general arousal resulting from the conflict detection, but they would not manage to label the detected logical violations explicitly. Bluntly put, people would sense that their response was wrong, but they would not manage to put their finger on it and explain verbally why their response is wrong. However, this post hoc characterization has not been tested directly. Establishing a possible link between autonomic modulation and the conflict detection might help to provide more solid conceptual ground for the idea that people literally "feel" the presence of conflict between their intuitions and logical considerations.

EXPERIMENT 1

Method

Participants. Thirty University of Leuven undergraduates who had not taken logic courses participated in return for a small monetary reimbursement.

Materials. The syllogistic reasoning task was based on the work of Sá, West, and Stanovich (1999). 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). Each problem consisted of a major premise, minor premise, and conclusion. The following item format was adopted (note that the technical labels in italics were not presented on screen):

Major premise: All fruits can be eaten
Minor premise: Hamburgers can be eaten
Conclusion: Hamburgers are fruits
Response alternatives: a. Conclusion follows logically
b. Conclusion does NOT follow logically

Table 1
Reasoning Accuracy (Percentage Correct [PC]) and Response Times (RTs, in Milliseconds) in the Different Experiments

	Accuracy				RTs			
	Conflict		No Conflict		Conflict		No Conflict	
Task	PC	*SE*	PC	*SE*	RT	*SE*	RT	*SE*
Experiment 1: Reasoning	52	6.0	89	2.7	4,033	339	3,314	232
Experiment 2: Belief task	24	6.1	95	2.8	3,151	305	2,504	224

Note—For comparison, we refer to the logical response as the correct response in Experiment 2.

To minimize the possibility that the content of the conflict and no-conflict problems affected the results, we constructed two problem sets in which the content was crossed (see the supplemental materials, Table S1). Each set was presented to half of the participants. The premise and conclusion believability of the conflict and no-conflict problems in each set was matched.

Each trial lasted 18,500 msec. First, a fixation cross was presented for 1,000 msec. Then, the major premise was presented for 3,000 msec. Next, the minor premise was presented for 2,000 msec. Finally, the conclusion and response options were presented. The complete problem remained on the screen for another 10,500 msec. After the 10,500 msec had elapsed, the screen was cleared, and after an additional 2,500-msec rest interval, the next problem was presented.

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. Before the start of the experiment, participants were familiarized with the presentation format and shown one example item.

Skin conductance recording. Standard Ag/AgCl electrodes (1-cm diameter) filled with a Unibase electrolyte were attached to the hypothenar palm of each participant's nondominant hand. The interelectrode distance was 2.5 cm. A Coulbourn skin-conductance coupler (V71-23) provided a constant 0.5 V across electrodes. The analogue conductance signal was passed through a 12-bit AD converter and digitized at 10 Hz. The resulting skin conductance signal was visually inspected, corrected for artifacts, and retained for analysis.

The SCR was quantified by a difference score between the maximum and the minimum skin conductance value within our time intervals of interest (e.g., Botvinick & Rosen, 2009; Dawson, Schell, & Filion, 2000).[1] Detection of a conflict between conclusion believability and validity can occur only after presentation of the conclusion and can be expected to be processed by the time a response has been given. Therefore, our main focus was the SCR in the interval between the presentation of the conclusion and the participant's response keypress; we refer to this interval as the *reasoning phase*. For completeness, we also looked at the SCR in the postresponse window between the response keypress and the end of the trial (*postresponse phase*), and in the interval between the start of the trial and the presentation of the minor premise (*reading phase*). Unless noted otherwise, however, all reported analyses concern the crucial reasoning phase.

Results and Discussion

Behavioral reasoning performance. Participants' performance on the reasoning task was as expected. People were typically biased when cued beliefs and logic conflicted, but had significantly less trouble in solving the no-conflict problems. Overall, correct response rates reached 52% on the conflict problems and 89% on the no-conflict problems [$F(1,29) = 27.38, p < .0001, \eta_p^2 = .49$]. As Table 1 shows, no-conflict problems were also solved faster than conflict problems were [$F(1,29) = 8.18, p < .01, \eta_p^2 = .22$]. These results closely replicate the findings in previous studies with similar syllogistic reasoning problems (e.g., De Neys, 2006; Markovits & Nantel, 1989).

SCR. The main question was whether, despite the bad reasoning performance on the conflict problems, people would nevertheless detect that their intuitive response was not warranted and show increased autonomic arousal when solving these problems. As Figure 1 indicates, results showed that the SCRs in the crucial reasoning phase were significantly higher when conflict problems rather than no-conflict problems were being solved [$F(1,29) = 5.70, p < .025, \eta_p^2 = .16$]. SCRs for the conflict and no-conflict problems did not differ, however, in the initial reading phase [$F(1,29) = 1.87, p = .18$] and in the postresponse phase [$F(1,29) < 1$]. Hence—consistent with the idea that the autonomic arousal results from conflict detection—SCRs increased only after the conflict was introduced, and leveled off after participants had responded near the end of the trial. This pattern indicates that the observed increased autonomic arousal is specifically tied to conflict detection.

Note that our time window of interest for the reasoning phase was the interval between presentation of the conclusion and the participant's response keypress. The behavioral data indicated that with an average response time (RT) of about 4,000 msec, participants needed more than 700 msec longer to solve the conflict problems than to solve the no-conflict ones. This implies that the reasoning phase was typically longer for the conflict trials. One might argue that it is the longer RT per se that drives the observed higher SCRs in the reasoning phase. To eliminate such a confound, we ran an additional analysis in which we controlled the length of the reasoning interval on the conflict trials. Only the skin conductance values during the first 3,000 msec after conclusion presentation were taken into account (i.e., a cutoff value about one *SE* below the average RT for the no-conflict problems). Hence, in this control analysis, the reasoning phase was actually slightly shorter for the conflict problems than for the no-conflict ones; however, the overall pattern of results was not affected. The SCRs were still significantly higher when conflict problems rather than no-conflict problems were being solved [$F(1,29) = 4.64, p < .05, \eta_p^2 = .14$].[2] This establishes that the longer conflict latencies in our standard analysis do not drive the higher SCRs.

The overall results support the idea that conflict detection during thinking is associated with increased autonomic arousal, but they do not yet allow us to decide between the lax- and flawless-detection views. Although reasoning accuracy on the conflict problems was low, some people did manage to respond correctly. Both the lax and flawless views entail that good reasoners, who

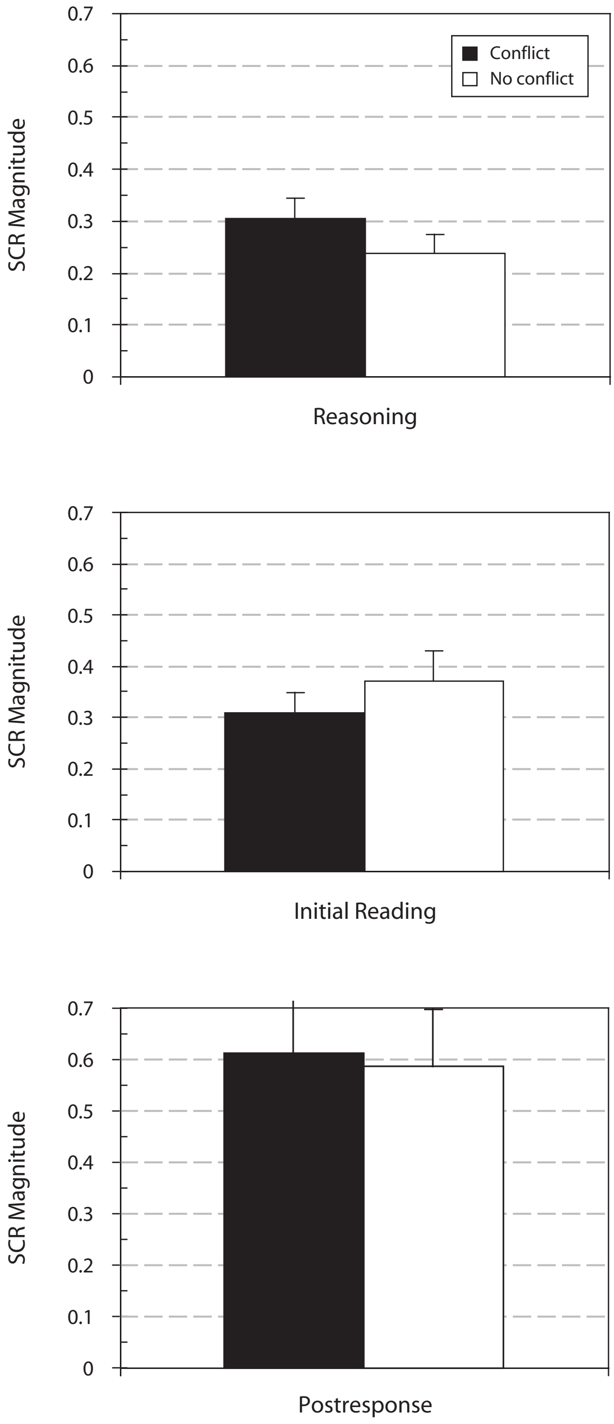

Figure 1. Average skin conductance response (SCR) magnitude (in μS) for conflict and no-conflict syllogisms during reasoning interval (top panel), initial reading (middle panel), and postresponse (bottom panel). Error bars indicate standard errors.

manage to solve the problems correctly, will detect the unwarranted nature of the intuitive response. Hence, although overall accuracy was low, a supporter of the lax-detection view could still argue that the observed overall SCR increase was simply driven by the good reasoners. To address this critique, we looked at the association between accuracy on the conflict problems and the extent of the observed SCR increase on these problems during the reasoning phase.

If only good reasoners were to show the increased arousal, there should be a clear link between the two factors. However, a correlation analysis showed that the SCR increase (i.e., SCR conflict problems − SCR no-conflict problems) did not depend on reasoning accuracy ($r = .15$, $p = .413$). To explore this issue further, Figure 2 shows the SCRs during the reasoning phase for good and bad reasoners on the basis of a median split on the conflict accuracy. Average reasoning accuracy in the good group ($n = 13$) was 89% and 24% in the bad group ($n = 17$). As Figure 2 illustrates, the crucial SCR increase on the conflict problems did not depend on the skill factor [$F(1,28) < 1$]. Overall, good reasoners did tend to have higher SCRs but this trend was not significant [$F(1,28) < 1$].

In an additional analysis, we looked at even more extreme skill groups. We contrasted performance of the very best and worst reasoners in our sample; people who solved none ($n = 6$) or all ($n = 7$) of the conflict problems correctly. Overall, SCRs were still significantly greater for conflict than for no-conflict problems [$F(1,11) = 6.77$, $p < .025$, $\eta_p^2 = .38$], but even for these extreme groups the increase did not differ for good and bad reasoners [$F(1,11) = 2.10$, $p = .17$]. These findings clearly establish that the observed overall SCR increase is not solely driven by the good reasoners. Consistent with the flawless-detection view, everybody seems to be detecting the conflict between cued intuitions and the logical appropriate response.

A last issue we need to address is the impact of possible learning effects. The initial studies that started focusing on conflict detection were typically quite lengthy. For example, in their fMRI study De Neys et al. (2008) presented almost 100 reasoning items. One might argue that the repeated presentation and repetitive nature of these studies cued participants to start paying attention

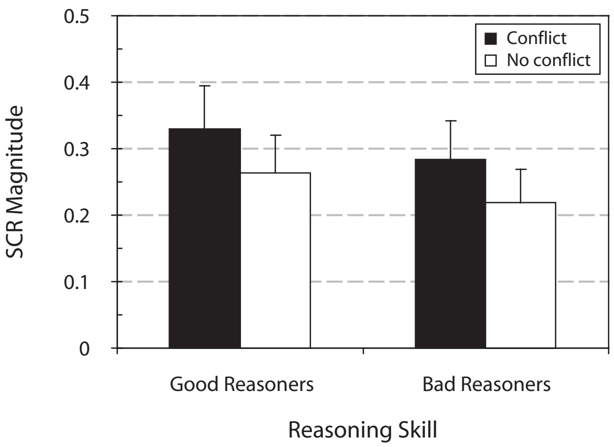

Figure 2. Average skin conductance response (SCR) magnitude (in μS) during the reasoning interval for conflict and no-conflict syllogisms of the best (good reasoners) and worst (bad reasoners) scoring half of the participants. Error bars indicate standard errors.

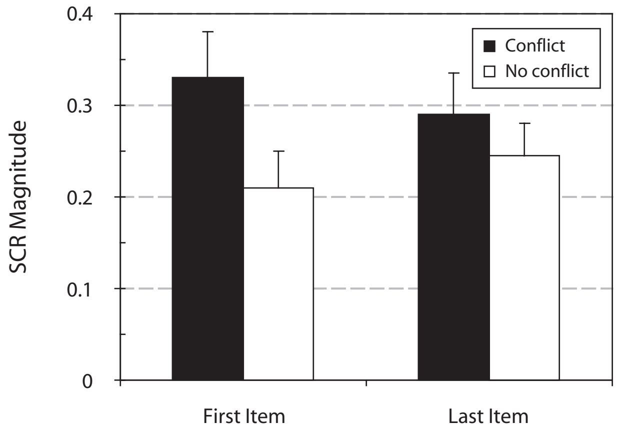

Figure 3. Average skin conductance response (SCR) magnitude (in μS) during the reasoning interval for the first and last presented conflict and no-conflict syllogisms. Error bars indicate standard errors.

to the conflict manipulation; hence, the flawless detection in these studies might simply be an artifact that results from a learning effect. Note that we already reduced the number of presented items in the present study to limit the impact of such a learning confound. However, to address the issue further, we repeated our analysis solely with the first presented conflict and no-conflict problem that every participant solved.

Contrary to the learning hypothesis, we replicated the overall pattern. As Figure 3 shows, right from the start of the experiment SCRs were higher for conflict than for no-conflict problems [$F(1,29) = 6.61$, $p < .025$, $\eta_p^2 = .19$]. Note that, as with the overall analysis, the increase did not depend on whether participants were good or bad reasoners and managed to solve the conflict problem correctly ($r_s = .06$, $p = .75$). We also contrasted the SCRs for the last conflict and no-conflict problem that participants solved. As Figure 3 shows, the SCR increase on the conflict problems tended to become less pronounced near the end of the experiment. Although this trend was not significant [$F(1,29) = 1.43$, $p = .24$], it clearly argues against the learning hypothesis. If anything, repeated presentation tended to result in autonomic habituation and decreased rather than boosted the observed effects.

EXPERIMENT 2

Experiment 1 established that dealing with conflicting logical and tempting intuitive responses during reasoning resulted in increased autonomic arousal, as reflected in increased SCRs. This visceral conflict response did not result from a learning effect and was shown by all reasoners. Thereby, the findings are consistent with the view that reasoning bias cannot be attributed to a conflict detection failure. However, the apparent omnipresence of the arousal signal also gives rise to possible alternative accounts. It could be argued, for example, that the increased SCRs do not result from a reasoning-related detection process per se, but simply from a superficial material characteristic or a more general process not related to reasoning. Although our study design minimized such a possible confound, it cannot be completely discarded. In a second control experiment, we therefore addressed this issue directly.

In Experiment 2, we presented participants the exact same material as in Experiment 1, but simply asked them to evaluate the believability of the conclusion. Hence, in this task there was no need to engage in a logical reasoning process. Consequently, since the task will not cue a logical response, there should also not arise a conflict between a cued logical and intuitive response; so if it is really the case that the observed increased SCRs in Experiment 1 result from a reasoning-related conflict detection process, we should no longer observe them in Experiment 2.

Method

Participants. Thirty University of Leuven undergraduates who had not taken logic courses participated in return for a small monetary reimbursement. The data of Experiment 1 were used as a baseline to test the impact of the task manipulation.

Materials. In the belief evaluation task, participants were presented with the same items as in Experiment 1. 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 that they 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 "it was fine to select the response that came first to mind and seemed intuitively most plausible." Previous studies indicated that some participants spontaneously engage in logical reasoning when presented with conditional syllogisms, 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. Except for these modifications, the presentation procedure was identical to Experiment 1.

In the skin conductance recording, the same procedure as in Experiment 1 was used. Data of 1 participant were lost due to equipment failure and were not included in the analyses.

Results and Discussion

Behavioral performance. Accuracy and response latencies established that the task manipulation was successful. As Table 1 indicates, participants gave overall more belief-based responses, when instructed to do this in Experiment 2, than in Experiment 1 [$F(1,57) = 5.44$, $p < .025$, $\eta_p^2 = .09$]. This tendency was more pronounced on the conflict than on the no-conflict problems [$F(1,57) = 12.38$, $p < .001$, $\eta_p^2 = .18$]. Overall, responses were also given faster in the belief evaluation task [$F(1,57) = 6.32$, $p < .025$, $\eta_p^2 = .15$]. These faster responses were equally clear for conflict and no-conflict problems [$F(1,57) < 1$]. The trend toward faster and more frequent belief-based responses indicates that participants indeed engaged in a more intuitive mode of processing in Experiment 2; this demonstrates that our instruction manipulation was successful.

SCRs. Figure 4 shows the average SCRs after conclusion presentation for conflict and no-conflict problems in the reasoning and belief evaluation task. An ANOVA established that the impact of the conflict factor clearly differed in both tasks [$F(1,57) = 4.65, p < .05, \eta_p^2 = .08$]. As Figure 3 shows, SCRs were overall lower in the belief evaluation task [$F(1,57) = 8.41, p < .01, \eta_p^2 = .13$], but contrary to Experiment 1, SCRs for conflict and no-conflict problems no longer differed in the belief evaluation task [$F(1,57) < 1$]. Hence, when people were not engaged in reasoning, and cued intuitions did not conflict with logical considerations, autonomic arousal did not increase when conflict problems were being solved. This finding establishes that the SCR increase we observed in Experiment 1 results from reasoning-related conflict detection.

SCR and believability × validity interaction. The believability and validity of the problems in our study were completely crossed. We were interested in the overall conflict between these two factors, and had no specific hypotheses about possible further lower level interactions. As in previous studies, our analyses therefore focused on the main conflict factor and collapsed data over the belief and validity levels (e.g., De Neys, 2006; Goel & Dolan, 2003; Stanovich & West, 2000; Tsujii & Watanabe, 2009). For completeness, we did enter the two factors separately in a 2 (believability) × 2 (validity) × 2 (experiment) mixed-model ANOVA on the SCRs in the reasoning phase. Table 2 gives an overview of the findings. There were main effects of the believability and experiment factors: Overall SCRs were lower in Experiment 2 than in Experiment 1 [$F(1,57) = 8.57, p < .005, \eta_p^2 = .13$], and higher for unbelievable than for unbelievable conclusions [$F(1,57) = 4.99, p < .05, \eta_p^2 = .08$]. The interaction between the three factors was also marginally significant [$F(1,57) = 3.70, p < .06, \eta_p^2 = .06$]. Other effects and interactions failed to reach (marginal) significance. We explored the three-way interaction with planned contrast. For the reasoning task in Experiment 1, results showed that there was a significant believability × validity interaction [$F(1,57) = 6.58, p < .025, \eta_p^2 = .10$]. For the belief evaluation task in Experiment 2, this interaction was not significant [$F(1,57) < 1$]. Further planned contrasts on

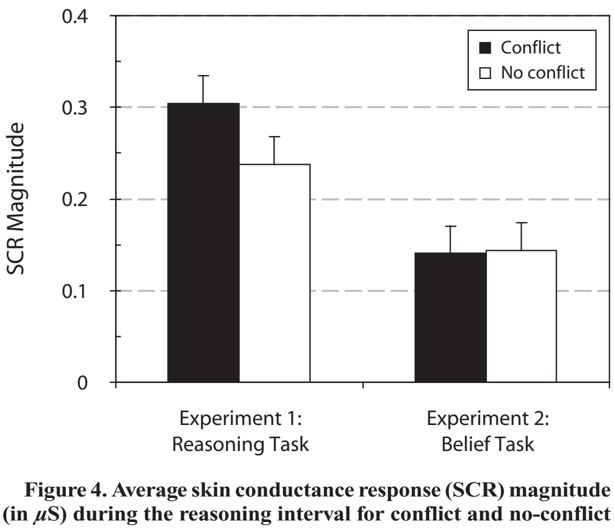

Figure 4. Average skin conductance response (SCR) magnitude (in μS) during the reasoning interval for conflict and no-conflict syllogisms in Experiments 1 and 2. Error bars indicate standard errors.

Table 2
Average Skin Conductance Response (in μS) During the Reasoning Interval As a Function of Conclusion Validity and Believability in Experiments 1 and 2

	Believability			
	Believable		Unbelievable	
Validity	*M*	*SE*	*M*	*SE*
Experiment 1: Reasoning				
Valid	.19	.03	.30	.04
Invalid	.31	.03	.30	.05
Experiment 2: Belief task				
Valid	.11	.03	.17	.04
Invalid	.11	.03	.17	.05

the Experiment 1 data indicated that the effect of validity was significant for the believable conclusions [$F(1,57) = 14.23, p < .001, \eta_p^2 = .20$], but not for the unbelievable ones [$F(1,57) < 1$]. This implies that the global conflict effect is especially driven by the believable problems. Being presented with unbelievable material seems to lead to a general SCR boost that blurs an additional effect of logical validity. However, on the believable problems it was still the case that SCRs were higher when the conclusion was logically invalid and beliefs and logic conflicted. Consistent with our hypothesis, this conflict-related SCR increase on the believable problems was completely absent when people were simply evaluating the conclusion believability in Experiment 2 [$F(1,57) < 1$].

GENERAL DISCUSSION

The present study shows that dealing with conflicting logical and intuitive responses during reasoning is accompanied by an increase in autonomic arousal. The increased autonomic arousal was reflected in an SCR boost right after the erroneous intuition was cued in conflict syllogisms but was absent when people were not engaged in a reasoning task. This establishes that the autonomic arousal is specifically tied to the detection of the conflict between logic and intuition. The finding that even the most biased reasoners showed the autonomic conflict response validated the view that intuitive bias cannot be attributed to a detection failure. Although not everyone might manage to discard the tempting intuitive response, all reasoners seem to be sensitive to the presence of the conflict.

The presence of a clear autonomic conflict response lends credence to the idea that conflict detection can be conceived as a "gut feeling" (e.g., Franssens & De Neys, 2009; see also Thompson, 2009, for related suggestions). People seem to detect that their intuitions conflict with more logical considerations at the visceral level. At this point the present conflict findings show an interesting link with the seminal work of Bechara, Damasio, and colleagues (e.g., Bechara, Damasio, Tranel, & Damasio, 1997; Damasio, 1994). In Bechara and Damasio's studies, participants were presented with a gambling task in which they could select cards from decks with different payoffs. Bechara and Damasio observed that participants needed about 80 trials before they could explicitly point out which decks were the good and bad ones. However, when the authors examined people's autonomic responses, they noticed that after a number of trials (i.e., between the 10th and 50th trial) people started showing an SCR increase (i.e., a "somatic marker") before they selected a card from the bad decks. Hence, although people had not yet acquired explicit knowledge of which decks were the bad ones, their autonomic response indicated that they implicitly acquired this knowledge. It will be evident that both Bechara and Damasio's work and the present study share the general idea that there is "more than meets the eye" in human thinking: Although people's overt response (i.e., their answer on the reasoning tasks) might be biased, this does not imply that they cannot detect this bias at a more implicit level. However, at the same time, we would like to point out that the autonomic signal that Bechara and Damasio focused on is different from the conflict-driven autonomic arousal signal that we report. First, Bechara and Damasio argued that the autonomic reaction in their gambling task resulted from the negative feedback (i.e., losing money) participants received after selecting bad cards. Clearly, in our classic reasoning task, such performance feedback was completely absent. Furthermore, our autonomic conflict response was present right from the start and did not result from a learning effect. Lastly, Bechara and Damasio observed their somatic markers only before participants made erroneous selections. As one would expect from a conflict-related detection signal, we observed our increased SCRs for both bad and good responses. Hence, although our findings and Bechara and Damasio's work both point to the importance of the autonomic response level in human thinking, the two bodies of work seem to deal with different autonomic signals.

The crucial implication of the present study is that people are less ignorant than the widespread intuitive bias suggests. Although people might not manage to label the logical principles being violated, the presence of the autonomic conflict response establishes that logical considerations are implicitly taken into account during reasoning. This implies that humans are no mere intuitive, illogical reasoners who disregard the logical implications of their judgments. Although our inferences are often biased, we do seem to sense that we are wrong.

We believe that our findings nicely illustrate the relevance of the cognitive neuroscience literature on cognitive control for the reasoning field (e.g., Botvinick et al., 2004; Ridderinkhof et al., 2004). The neuroscience toolbox allows reasoning researchers to start scratching below the response surface and address the key theoretical debates. In our view, a further continuation of the crosstalk between the fields might prove especially fruitful. For example, one interesting line for further study is the link between the presently observed autonomic response and previously established ACC activation during biased reasoning (De Neys et al., 2008). At least with the elementary cognitive control tasks, it has been suggested that the autonomic conflict signal originates in the ACC (e.g., Critchley, Tang, Glaser, Butterworth, & Dolan, 2005; Hajcak et al., 2003). Combining fMRI and SCR recordings during reasoning might allow us to examine how strongly both signals covary, and whether one or the other specific subdivision within the ACC (e.g., dorsal or rostral ACC) drives the autonomic signal. Such research could also help to further distinguish the presently observed autonomic conflict signal from Bechara and Damasio's somatic marker signals (which have been shown to originate from the ventromedial prefrontal cortex; e.g., Bechara et al., 1997). Clearly, looking at the reasoning performance and SCRs of patients with lesions in the identified brain areas could then be used to further specify the neural underpinning of the autonomic conflict signal.

In the present study, we focused on popular syllogistic reasoning problems to study the conflict between intuitive and logical thinking. As we noted in the introduction, such conflicts arise in a wide range of classic reasoning and decision-making tasks. Clearly, it would be worthwhile to test the generalizability of the present findings across these tasks. Such generalization might be especially interesting, given the possible distortion that seemed to be created by the presence of unbelievable material in the present study. We mentioned that, when working with syllogisms, one typically creates conflict by crossing the problem believability with logical validity (e.g., one can construct both believable–invalid problems and unbelievable–valid problems). When we examined both factors separately in Experiment 2, we observed that the conflict findings were less clear for the unbelievable problems. Unbelievable material gave rise to an overall SCR boost that might have blurred an additional effect of logical validity. It is interesting to note that the conflict in other reasoning tasks typically involves a conflict between believable intuitive material (e.g., information that fits with stored stereotypes; see Kahneman & Frederick, 2005) and normative logical or probabilistic considerations. Given the possible distorting impact of unbelievable material, this could imply that the presently observed conflict-related autonomic response might be even clearer in nonsyllogistic reasoning tasks. This underscores the importance of generalizing our findings in future studies. Overall, it must be remembered that the present study was only the first to look for possible conflict-related autonomic arousal during thinking. Clearly, this pioneering status also implies that, in the absence of future replication, the findings need to be interpreted with caution.

Another issue that will need further study is the precise conceptualization of the observed conflict-related autonomic arousal. We noted that the presence of an autonomic conflict response supports the idea that conflict detection can be conceived as a "gut" conflict feeling, and that the bias is detected at an implicit level. It should be clear that our use of the label *implicit* here refers to the well-established fact that biased reasoners do not explicitly refer to any violation of logical principles when asked for a verbal response justification. Hence, the crucial logical knowledge that must be present in order to detect a conflict with one's intuitive beliefs has been characterized as implicit knowledge. Studies that started focusing on conflict detection during thinking have, therefore, typically referred to the detection as an implicit process (e.g., De Neys & Franssens, 2009; De Neys & Glumicic, 2008; Franssens & De Neys, 2009). However, this does not imply that the arousal generated by the conflict is also implicit, in the sense that it is not consciously experienced. We do believe that people are perceiving the arousal consciously, and that this arousal informs them that their judgment is questionable; our point is simply that people will not manage to explain why their judgment is logically questionable. Note that the idea that people consciously experience the outcome of more implicit processing is quite generally accepted within the reasoning field (e.g., Evans, 2008). Consistent with this idea, recent work by our group showed that, although reasoners do not explicitly mention logical principles, they are indicating that they doubt their response; this is shown, for example, by decreased confidence ratings for biased conflict responses (De Neys, Cromheeke, & Osman, 2009). At the same time, however, we need to acknowledge that the mere idea of an unconscious conflict feeling has not been given serious consideration in the research on conflict detection during thinking. Recent work on unconscious feelings and levels of awareness in intuitive judgments (e.g., Topolinski & Strack, 2009; Winkielman & Berridge, 2004) indicates that the possibility of unconscious autonomic arousal cannot be a priori excluded. An interesting idea to address this issue more directly in further studies would be to specifically ask people to report any experienced arousal. Hence—although we do believe that the conflict-related arousal feeling is consciously perceived—it will be clear that this claim needs to be further validated. On the basis of the present findings per se, the safest course of action would be to refrain from drawing strong conclusions about the ultimate experiential status of the observed conflict arousal.

Although we acknowledge that the study of the conflict detection process during reasoning is still in its infancy, we do want to point to some possible practical implications of the findings. We have argued previously that specifying the nature of the intuitive bias is crucial to designing more effective intervention strategies to debias and improve human reasoning and decision making (De Neys et al., 2008). Intuitively, the bias-as-detection-failure view is very appealing. Most people seem to assume that giving a biased response implies that one does not know the correct answer. This is reflected in the design of experimental intervention studies, in which participants are typically tutored about logic or probability theory. Broader sociological and governmental intervention attempts often share this implicit provide-the-knowledge idea: As a society, we spend billions of dollars on campaigns to inform the public about the dangers of anything from smoking to drug use to obesity. Unfortunately, these interventions have typically not been very effective (e.g., Hornik, Jacobsohn, Orwin, Piesse, & Kalton, 2008; Kahneman, Slovic, & Tversky, 1982; Paek & Gunther, 2007). Given our findings, this failure is not necessarily surprising. Bluntly put, if people already detect the unwarranted nature of their intuitive decisions, it is clear that simply telling them “smoking kills” or “fast food equals fat food” is not going to have much of an impact on their decision to light another cigarette or order a second Big Mac. Although we realize that caution is needed when generalizing from studies with formal laboratory-based reasoning tasks, our findings suggest that these interventions have been targeting the wrong component of the thinking process. Future programs might be more effective if they tried to help people sidestep bad intuitions (e.g., strengthening people’s inhibitory skills; see Houdé, 2007) rather than merely inform them that the intuitions are bad per se. In general, the available conflict detection evidence indicates that people are less ignorant than their biased judgments suggest.

AUTHOR NOTE

Correspondence concerning this article should be addressed to W. De Neys, Laboratoire CLLE, Maison de la Recherche, 5 allée A. Machado, 31058 Toulouse Cedex 9, France (e-mail: wim.deneys@univ-tlse2.fr).

REFERENCES

Alexander, B. (2009). Amid swine flu outbreak, racism goes viral. *NBC Los Angeles*. Retrieved June 11, 2009, from www.msnbc.msn.com/id/30467300/.

Ballantyne, C. (2009). Will Egypt’s plans to kill pigs protect it from swine—sorry, H1N1 flu? Retrieved June 16, 2009, from www.scientificamerican.com/blog/60-second-science/post.cfm?id=will-egypts-plans-to-kill-pigs-prot-2009-05-01.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, **275**, 1293-1295.

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

Botvinick, M. M., & Rosen, Z. B. (2009). Anticipation of cognitive demand during decision-making. *Psychological Research*, **73**, 835-842.

Critchley, H. D., Tang, J., Glaser, D., Butterworth, B., & Dolan, R. J. (2005). Anterior cingulate activity during error and autonomic response. *NeuroImage*, **27**, 885-895.

Damasio, A. R. (1994). *Descartes’ error: Emotion, reason, and the human brain*. New York: Putnam.

Dawson, M. E., Schell, A. M., & Filion, D. L. (2000). The electrodermal system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 200-223). New York: Cambridge University Press.

De Neys, W. (2006). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science*, **17**, 428-433.

De Neys, W., Cromheeke, S., & Osman, M. (2009). *Biased but in doubt: Conflict and decision confidence*. Manuscript submitted for publication.

De Neys, W., & Franssens, S. (2009). Belief inhibition during thinking: Not always winning but at least taking part. *Cognition*, **113**, 45-61.

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 & 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.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, **49**, 709-724.

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. (2008). Dual-processing accounts of reasoning, judgment and social cognition. *Annual Review of Psychology*, **59**, 255-278.

Evans, J. St. B. T., & Over, D. E. (1996). *Rationality and reasoning*. Hove, U.K.: Psychology Press.

Franssens, S., & De Neys, W. (2009). The effortless nature of conflict detection during thinking. *Thinking & Reasoning*, **15**, 105-128.

Goel, V., & Dolan, R. J. (2003). Explaining modulation of reasoning by belief. *Cognition*, **87**, B11-B22.

Hajcak, G., McDonald, N., & Simons, R. F. (2003). To err is autonomic: Error-related brain potentials, ANS activity, and post-error compensatory behavior. *Psychophysiology*, **40**, 895-903.

Hornik, R., Jacobsohn, L., Orwin, R., Piesse, A., & Kalton, G. (2008). Effects of the National Youth Anti-Drug Media Campaign on youths. *American Journal of Public Health*, **98**, 2229-2236.

Houdé, O. (2007). First insights on "neuropedagogy of reasoning." *Thinking & Reasoning*, **13**, 81-89.

Kahneman, D. (2002, December). *Maps of bounded rationality: A perspective on intuitive judgment and choice*. Nobel Prize lecture. Retrieved January 11, 2006, from http://nobelprize.org/nobel_prizes/economics/laureates/2002/kahnemann-lecture.pdf.

Kahneman, D., & Frederick, S. (2005). A model of heuristic judgment. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 267-293). New York: Cambridge University Press.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.) (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Kobayashi, N., Yoshino, A., Takahashi, Y., & Nomura, S. (2007). Autonomic arousal in cognitive conflict resolution. *Autonomic Neuroscience*, **132**, 70-75.

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

Oakhill, J., Johnson-Laird, P. N., & Garnham, A. (1989). Believability and syllogistic reasoning. *Cognition*, **31**, 117-140.

Paek, H.-J., & Gunther, A. C. (2007). How peer proximity moderates indirect media influence on adolescent smoking. *Communication Research*, **34**, 407-432.

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science*, **306**, 443-447.

Sá, W. C., 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.

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

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

Thompson, V. A. (2009). Dual process theories: A metacognitive perspective. In J. St. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 171-195). New York: Oxford University Press.

Topolinski, S., & Strack, F. (2009). Scanning the "Fringe" of consciousness: What is felt and what is not felt in intuitions about semantic coherence. *Consciousness & Cognition*, **18**, 608-618.

Tsujii, T., & Watanabe, S. (2009). Neural correlates of dual-task effect on belief-bias syllogistic reasoning: A near-infrared spectroscopy study. *Brain Research*, **1287**, 118-125.

van Veen, V., & Carter, C. S. (2006). Conflict and cognitive control in the brain. *Current Directions in Psychological Science*, **15**, 237-240.

Wason, P. C., & Evans, J. St. B. T. (1975). Dual processes in reasoning? *Cognition*, **3**, 141-154.

Winkielman, P., & Berridge, K. C. (2004). Unconscious emotion. *Current Directions in Psychological Science*, **13**, 120-123.

NOTES

1. A secondary analysis measuring SCR as the difference between the average skin conductance values during the reading and reasoning phases yielded comparable results.

2. In addition, a correlational analysis on the original data established that there was no association between an individual's reasoning time on the conflict problems ($r = .11$, $p = .53$) or no-conflict problems ($r = -.10$, $p = .57$) and the observed SCR increase.

SUPPLEMENTAL MATERIALS

A detailed discussion of the materials for this article may be downloaded from http://cabn.psychonomic-journals.org/content/supplemental.

(Manuscript received July 30, 2009;
revision accepted for publication December 7, 2009.)

OPEN ACCESS Freely available online

PLoS one

Biased but in Doubt: Conflict and Decision Confidence

Wim De Neys[1], Sofie Cromheeke[2], Magda Osman[3]*

1 Centre National de la Recherche Scientifique (CNRS), University of Toulouse, Toulouse, France, 2 Department of Psychology, University of Leuven, Leuven, Belgium, 3 School of Biological and Chemical Sciences, Queen Mary University of London, London, United Kingdom

Abstract

Human reasoning is often biased by intuitive heuristics. A central question is whether the bias results from a failure to detect that the intuitions conflict with traditional normative considerations or from a failure to discard the tempting intuitions. The present study addressed this unresolved debate by using people's decision confidence as a nonverbal index of conflict detection. Participants were asked to indicate how confident they were after solving classic base-rate (Experiment 1) and conjunction fallacy (Experiment 2) problems in which a cued intuitive response could be inconsistent or consistent with the traditional correct response. Results indicated that reasoners showed a clear confidence decrease when they gave an intuitive response that conflicted with the normative response. Contrary to popular belief, this establishes that people seem to acknowledge that their intuitive answers are not fully warranted. Experiment 3 established that younger reasoners did not yet show the confidence decrease, which points to the role of improved bias awareness in our reasoning development. Implications for the long standing debate on human rationality are discussed.

Citation: De Neys W, Cromheeke S, Osman M (2011) Biased but in Doubt: Conflict and Decision Confidence. PLoS ONE 6(1): e15954. doi:10.1371/journal.pone.0015954

Editor: Hans Op de Beeck, University of Leuven, Belgium

Received September 22, 2010; **Accepted** November 30, 2010; **Published** January 25, 2011

Funding: This research was supported by grants from the Centre National de la Recherche Scientifique (CNRS), Fonds Wetenschappelijk Onderzoek Vlaanderen (FWO), and Engineering and Physical Sciences Research Council (EPSRC - EP/F069421/1). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: wim.deneys@univ-tlse2.fr

Introduction

Human judgment is often biased by erroneous intuitions. Consider, for example, the success of the popular "Buy One, Get Second One 50% Off" sale you often see at retail stores. If you buy one item you get the opportunity to buy a second, similar one for only half of the original price. Even when we do not need the second item, we will often be tempted to buy it simply because our intuition is telling us that by not taking the offer we are missing out on a unique opportunity to get something for "only half of the original price". From a normative point of view, however, this behavior is quite irrational. If you do not need a specific good, spending any money to obtain it is a waste of scarce financial resources. Hence, while we intuitively think that we are saving money, the store marketeers are actually tricking us to spend more than we should.

Decades of reasoning and decision-making research have shown that similar intuitive thinking is biasing people's judgment in a wide range of situations and tasks [1,2]. In general, human reasoners seem to have a strong tendency to base their judgment on fast intuitive impressions rather than on more demanding, deliberative reasoning. Although this intuitive or so-called "heuristic" thinking might sometimes be useful, it will often cue responses that conflict with normative logical or probabilistic considerations and bias our decision-making.

Whereas it is well established that human judgment is often biased, the nature of this bias is far less clear. A central question is whether or not people know that they are biased and detect that their intuitive conclusions are not logically warranted. Some influential authors have argued that the widespread heuristic bias can be attributed to a failure to monitor our intuition [3]. Because of lax monitoring people would simply fail to detect that the intuitive response conflicts with logical considerations. However, others have suggested that there is nothing wrong with the detection process (e.g., [4–6]). According to these authors, people do notice that their intuitive response conflicts with traditional normative considerations. The problem, however, is that despite this knowledge they will not always manage to inhibit and discard the tempting intuitive beliefs. Thus, people "behave against their better judgment" [4] when they give an unwarranted heuristic response: They detect that they are biased but simply fail to block the biased response. In sum, according to this flawless detection view biased decisions are attributed to an inhibition failure rather than a conflict detection failure per se.

Clarifying the efficiency of the detection process and the nature of the heuristic bias is paramount for the development of reasoning and decision-making theories. The issue has also far-reaching implications for our view of human rationality (e.g., [7,8]). Unfortunately, deciding between the two views has not been easy [9,10]. Consistent with the lax detection view, it has long been established that reasoners' online verbalizations and retrospective response justifications do not indicate that they are taking any traditional logical or probabilistic considerations into account during reasoning (e.g., [11,12]). For example, in one study De Neys and Glumicic [13] asked participants to think aloud while they were solving problems that were modelled after Kahneman and Tversky's [2] classic base-rate neglect problems. Consider the following example:

> A psychologist wrote thumbnail descriptions of a sample of 1000 participants consisting of 995 females and 5 males. The

description below was chosen at random from the 1000 available descriptions.

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.

Which one of the following two statements is most likely?

a. Jo is a man
b. Jo is a woman

From a probabilistic point of view, given the size of the two groups in the sample, it will be more likely that a randomly drawn individual will be female (i.e., the largest group in the sample). However, intuitively many people will be tempted to respond that the individual is a male based on stereotypical beliefs cued by the description ("Jo is an engineer and drinks beer").

The central question for De Neys and Glumicic [13] was whether verbal protocols would indicate that when people selected the intuitive response option ("a. Jo is a man") they at least referred to the group size information during the reasoning process (e.g., " … because Jo's drinking beer and loud I guess Jo'll be a guy, *although there were more women* …"). Such a basic sample size reference during the reasoning process can be considered as a minimal indication of successful bias detection: It indicates that people are not simply neglecting the normative base-rate information. Results clearly showed, however, that except for the few participants who gave the probabilistic base-rate response ("b. Jo is a woman), people hardly ever mentioned the base-rates. Hence, consistent with the lax detection view and numerous classic verbalisation studies, the explicit protocols suggested that biased reasoners are indeed mere intuitive thinkers who do not detect that their intuition conflicts with normative considerations.

Studies that started looking at more implicit detection measures, however, have presented support for the flawless detection view (e.g., [14–16]). For example, De Neys et al. [15] used fMRI to monitor the activation of a specific brain area, the anterior cingulate cortex (ACC), which is believed to mediate conflict detection during thinking. Participants again solved the classic base-rate problems in which the base-rates and personality description cued a conflicting response. Participants were also presented with no-conflict control versions in which the base-rates were switched around so that both the base-rates and description cued the same response. If people indeed neglected the base-rates, as the explicit protocols suggested, and did not detect that base-rates and description cued inconsistent responses, the conflict and control problems should not be processed any differently. Results showed, however, that the ACC was much more activated when people solved the classic conflict versions than when they solved the control versions without such conflicts. This increased activation was equally clear for correctly and incorrectly solved conflict problems. Hence, even when people were biased, the ACC seemed to signal the intrinsic conflict between the cued intuitive and base-rate response. Bluntly put, although people were not explicitly referring to the base-rate information, their brains did seem to pick up that their response was not consistent with it. Further work with the base-rate task and other logical reasoning problems showed that this increased ACC activation for biased responses is also accompanied by an increased autonomic activation [17], increased response decision-time [13,18,19], and altered accessibility of stored information that is associated with the cued logical/probabilistic and intuitive responses (e.g., [13,14,16]).

In sum, although it is clear that people do not explicitly detect that they are erring, available evidence suggests that they do seem to be sensitive to the presence of conflict between cued intuitive and normative logical or probabilistic principles at a more implicit level. The lack of explicitation has been explained by arguing that the neural conflict detection signal should be conceived as an implicit "gut" feeling. The signal would inform people that their intuition is not fully warranted but people would not always manage to verbalize the experience and explicitly label the logical principles that are being violated [16] (see also [20] for related suggestions). Although this hypothesis is not unreasonable, it faces a classic caveat. Without discarding the possible value of implicit processing, the lack of explicit evidence does open the possibility that the implicit conflict signal is a mere epiphenomenon. That is, the implicit conflict detection research clearly established that some part of our brain is sensitive to the presence of conflict in classic reasoning tasks. However, this does not necessarily imply that this conflict signal is also being used in the reasoning process. In other words, showing that the presence of conflict is detected does not suffice to argue that reasoners also "know" that their intuition is not warranted. Indeed, a critic might utter that the fact that despite the clear presence of a conflict signal people do not report experiencing a conflict and keep selecting the erroneous response, questions the value of this signal. Hence, what is needed to settle the bias debate is some minimal (nonverbal) indication that this signal is no mere epiphenomenon but has a functional impact on the reasoning process. This issue is the focus of the present study.

A straightforward way to assess the functional relevance of the implicit conflict signal is to examine people's decision confidence after they solve a reasoning problem. If the detection signal is not merely epiphenomenal, but actually informs people that their intuitive response is not fully warranted, people's decision confidence should be affected. That is, if people detect that they are biased but simply fail to verbalize the experience, we should at the very least expect to see that they do not show full confidence in their judgments.

Of course, people might never show full confidence and there might be myriad reason for why individuals differ in their confidence ratings (e.g., [21,22]). Note, however, that our main research question does not concern people's absolute confidence level. As with the initial detection studies, in the present study we will present participants classic conflict problems and newly constructed no-conflict control problems. The only difference between the two types of problems is that cued intuitions conflict with traditional normative principles in the conflict versions while intuition and normative principles cue the same response in the no-conflict versions. The aim of the confidence contrast for the two types of problems is to help decide the detection debate. If detection of the intrinsic conflict on the classic versions is functional for the reasoning process and informs people that their intuitive response is questionable, participants should show lower confidence ratings after solving conflict problems as compared to no-conflict problems. If people do not detect the presence of conflict or the signal has no impact on the reasoning process, confidence ratings for the two types of problems should not differ.

We tested the confidence predictions in two initial experiments. In Experiment 1 people were presented with problems based on the classic base-rate task [2]. Experiment 2 tested the predictions with another well-studied reasoning task, the conjunction fallacy [23], to examine the generality of the findings. In Experiment 3 we tried to validate the findings by testing the performance of a population of reasoners who have been shown to have suboptimal conflict detection skills. Developmental studies in the cognitive control field have established that basic conflict monitoring abilities are not fully developed before late adolescence and young adulthood (e.g., [24–26]). Therefore, in Experiment 3 we presented our reasoning problems to a group of early and late adolescents and also asked them to rate their decision confidence. Given that conflict detection should be less efficient for young adolescents, we predict that any possible confidence decrease after solving conflict problems with adults or late adolescents should be absent (or less clear at least) in early adolescents.

Methods

Ethics statement

All experiments in this study were conducted in accordance with the Declaration of Helsinki and approved by the local ethics committee of the University of Leuven. Written informed consent was obtained from all participants (or their parent or guardian).

Experiment 1: Base-rate task

Participants. A total of 247 undergraduates who were taking an introductory psychology course at the University of Leuven (Belgium) participated in return for course credit. Participants provided written informed consent and the study was approved by the local ethics committee of the University of Leuven.

Material. Participants solved a total of six base-rate problems. Three of these were classic 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 three no-conflict problems the description and the base-rates agreed.

Problems were based on a range of stereotypes (e.g., involving gender, age, nationality, see Appendix S1 for an overview). Descriptions were selected on the basis of an extensive pilot study [16]. Selected descriptions for the conflict and no-conflict problems moderately but consistently cued one of the two groups. This point is not trivial. 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., [27]) and combine the base-rates with the diagnostic value of the description, this can lead to complications when the description is extremely diagnostic. For example, imagine that we have a sample of males and females and the description would state that the randomly drawn individual ''gave birth to two children''. Now, by definition, no matter what the base-rates in the sample are, one would always need to conclude that the person is a woman. We limited the impact of this problem by only selecting descriptions that were judged to have moderate diagnostic value. Given these restrictions one may generally conclude that the response that is cued by the base-rates should be selected if participants manage to refrain from giving too much weight to the intuitive answer cued by the description.

To make sure that the contrast between conflict and no-conflict problems was not affected by the selected material, the descriptions for the conflict and no-conflict problems were completely crossed. That is, problems that were presented as conflict problems to one half of the participants were presented as no-conflict problems to the other half of the participants (and vice versa) by switching the base-rates around. The order of the two response options ('a' and 'b') was also 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.

Each problem was presented on a separate page in a booklet. After participants had solved a problem they found a rating scale ranging from 0% (completely unconfident) to 100% (completely confident) with 5% units (see Figure S1 for an example) and the following instructions on the next page:

> Bellow you find a scale from 0% to 100%. Please indicate how confident you are that the answer you just gave was the right one. Circle the number that matches your feeling of confidence:

Procedure. Participants were tested at the same time during a regular course break. On the first page of the booklet they received the following general instructions:

> 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 six base-rate problems were presented in one of four pseudo-random orders. We made sure that half of the presented booklets started with a conflict problem, while the other half started with a no-conflict problem.

Experiment 2: Conjunction fallacy task

In Experiment 2 we investigated the generality of our findings by testing the same hypotheses with a different reasoning task. Participants were presented with problems that were based on the classic conjunction-fallacy task (e.g., the ''Linda-Problem'', see [23]). Together with the base-rate task, the conjunction fallacy is probably one of the most popular examples of the biasing impact of heuristics on people's decision-making. In the task people typically read a short personality sketch, for example, 'Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.' Participants are then asked to rank statements according to their probability, for example '(A) Linda is a bank teller', and '(B) Linda is a bank teller and is active in the feminist movement'.

The conjunction rule, the simplest and most fundamental law of probability [23], holds that the probability of a conjunction of two events cannot exceed that of either of its constituents (i.e., $p(A\&B) \leq p(A), p(B)$). Thus, there should always be more individuals that are simply bank tellers than individuals that are bank teller and in addition also active in the feminist movement. However, people typically violate the conjunction rule and intuitively conclude that statement B is more probable than statement A based on the intuitive match with the stereotypical description.

As in Experiment 1, we presented people with both the classic conflict versions and newly constructed no-conflict control problems. After each problem people were again asked to indicate their response confidence.

Participants. A total of 147 undergraduates who were taking an introductory psychology course at the University of Leuven (Belgium) participated in return for course credit. None had participated in Experiment 1. Participants provided written informed consent and the study was approved by the local ethics committee of the University of Leuven.

Material. Participants solved two conjunction problems each. In each problem participants first read a short personality description of a character (based on the classic "Linda" or "Bill" descriptions, see [23]). Next, they were given two statements about the character and were asked to indicate which one of the two was most probable. One statement always consisted of a conjunction of two characteristics (one characteristic that was likely given the description and one that was unlikely). The other statement contained only one of these characteristics. Consider the following example:

Bill is 34. He is intelligent, punctual but unimaginative and somewhat lifeless. In school, he was strong in mathematics but weak in social studies and humanities.

Which one of the following statements is most likely?

a. Bill plays in a rock band for a hobby
b. Bill is an accountant and plays in a rock band for a hobby

We manipulated the conflict nature of the problems by changing the content of the non-conjunctive statement. In the classic conflict versions we presented the unlikely characteristic (e.g., Bill plays in a rock band for a hobby) as the non-conjunctive statement (see example above). In the no-conflict versions we presented the likely characteristic (e.g., Bill is an accountant) as non-conjunctive statement (see example bellow):

Bill is 34. He is intelligent, punctual but unimaginative and somewhat lifeless. In school, he was strong in mathematics but weak in social studies and humanities.

Which one of the following statements is most likely?

a. Bill is an accountant
b. Bill is an accountant and plays in a rock band for a hobby

Intuitively, people will tend to select the statement that best fits with the stereotypical description (i.e., the most representative statement, see [23]). Clearly, the fit will be higher for the likely than the unlikely characteristic with the conjunctive statement falling in between. Normative considerations based on the conjunction rule always cue selection of the non-conjunctive statement. Hence, on our no-conflict problems both intuition and normative considerations will cue selection of the non-conjunctive response whereas people will be intuitively tempted to pick the conjunctive statement on the conflict problems.

Each participant solved one conflict and one no-conflict problem. To make sure that the content of the problems did not affect the findings we crossed the scenario content and conflict status. For half of the participants the conflict problem was based on the Bill scenario and the no-conflict problem on the Linda Scenario (and vice versa for the other half). As in Experiment 1, the order of the two response options ('a' and 'b') was also counterbalanced. For one of the problems the correct response (i.e., the non-conjunctive statement) was option 'a' whereas for the other problem the second response option ('b') was the correct one.

Each problem was presented on a separate page in a booklet and followed by the same confidence rating scale as in Experiment 1.

Procedure. Participants were tested at the same time during a regular course break. As in Experiment 1 we made sure that half of the presented booklets started with a conflict problem, while the other half started with a no-conflict problem. The scenario content of the first problem was also counterbalanced.

Experiment 3: Developmental study

Participants. A total of 109 young (Mean age = 13.14 year, SE = .10) and 126 late (Mean age = 16.32, SE = .08) adolescents volunteered to participate. Young adolescents were recruited from a suburban middle school and the late adolescents were students at an associated high school. Informed consent was obtained from the participants' parents or guardian. The study was approved by the local ethics committee of the University of Leuven and the school boards.

Material. All participants were presented with one booklet with four base-rate problems and one booklet with four conjunction problems. Half of the problems in each booklet were conflict problems and the other half no-conflict control problems. Problems were constructed as in Experiment 1 and 2 with the same randomization procedures, instructions, and confidence rating scales. The only difference was the exact content of the problems. The materials were selected based on a pilot study [28] in which young and late adolescents rated the stereotypicality of a large number of descriptions. We made sure to select stereotypical descriptions and characteristics that were familiar for both age groups. A complete overview of all problems can be found in the Appendix S1.

Procedure. Participants were tested during a standard one-hour course break in which they remained in their classroom. Participants were presented with two booklets. Half of the participants started with the conjunction booklet and the other half with the base-rate booklet. Participants were given a five minute break after they finished solving the first booklet.

Results

Experiment 1: Base-rate task

Accuracy. The accuracy on the base-rate problems replicated the findings in previous studies (e.g., [13,15]). Participants seemed to neglect the base-rate information and erred on the vast majority of the conflict problems. On average, only 20% (SE = 1.81) of these problems were solved correctly. Also, as expected, people had few difficulties when intuitive beliefs and base-rates pointed towards the same conclusion. Correct response rates on the no-conflict control problems reached 95% (SE = .83), $F(1, 246) = 1443.54$, $p < .0001$, $\eta^2_p = .85$.

Response confidence. Our main question concerned people's decision confidence. If despite the poor performance on the conflict problems, people detect that their intuitive response conflicts with the base-rates, and know that their answer is questionable, then their confidence should be affected. As Figure 1 shows, overall confidence ratings were indeed about 10% lower for the classic conflict problems than for the control no-conflict problems, $F(1, 246) = 54.98$, $p < .0001$, $\eta^2_p = .18$. Recall that the only difference between the conflict and no-conflict problems is the (in)consistency of the cued intuitive and base-rate response. If this intrinsic conflict was not detected or merely epiphenomenal, confidence ratings for the two types of problems should not have differed.

Although people typically erred on the conflict problems, on some occasions people did manage to give a correct response. A proponent of the bias-as-detection-failure view might therefore argue that it is those responses that are driving the overall confidence effect. Hence, it is still possible that there is no actual confidence effect for the intuitive responses. To eliminate such a possible confound we ran a separate analysis that was restricted to confidence ratings for incorrectly solved conflict problems. Results

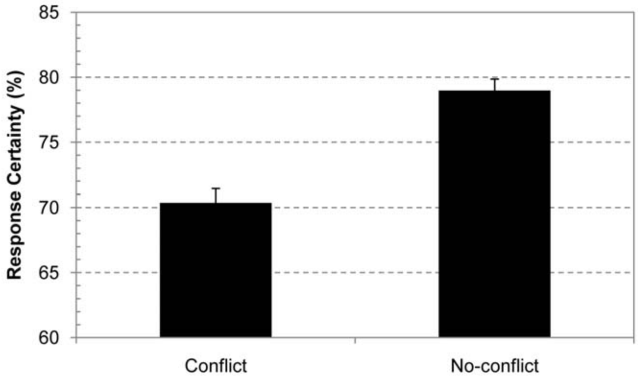

Figure 1. Response confidence for conflict and no-conflict base-rate problems. Average response confidence after solving conflict and no-conflict base-rate problems. Error bars are standard errors.
doi:10.1371/journal.pone.0015954.g001

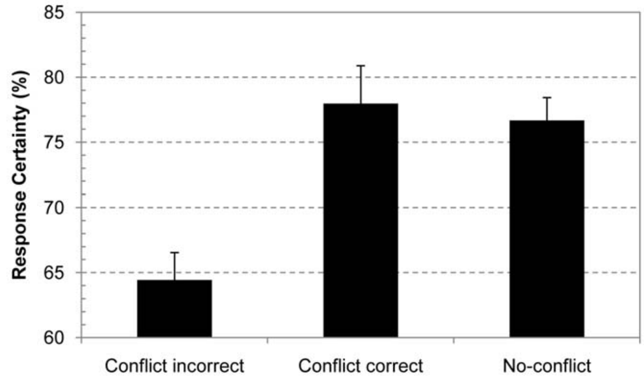

Figure 2. Response confidence for first-presented base-rate problem. Average response confidence for different types of responses on the first presented base-rate problem. Error bars are standard errors.
doi:10.1371/journal.pone.0015954.g002

showed that we found the same 10% confidence decrease for the incorrectly solved conflict problems as in the overall analysis, $F(1, 230) = 59.35$, $p < .0001$, $\eta^2_p = .21$. Note that in an additional control analysis we also made sure to remove the few trials in which the no-conflict problems were not solved correctly. However, as with all confidence analyses in the present study that took response accuracy into account, results were not shown to be affected by the elimination of these trials.

A second issue we need to address is the within-subject nature of the present conflict manipulation and the impact of possible learning effects. The initial studies that started focusing on conflict detection during thinking were typically quite lengthy. For example, in their fMRI study De Neys et al. [15] presented almost 100 problems. One might argue that the repeated presentation and repetitive nature of these studies cued participants to start paying attention to the conflict manipulation. Hence, the detection findings in these studies might simply be an artifact of learning. Note that we already reduced the number of presented items in the present study to limit the impact of such a learning confound. However, it has been argued that the purest test case in this respect concerns a between-subject experiment in which each subject solves only one single problem [29,30]. To address this issue we ran an additional analysis in which we included only the confidence rating of the first problem that each participant solved (recall that this was a conflict problem for half of the participants and a no-conflict problem for the remaining half). As Figure 2 shows, results replicated the main finding of the overall analysis: The group of people who gave an intuitive response on the conflict problem were significantly less confident about their decision than the group of people who solved the no-conflict problem, $F(1, 192) = 18.86$, $p < .0001$, $\eta^2_p = .09$. This establishes that the observed overall confidence decrease on the conflict problems does not result from a learning confound.

In the present study we were less concerned with confidence ratings for correctly solved conflict problems per se. The typical low accuracy rates on the conflict trials imply that the ratings for correctly solved conflict problems will be based on a small number of observations which might compromise the reliability of the data. Nevertheless, with this caution in mind, for completeness we also examined the confidence data of the group of people who gave the correct base-rate response on the crucial first conflict problem and included these in Figure 2. As the figure indicates, for people who solved the conflict problem correctly, confidence ratings did not seem to differ from the no-conflict ratings, $F(1, 154) < 1$. This does suggest that good reasoners who reason in line with the normative standards also seem to know that they are right and show high response certainty. By itself this does not come as a surprise since after being confronted with the initial conflict these people manage to override the intuitive response and resolve the conflict. Nevertheless, as we noted, caution is needed when interpreting findings for the infrequent correct conflict responses. The main question in the present study concerns the confidence ratings for the common incorrect conflict responses. The decreased confidence on these problems compared to no-conflict control problems supports the claim that biased reasoners detect that their intuitive response on the classic conflict problems conflicts with the cued normative response.

Experiment 2: Conjunction fallacy task

Accuracy. Participants' accuracy on the conjunction problems was as expected. In line with previous findings [23], the vast majority of participants committed the conjunction fallacy on the classic conflict problems. Correct response rates reached only 24% (SE = 3.5). However, as with the base-rate problems in Experiment 1, performance on the no-conflict control versions was much better with almost 96% (SE = 1.6) correct responses, Wilcoxon matched pairs test, $n = 147$, $Z = 8.73$, $p < .0001$.

Response confidence. As Figure 3 shows, the confidence results nicely replicated the findings with the base-rate problems in Experiment 1. Despite the low accuracy, overall confidence ratings were again about 10% lower for the classic conflict problems than for the control no-conflict problems, $F(1, 146) = 24.49$, $p < .0001$, $\eta^2_p = .14$. As in Experiment 1, this effect was equally clear when the analysis was restricted to incorrectly solved conflict trials, $F(1, 106) = 13.72$, $p < .0005$, $\eta^2_p = .12$.

Finally, we also restricted the analysis to the first presented item and contrasted the confidence of the group of people who gave an incorrect conflict response and the confidence of people who solved a no-conflict control problem first. Figure 4 shows the results. Despite the smaller sample size the confidence effect was still marginally significant in this between-subject analysis, $F(1, 120) = 2.85$, $p < .095$, $\eta^2_p = .02$. As in Experiment 1, the between-subject confidence contrast on the first item was not significant for the correctly solved conflict items, $F(1, 89) = 1.77$, $p = .19$.

Experiment 3: Developmental study

Experiment 1 and 2 established that biased reasoners showed decreased confidence in their answers after solving conflict

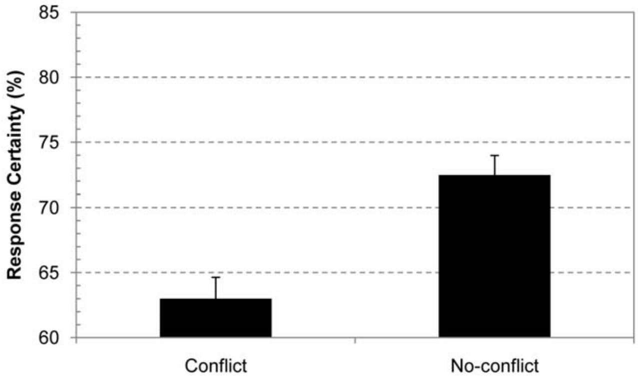

Figure 3. Response confidence for conflict and no-conflict conjunction problems. Average response confidence after solving conflict and no-conflict conjunction problems. Error bars are standard errors.
doi:10.1371/journal.pone.0015954.g003

problems. Consistent with the flawless detection view, this suggests that biased reasoners at least acknowledge that their intuitive answer is questionable. In Experiment 3 we tried to validate the findings by testing the confidence contrast for conflict and no-conflict problems in a group of young and late adolescents. Given that elementary conflict monitoring skills are not fully developed before late adolescence (e.g., [25,26]) we predicted that conflict detection during thinking will be less successful for the youngest reasoners. If young adolescents do not yet detect that their intuitive response conflicts with the cued normative response, they should not treat the conflict and no-conflict problems any differently and show similar confidence in their responses for both types of problems. Therefore, the decreased confidence after solving conflict problems should be far less pronounced for early than for late adolescents.

Accuracy. We ran a 2 (Conflict; conflict or no-conflict problem)×2 (Task; base-rate or conjunction task)×2 (Age Group; young or late adolescents) mixed model ANOVA on the mean accuracy scores. The first two factors were within-subjects factors and the Age Group was a between-subjects factor. Results showed that there was a main effect of the Conflict factor. Just as with the adults in Experiment 1 and 2, accuracy was near perfect on the no-conflict problems but significantly lower on the classic conflict problems, $F(1, 233) = 2371.46$, $p < .0001$, $\eta^2_p = .91$. There was also a main effect of Age group, $F(1, 233) = 8.03$, $p < .01$, $\eta^2_p = .03$, and the Age and Conflict factors interacted, $F(1, 233) = 4.56$, $p < .05$, $\eta^2_p = .02$. Planned contrasts showed that age did not affect accuracy on the no-conflict problems, $F(1, 233) < 1$, but performance on the conflict problems did increase slightly for late adolescents, $F(1, 233) = 4.56$, $p < .05$, $\eta^2_p = .02$. However, despite the developmental increase even the oldest age group was typically biased with accuracies on the conflict problems below 20%.

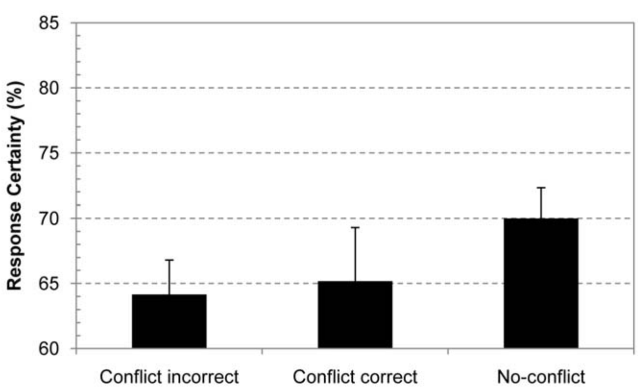

Figure 4. Response confidence for first-presented conjunction problem. Average response confidence for different types of responses on the first presented conjunction problem. Error bars are standard errors.
doi:10.1371/journal.pone.0015954.g004

The accuracy findings were very similar for the base-rate and conjunction problems. Neither the Task factor nor any of its interactions with the other factors reached significance. A complete overview of the accuracy findings can be found in Table 1.

Response confidence. We also ran a 2 (Conflict; conflict or no-conflict problem)×2 (Task; base-rate or conjunction)×2 (Age Group; young or late adolescents) mixed model ANOVA on the mean confidence ratings. Figure 5 shows the results. There was a main effect of Conflict with overall lower confidence ratings for the conflict than for the no-conflict problems, $F(1, 233) = 78.75$, $p < .0001$, $\eta^2_p = .26$. However, as predicted, this effect interacted with Age Group, $F(1, 233) = 12.84$, $p < .0005$, $\eta^2_p = .05$. Although the conflict contrast was significant for both young, $F(1, 233) = 13.05$, $p < .0005$, $\eta^2_p = .05$, and late adolescents, $F(1, 233) = 83.64$, $p < .0001$, $\eta^2_p = .26$, the confidence decrease was much smaller in the youngest age group (i.e., 4% vs. 10%, $t(233) = 3.58$, $p < .0005$, $d = .47$). The main effect of Age Group was not significant, $F(1, 233) < 1$.

There was also a main effect of the Task factor, $F(1, 233) = 78.07$, $p < .0001$, $\eta^2_p = .25$. As Figure 5 shows, confidence ratings for the base-rate problems seemed to be overall higher than ratings for the conjunction problems. However, neither the higher-order interaction between the Task, Conflict and Age factors, $F(1, 233) < 1$, nor any of the other interactions with the Task factor reached significance. As Experiment 1 and 2 already suggested, this establishes that the impact of conflict on the confidence measure is very similar in the two types of tasks. This consistency across reasoning tasks further supports the generality of the findings.

We also repeated the above analysis but made sure to exclude confidence ratings for correctly solved conflict trials. As in Experiment 1 and 2, the pattern remained unchanged. There was a significant main effect of the Conflict, $F(1, 206) = 78.59$, $p < .0001$, $\eta^2_p = .28$, and Task, $F(1, 206) = 55.77$, $p < .0001$, $\eta^2_p = .21$, factors. Once again, the conflict effect was less pronounced in the youngest age group, $F(1, 206) = 9.07$, $p < .005$, $\eta^2_p = .04$. Other effects and interactions were not significant.

Finally, we also ran a between-subjects analysis on the confidence ratings for the first presented problem. The analysis focused on the contrast between the confidence ratings of the group of students who failed to solve the first conflict problem and those who solved a no-conflict problem (given that there were only six out of 109 young adolescents who responded correctly on the first presented conflict item we refrained from analyzing these confidence responses, see Table 1 for complete overview). The confidence data was entered in a 2 (Conflict; incorrect conflict or no-conflict problem)×2 (Task; base-rate or conjunction)×2 (Age Group; young or late adolescents) between-subjects ANOVA. The pattern for the first item was consistent with the overall analysis. There was a main effect of the Task, $F(1, 201) = 10.53$, $p < .005$, $\eta^2_p = .05$, and Conflict factors, $F(1, 201) = 10.02$, $p < .005$, $\eta^2_p = .05$, and the Conflict and Age Group factors tended to interact, $F(1, 201) = 3.01$, $p < .085$, $\eta^2_p = .02$. Other effects and

Table 1. Overall accuracy and response confidence on the first item in two age groups.

		Base-rate task				Conjunction fallacy task			
		Young adolescents		Late Adolescents		Young adolescents		Late Adolescents	
Measure	Problem	Mean (SE)	n	Mean (SE)	n	Mean (SE)	n	Mean (SE)	n
Accuracy	Conflict	7 (2.5)	109	16 (2.3)	126	11 (2.8)	109	18 (2.7)	126
	No-conflict	97 (1.3)	109	95 (1.2)	126	92 (1.6)	109	96(1.5)	126
Confidence	Conflict incorrect	83 (4.2)	21	67 (4.0)	22	66 (3.7)	27	63 (3.5)	30
	No-conflict	82 (3.8)	25	82 (3.6)	28	73 (3.9)	30	75 (3.8)	35
	Conflict correct	59 (11.4)	4	73 (9.3)	6	55 (16.1)	2	46 (10.2)	5

doi:10.1371/journal.pone.0015954.t001

interactions were not significant. The Conflict×Age Group interaction is illustrated in Figure 6. Planned contrasts showed that the conflict contrast was significant for the oldest age group, $F(1, 201) = 5.62$, $p < .025$, $\eta^2_p = .03$, but not for the young adolescents, $F(1, 201) < 1$. Hence, on the first item confidence of young adolescents did not yet decrease when they gave a biased conflict response. This suggests that contrary to older reasoners, young adolescents do not yet detect that their intuitive response is unwarranted.

Discussion

Consistent with decades of reasoning and decision making research, reasoning accuracies in the present study showed that people are typically biased and fail to select the normatively correct response on classic reasoning problems. However, our confidence measure indicated that despite this resounding bias, adults and older adolescents are detecting that their intuitive response is questionable. Three experiments established that reasoners' decision confidence on classic conflict problems was consistently lower than their confidence on the control no-conflict problems. The only difference between the conflict and no-conflict problems was that the cued intuitive response conflicted with traditional normative considerations on the classic versions. If reasoners were not detecting this conflict or the detection was merely epiphenomenal, their response confidence should not have decreased. This establishes that although people do typically not manage to discard a biased intuitive answer, they at least seem to be aware that their intuitive response is not fully warranted. Our developmental evidence in Experiment 3 suggested that it is precisely this bias awareness that younger reasoners lack.

The confidence findings help to clarify the nature of heuristic bias and validate the flawless detection view. We noted that although people hardly ever explicitly refer to normative considerations during reasoning, more implicit detection measures such as the activation of the anterior cingulate cortex or autonomic skin-conductance levels already indicated that our brain is sensitive to the presence of conflict between cued intuitive and normative considerations (e.g., [15,17]). The present findings establish that this detection signal is not epiphenomenal. Giving an intuitive response that conflicts with more normative considerations does not simply result in some fancy brain-activation but

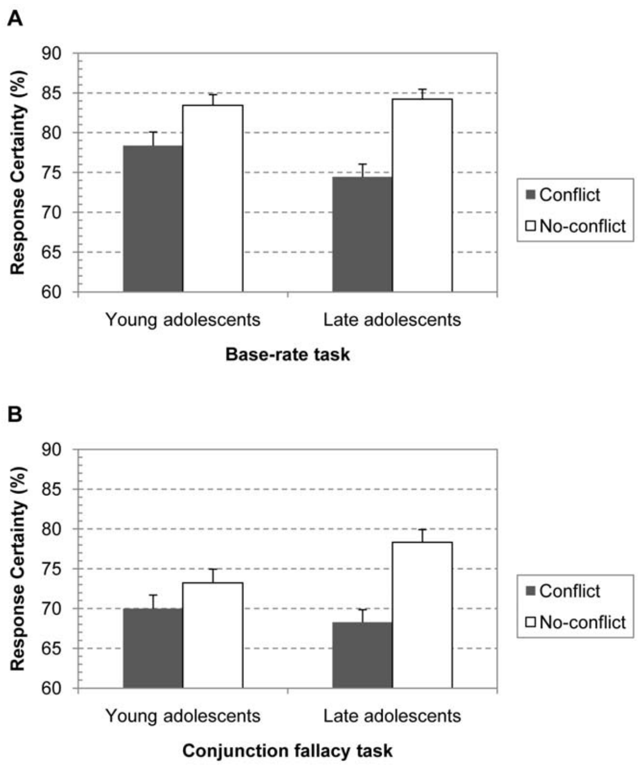

Figure 5. Response confidence in different age groups. Average response confidence after solving conflict and no-conflict base-rate (A) and conjunction (B) problems in the different age groups. Error bars are standard errors.
doi:10.1371/journal.pone.0015954.g005

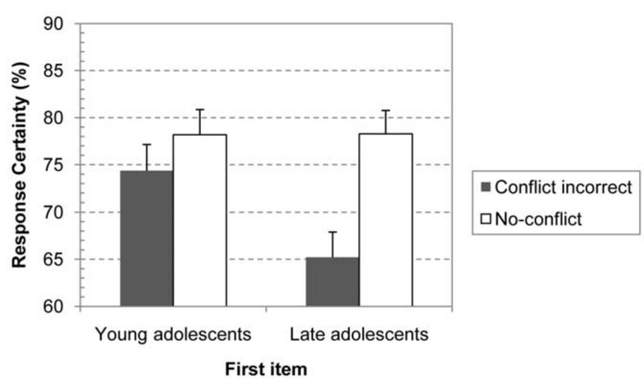

Figure 6. Response confidence for first problem in different age groups. Developmental impact on the response confidence of incorrect conflict responses on the first presented problem. Error bars are standard errors.
doi:10.1371/journal.pone.0015954.g006

directly affects our judgment. People literally indicate that their intuitive response is not fully warranted. Clearly, the well-established lack of explicit verbalization suggests that this knowledge is implicit in nature. People will not manage to label and identify the exact normative principles that are being violated. However, whenever their intuitive answer conflicts with more normative considerations they do seem to acknowledge that their response is questionable. The fact that this conflict is affecting their judgment implies that reasoners at least implicitly adhere to these normative principles.

At a more general level our findings help to sketch a more optimistic view of human rationality. Note that over the decades, the continuous confrontation with the strikingly low accuracy of educated adults on the classic reasoning tasks has led researchers to question human rationality and traditional normative standards [8,31]. In a nutshell, some researcher argued that the widespread bias implied that humans are illogical and irrational intuitive reasoners (e.g., [32,33]). Others argued that the low accuracy pointed to the invalidity of the traditional logical or probabilistic normative rules [34–36]. According to this latter view, humans are adhering to other norms than the traditional normative logical standards when solving classic reasoning tasks. People would interpret tasks such as the base-rate or conjunction fallacy task as a type of social classification problem in which they try to determine to which social group a character belongs. Given this alternative task interpretation the intuitive response would be perfectly valid. These issues have resulted in a debate that has raged through the field for decades without clear solution [8]. The present findings shed light on this and support a conclusion that might help to save human rationality and the traditional normative standards: The lower confidence implies that people are at least implicitly taking the normative principles into account when solving the classic conflict problems. If adult reasoners would not master the normative principles or would not consider these to be relevant, there would be nothing to conflict with their responses, and so people's response confidence should not be affected. It has previously been argued that the whole rationality discussion in the reasoning field has been biased by an almost exclusive focus on accuracy rates and the output of the reasoning process (e.g., [37–40]). The present work underscores this point and indicates that if we scratch below the accuracy surface, people are more normative than their biased responses suggest.

This being said, it is important to address some potential critiques with respect to our study. As we stated, our findings imply that people show some minimal sensitivity to base-rates and the conjunction rule in classic reasoning tasks. One might wonder whether this point has been demonstrated in past studies. It is true that a number of manipulations and interventions (e.g., making base-rates more extreme or making the description less diagnostic) have been shown to increase people's reasoning accuracy (e.g., [41]). This indicates that it is possible to have people select the correct response and take base-rates into account, for example. However, that is not the issue here. The question is: Are people taking the base-rates and conjunction rule into account when they give an intuitive response? This question cannot be answered by looking at accuracy rates per se. Indeed, even if, for example, people show perfect accuracy when the base-rates are made more extreme, this can never establish whether or not they were taken into account initially. This is precisely the reason why the diametric accounts on conflict detection persist in the reasoning and decision-making literature. The present confidence data and study design are critical to address this question.

We do believe that there is an interesting link between the present findings and an earlier study on metacognitive uncertainty during syllogistic reasoning by Quayle and Ball [42]. These authors observed that although people often judged invalid syllogistic conclusions to be valid, their subjective confidence ratings for these erroneous judgments were typically lower than for valid problems. Although Quayle and Ball did not manipulate the conflict nature of their problems, the results do seem to fit with the basic idea that people are sensitive to normative violations and might be more logical than their erroneous responses suggest. This strengthens the generality of our claims with respect to the validity of traditional normative standards.

In our work we have been specifically contrasting the lax and flawless views on conflict detection during thinking. We noted that the present confidence findings are consistent with the flawless detection view. However, one might want to consider alternative conceptualizations. For example, the present findings also fit with a "weighing view". The idea behind the weighing view is that people are simply weighing competing arguments when solving the conflict problems. People would consider the normative response on the conflict problems, find it unpersuasive or weaker than the intuitive response and therefore go with the intuitive response. The flawless detection view entails that people notice that their intuitive response conflicts with the normative response, try to block it but fail to do so because of the compelling nature of the intuitive response. The weighing view also entails that people experience a conflict, but suggests that precisely because people find the intuitive response so compelling, they simply see no need to engage in an inhibition process. Hence, the difference between the two views lies in the postulation of an additional inhibition process.

It is important to stress that the flawless detection and weighing views make similar claims with respect to reasoners' conflict sensitivity and subjective knowledge state. Note that the flawless detection view does not entail that biased reasoners are 100% convinced that the normative response is correct. The whole point is that people will be in doubt. If people detect the conflict and this has any functional impact on their reasoning process, they should show decreased response confidence for intuitive responses on the conflict (vs. no-conflict) problems. This implies that the normative considerations have a minimal impact on people's judgment. Hence, it does not necessarily need to be the case that people consider the intuitive response less appropriate than the normative response per se. The point is that reasoners consider the intuitive response less compelling than the intuitive response on the no-conflict problems. In this respect the flawless detection and weighing view are consistent and point to the same implications: If reasoners decide after weighing to go with the intuitive response, the weighing at least implies that the normative information has been given some minimal consideration. If people would find the normative response on the conflict problems completely unconvincing, their response confidence should not be affected.

For completeness, one might note that the postulation of an additional inhibition process has gained some credence from recent findings. For example, De Neys and Franssens [14], probed memory activations after reasoning to examine the inhibition process. In their study participants solved conflict and no-conflict versions of the base-rate problems (and related syllogistic reasoning problems). After each problem participants were presented with a lexical decision task in which they had to judge whether a presented letter string was a word or not. Half of the presented words were strongly associated with the intuitive response that was cued in the reasoning problem. Results showed that lexical decision times for these target words were longer after solving conflict vs. no-conflict problems. This classic inhibition effect was less pronounced but still significant when people gave the intuitive response on the conflict problems. This seems to argue against a mere weighing view. If people were not at least engaging in an inhibition process and tried to discard the intuitive response it becomes harder to explain why words that are closely associated with the cued intuition become less accessible in memory after the reasoning task (i.e., less accessible than after solving no-conflict problems). Nevertheless, it should be clear that with respect to reasoners' subjective knowledge state the flawless detection view is consistent with a weighing view. Both views entail that the intuitive response should be less compelling on conflict problems than when it does not conflict with normative considerations on the no-conflict problems. This implies that even when people give an intuitive response on the classic conflict problems they give some minimal weight to normative considerations such as the conjunction rule or the role of the base-rates. It is this critical norm sensitivity that the present confidence data establish.

We stated that our present confidence findings fit with the early flawless detection claims by Sloman and Epstein [4,6,43]. It should be noted, however, that the claims of these authors were rooted in specific dual process models of reasoning. For example, Sloman [6] has suggested that people will detect conflicts because they always simultaneously engage in more automatic intuitive processing and demanding analytic-logical processing. One implication of this view is that the detection is assumed to result from time-consuming and resource demanding analytic computations. For completeness, we should stress that we do not subscribe to these further dual process assumptions. The present confidence findings imply that people are taking traditional normative principles into account when solving the classic conflict problems. However, there is no need to assume that the activation of these principles itself is especially demanding in cognitive terms. We have pointed to a number of theoretical paradoxes associated with this assumption [13] (see also [44]), and have provided empirical data that indicates that the detection process is indeed quite effortless [16]. The interested reader can find an extensive discussion of the implications of our findings for dual process theories in De Neys and Glumicic [13]. The basic point we want to note here is that while we agree with Sloman and Epstein that detection is flawless, we do not necessarily share their specific dual process assumptions as put forward in their original models.

To avoid possible misinterpretations it is perhaps also informative to underline that our claims with respect to the norm validity are situated at the psychological processing level. Our study indicates that people are sensitive to violations of traditional norms during thinking. As we explained, this finding argues against the claim that people consider these traditional norms to be irrelevant for their judgment. However, clearly, the fact that people adhere to a certain norm does not by itself entail that the norm is valid. From an epistemological/philosophical point of view, it might still be that other norms are more appropriate. In other words, our claim with respect to the validity of traditional norms does not entail that these norms are ultimately correct, but rather that human reasoners *consider* them to be correct. It is this demonstrated adherence to the traditional normative principles that is crucial to counter the idea that people do not know these principles or do not consider them relevant to solve classic reasoning problems.

Finally, we would like to highlight that the present study might have interesting implications for the developmental field. Just as with the debate on human rationality, the apparent omnipresence of intuitive bias resulted in quite pessimistic developmental views. As Markovits and Barrouillet [45] noted, the demonstration of the widespread bias in human reasoning since the 1960s seemed to point to a developmental standstill in human reasoning (see [46] for studies criticizing this idea). In other words, if the vast majority of educated university students fail to solve basic reasoning problems, there surely does not seem to be a lot of development going on. At first sight, our developmental study might have seem to strengthen this conclusion. Although there was a slight performance increase when contrasting early and late adolescents' accuracy rates, even in late adolescence accuracy was only proximately 15%. However, looking closely at the conflict detection process and the confidence data suggests that the lack of development is more apparent than real. Although both adults and adolescents are indeed biased most of the time, our findings indicate that a possible important difference between the age groups is that adults at least detect that their responses are biased. Consistent with recent insights in the developmental field (e.g., [39,46–48]) this differential bias awareness argues against the idea of a developmental standstill in human reasoning. Nevertheless, our developmental findings will need further validation. For example, although our confidence measure allowed us to document the differential bias awareness, it is not clear whether younger adolescents also lack the implicit neuronal conflict signal or merely its translation into a decreased response confidence (i.e., it might be that younger adolescents also showed implicit conflict-related brain activity but this activity might still be epiphenomenal). Clearly, directly studying the conflict-related brain activity of younger reasoners in an fMRI study would be very useful in this respect. Likewise, it would be interesting to further clarify whether the lack of conflict awareness primarily results from limited basic conflict monitoring skills per se or whether it is also affected by a possible less developed normative knowledge (e.g., see [28]). These outstanding questions will need to be addressed in more focused and fine-grained developmental studies.

In sum, the present paper indicated that although human reasoners might typically fail to refrain from giving biased responses, they do seem to acknowledge that their intuitive responses are not fully warranted. This implies that at least by the end of adolescence, human reasoners are more sensitive to normative standards than the historical omnipresence of the intuitive response bias suggests.

Supporting Information

Appendix S1 Overview of the problem content in Experiments 1 and 3.
(DOC)

Figure S1 Example of the confidence rating scale
(TIF)

Author Contributions

Conceived and designed the experiments: WDN MO. Performed the experiments: SC WDN. Analyzed the data: WDN SC. Wrote the paper: WDN MO.

References

1. Evans J (2003) In two minds: dual-process accounts of reasoning. Trends in Cognitive Sciences 7: 454–459.
2. Kahneman D, Tversky A (1973) On the Psychology of Prediction. Psychological Review 80: 237–251.
3. Kahneman D, Frederick S (2002) Representativeness revisited: Attribute substitution in intuitive judgement. In: Gilovich T, Griffin D, Kahneman D, eds. Heuristics and Biases: The Psychology of Intuitive Judgement. New York: Cambridge University Press. pp 49–81.

4. Denes-Ray V, Epstein S (1994) Conflict between intuitive and rational processing: When people behave against their better judgment. Journal of Personality and Social Psychology 66: 819–829.
5. Houde O (2007) First insights on "neuropedagogy of reasoning". Thinking & Reasoning 13: 81–89.
6. Sloman SA (1996) The empirical case for two systems of reasoning. Psychological Bulletin 119: 3–22.
7. De Neys W (2009) Beyond response output: More logical than we think. Behavioral and Brain Sciences 32: 87–88.
8. Stanovich KE, West RF (2000) Individual differences in reasoning: Implications for the rationality debate? Behavioral and Brain Sciences 23: 645–726.
9. Evans J (2007) On the resolution of conflict in dual process theories of reasoning. Thinking & Reasoning 13: 321–339.
10. Evans J (2008) Dual-processing accounts of reasoning, judgment, and social cognition. Annual Review of Psychology 59: 255–278.
11. Evans JSBT, Over DE (1996) Rationality and reasoning. Hove, UK: Psychology Press.
12. Wason PC, Evans J (1975) Dual processes in reasoning. Cognition 3: 141–154.
13. De Neys W, Glumicic T (2008) Conflict monitoring in dual process theories of reasoning. Cognition 106: 1248–1299.
14. De Neys W, Franssens S (2009) Belief inhibition during thinking: Not always winning but at least taking part. Cognition 113: 45–61.
15. 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.
16. Franssens S, De Neys W (2009) The effortless nature of conflict detection during thinking. Thinking & Reasoning 15: 105–128.
17. De Neys W, Moyens E, Vansteenwegen D (2010) Feeling we're biased: Autonomic arousal and reasoning conflict. Cognitive, Affective, & Behavioral Neuroscience 10: 208–216.
18. Bonner C, Newell BR (2010) In conflict with ourselves? An investigation of heuristic and analytic processes in decision making. Memory & Cognition 38: 186–196.
19. Stupple EJN, Ball LJ (2008) Belief-logic conflict resolution in syllogistic reasoning: Inspection-time evidence for a parallel-process model. Thinking & Reasoning 14: 168–181.
20. Thompson VA (2009) Dual process theories: A metacognitive perspective. In: Evans JBST, Frankish K, eds. In two minds: Dual processes and beyond. Oxford, UK: Oxford University Press.
21. Kruger J, Dunning D (1999) Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. Journal of Personality and Social Psychology 77: 1121–1134.
22. Shynkaruk JM, Thompson VA (2006) Confidence and accuracy in deductive reasoning. Memory & Cognition 34: 619–632.
23. Tversky A, Kahneman D (1983) Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. Psychological Review 90: 293–315.
24. Davies PL, Segalowitz SJ, Gavin WJ (2004) Development of response-monitoring ERPs in 7-to 25-year-olds. Developmental Neuropsychology 25: 355–376.
25. Fitzgerald KD, Perkins SC, Angstadt M, Johnson T, Stern ER, et al. (2010) The development of performance-monitoring function in the posterior medial frontal cortex. Neuroimage 49: 3463–3473.
26. Santesso DL, Segalowitz SJ (2008) Developmental differences in effor-related ERPs in middle- to late-adolescent males. Developmental Psychology 44: 205–217.
27. 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.
28. Teppers E (2009) Ontwikkeling van conflictdetectie tijdens redeneren [development of conflict detection during reasoning]. Leuven, Belgium: University of Leuven.
29. Kahneman D (2000) A psychological point of view: Violations of rational rules as a diagnostic of mental processes. Behavioral and Brain Sciences 23: 681–683.
30. Stanovich KE, West RF (2008) On the relative independence of thinking biases and cognitive ability. Journal of Personality and Social Psychology 94: 672–695.
31. Stein E (1996) Without good reason: The rationality debate in philosophy and cognitive science. Oxford, UK: Oxford University Press.
32. Wason PC (1968) Reasoning about a rule. Quarterly Journal of Experimental Psychology 20: 273–281.
33. Wason PC (1983) Realism and rationality in the selection task. In: Evans JBST, ed. Thinking and reasoning. London: Routledge. pp 45–75.
34. Hertwig R, Gigerenzer G (1999) The 'conjunction fallacy' revisited: How intelligent inferences look like reasoning errors. Journal of Behavioral Decision Making 12: 275–305.
35. Oaksford M, Chater N (2007) Baysian Rationality: The probabilisitc approach to Human Reasoning. Oxford, UK: Oxford University Press.
36. Gigerenzer G (1996) On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). Psychological Review 103: 592–596.
37. De Neys W (2006) Dual processing in reasoning - Two systems but one reasoner. Psychological Science 17: 428–433.
38. Hoffrage U (2000) Why the analyses of cognitive processes matter. Behavioral and Brain Sciences 23: 679–680.
39. Reyna VF, Lloyd FJ, Brainerd CJ (2003) Memory, development, and rationality: An integrative theory of judgement and decision-making. In: Schneider S, Shanteau J, eds. Emerging perspectives on judgment and decision research. New York: Cambridge University Press.
40. Osman M (2004) An evaluation of dual-process theories of reasoning. Psychonomic Bulletin & Review 11: 988–1010.
41. Koehler JJ (1996) The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. Behavioral and Brain Sciences 19: 1–17.
42. Quayle JD, Ball LJ (2000) Working memory, metacognitive uncertainty, and belief bias in syllogistic reasoning. Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology 53: 1202–1223.
43. Epstein S (1994) Integration of the cognitive and the psychodynamic unconscious. American Psychologist 49: 709–724.
44. Evans JBST (2009) How many dual process theories do we need: One, two or many? In: Evans JBST, Frankish K, eds. In two minds: Dual processes and beyond. Oxford, UK: Oxford University Press.
45. Markovits H, Barrouillet P (2004) Introduction: Why is understanding the development of reasoning important? Thinking & Reasoning 10: 113–121.
46. Jacobs JE, Klaczynski PA (2005) The development of judgment and decision making in children and adolescents. Mahwah, NJ: Lawrence Erlbaum.
47. Brainerd CJ, Reyna VF, Ceci SJ (2008) Developmental reversals in false memory: A review of data and theory. Psychological Bulletin 134: 343–382.
48. Reyna VF, Farley F (2006) Risk and rationality in adolescent decision making - Implications for theory, practice, and public policy. Psychological Science. pp 1–44.

Cognition 113 (2009) 45–61

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

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.

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 overrely 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 Elsevier B.V. All rights reserved.
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 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) $\times$ 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.

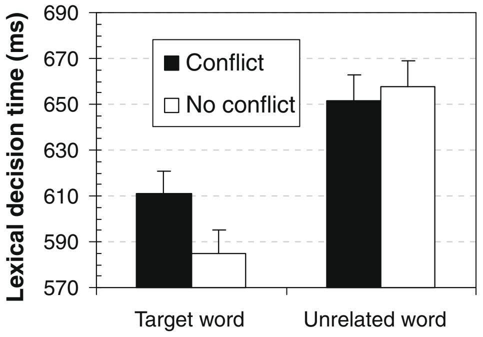

Fig. 1. Average lexical decision time for words that were related (i.e., targets) and unrelated to cued beliefs after solving conflict and no-conflict syllogisms. Error bars are standard errors.

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.

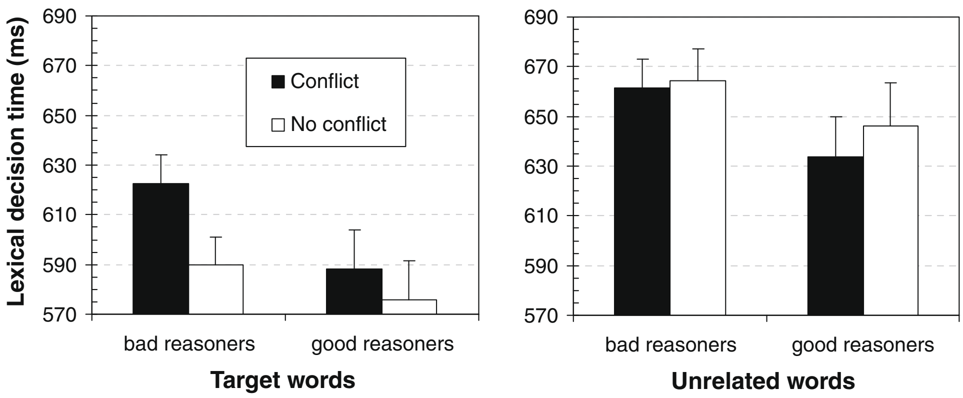

Fig. 2. Lexical decision times for the worst (bad reasoners) and best (good reasoners) scoring half of participants in Experiment 1. Error bars are standard errors.

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

[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.

response option ('b') was the correct one. A complete overview of all eight problems can be found in the Appendix A.

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.

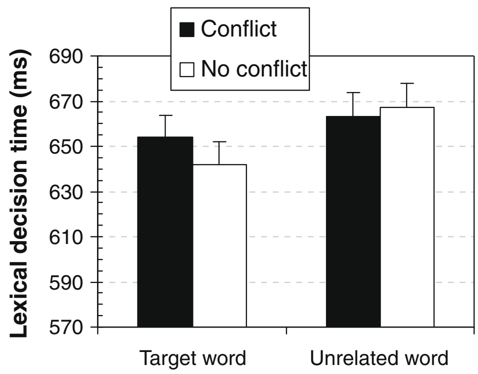

Fig. 3. Average lexical decision time for words that were related (i.e., targets) and unrelated to cued beliefs after solving conflict and no-conflict base-rate problems. Error bars are standard errors.

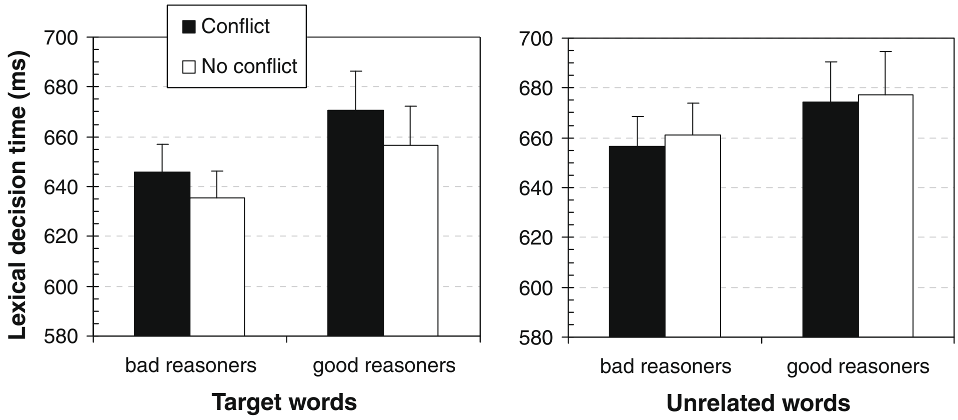

Fig. 4. Lexical decision times for the worst (bad reasoners) and best (good reasoners) scoring half of participants in Experiment 2. Error bars are standard errors.

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

[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, $F(1, 91) < 1$].

presented 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 a 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 -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

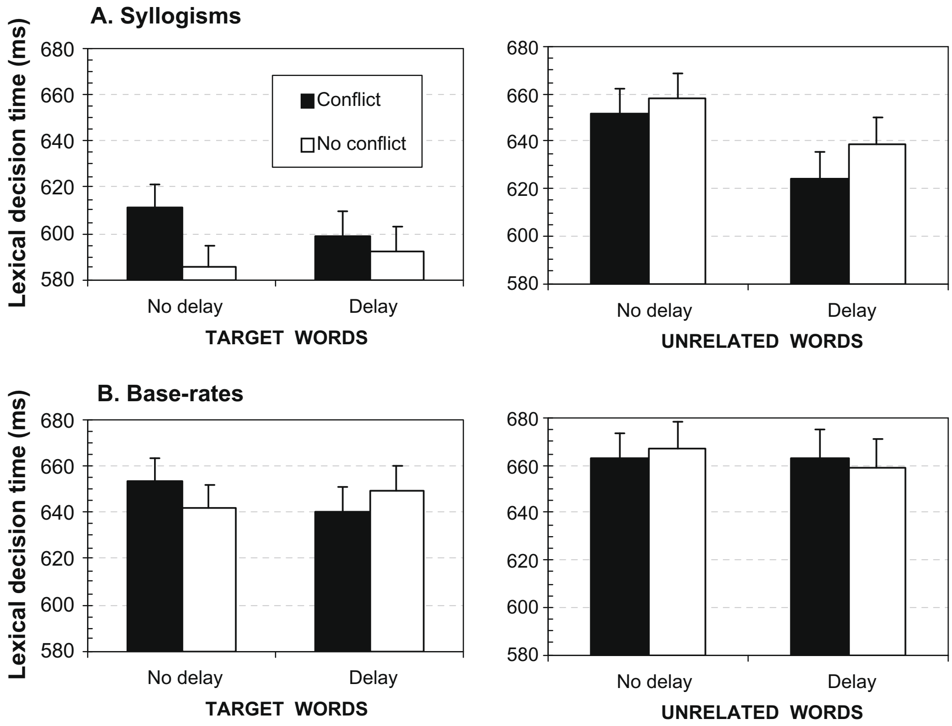

Fig. 5. The impact of a one-minute delay on the average lexical decision times after solving syllogisms (top panel) and base-rate problems (bottom panel). Error bars are standard errors.

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, 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.

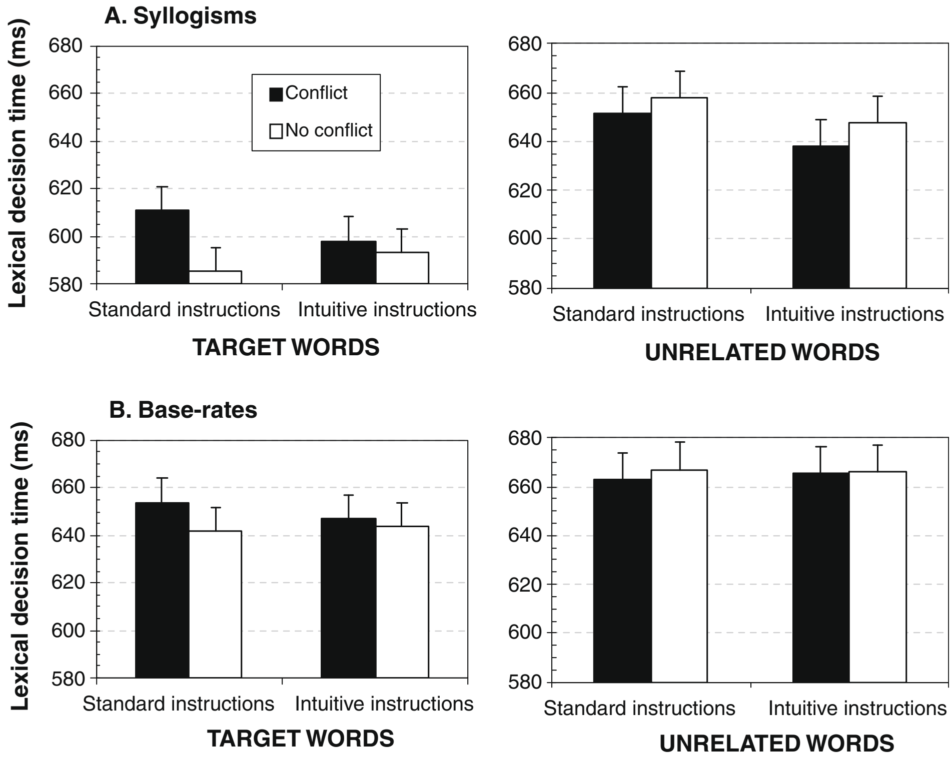

Fig. 6. The impact of the explicit instruction to think intuitively on lexical decision times after solving syllogisms (top panel) and base-rate problems (bottom panel). Error bars are standard errors.

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.

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 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

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 5 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 5 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 5 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 5 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 5 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 5 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 5 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.

Perspectives on Psychological Science
7(1) 28–38

DOI: 10.1177/1745691611429354
http://pps.sagepub.com

Bias and Conflict: A Case for Logical Intuitions

Wim De Neys
Centre National de la Recherche Scientifique (CNRS), University of Toulouse, France

Abstract

Human reasoning has been characterized as often biased, heuristic, and illogical. In this article, I consider recent findings establishing that, despite the widespread bias and logical errors, people at least implicitly detect that their heuristic response conflicts with traditional normative considerations. I propose that this conflict sensitivity calls for the postulation of logical and probabilistic knowledge that is intuitive and that is activated automatically when people engage in a reasoning task. I sketch the basic characteristics of these intuitions and point to implications for ongoing debates in the field.

Keywords

bias, cognition, conflict detection, heuristics

Half a century of reasoning and decision-making research has shown that human judgment is often biased (e.g., Kahneman & Frederick, 2005). People seem to overrely on stereotypical intuitions and so-called heuristic thinking instead of more demanding, deliberative reasoning when making decisions (e.g., Evans, 2003, 2008). The received view is that although intuitive heuristics can sometimes be useful, they often cue responses that conflict with traditional logical or probabilistic normative principles and bias our decisions (e.g., Evans, 2010).

This bias has been demonstrated with a number of classic tasks that can be considered the "fruit flies" of the reasoning and decision-making field. Box 1 presents some examples of the most famous of these classic tasks. Literally hundreds of studies have used these tasks, and they have been the basis for most of the theorizing in the field (Bonnefon, 2011). Giving the correct response in the tasks requires only the application of some very basic logical or probabilistic principles. However, the tasks are constructed such that they intuitively cue a tempting stereotypical or belief-based heuristic response that conflicts with these principles. The striking finding has been that although the studies have been run with educated, university students, the vast majority of participants nevertheless fail to solve the problems correctly and pick the heuristic response. These findings have contributed to the widespread belief that traditional logical or probabilistic considerations play little role in our reasoning (e.g., Gigerenzer, 1996; Hertwig & Gigerenzer, 1999).

Consider again the classic tasks in Box 1 for a minute. Presumably, as most people, you were probably biased and picked the heuristic response the first time you encountered them. However, you might have picked the incorrect response, but were you actually fully convinced that your answer was right? That is, the problems might have tempted you to pick the heuristic response, but were you convinced that your answer was correct or did you feel that there was something tricky about the problem, that you were missing out on something? Recent studies on conflict sensitivity during biased reasoning suggest you probably did sense that something wasn't right and questioned your response (e.g., Bonner & Newell, 2010; De Neys, Cromheeke, & Osman, 2011; De Neys, Moyens, & Vansteenwegen, 2010). Using a range of methods these studies showed that people are especially sensitive to violations of the traditional logical and probabilistic principles in the classic tasks. For example, giving an unwarranted heuristic response in these tasks has been shown to affect a reasoner's autonomic arousal (e.g., De Neys et al., 2010), response times (e.g., Bonner & Newell, 2010), and subjective response confidence (e.g., De Neys et al., 2011). In this article, I point to the fundamental implications of this conflict sensitivity. My basic idea is that despite their erroneous responses, people have implicit knowledge of the logical and probabilistic normative principles that are evoked in the classic problems and automatically activate this knowledge when faced with the reasoning problem. Bluntly put, contrary to conventional wisdom, I argue

Corresponding Author:
Wim De Neys, Centre National de la Recherche Scientifique (CNRS), University of Toulouse, France, Laboratoire CLLE (CNRS, Université de Toulouse), Maison de la Recherche, 5, allée A. Machado, 31058 Toulouse Cedex 9, France
E-mail: Wim.Deneys@univ-tlse2.fr

Box 1. Illustrations of Some of the Most Popular "Fruit Flies" Tasks in the Reasoning and Decision-Making Field

A. Classic "Conflict" versions	B. Control "No conflict" versions
Conjunction fallacy task: Bill is 34. He is intelligent, punctual but unimaginative and somewhat lifeless. In school, he was strong in mathematics but weak in social studies and humanities. Which one of the following statements is most likely? a. Bill plays in a rock band for a hobby* b. Bill is an accountant and plays in a rock band for a hobby+	Conjunction fallacy task: Bill is 34. He is intelligent, punctual but unimaginative and somewhat lifeless. In school, he was strong in mathematics but weak in social studies and humanities. Which one of the following statements is most likely? a. Bill is an accountant*+ b. Bill is an accountant and plays in a rock band for a hobby
Base-rate neglect task: A psychologist wrote thumbnail descriptions of a sample of 1000 participants consisting of 995 females and 5 males. The description below was chosen at random from the 1,000 available descriptions. 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. Which one of the following two statements is most likely? a. Jo is a man* b. Jo is a woman+	Base-rate neglect task: A psychologist wrote thumbnail descriptions of a sample of 1,000 participants consisting of 995 males and 5 females. The description below was chosen at random from the 1,000 available descriptions. 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. Which one of the following two statements is most likely? a. Jo is a man*+ b. Jo is a woman
Syllogistic reasoning task: Premises: All vehicles have wheels Boats are vehicles Conclusion: Boats have wheels a. The conclusion follows logically* b. The conclusion does not follow logically+	Syllogistic reasoning task: Premises: All vehicles have wheels Bikes are vehicles Conclusion: Bikes have wheels a. The conclusion follows logically*+ b. The conclusion does not follow logically

Note. The left panel (A) shows the classic versions and the right panel (B) shows the newly constructed control versions. The classic versions cue a heuristic response that conflicts with the correct logical response (i.e., the response considered correct according to standard logic or probability theory principles). In the control versions, small content transformations guarantee that the cued heuristic response is consistent with the logical response. * = logical response, + = heuristic response

that people are actually intuitive logicians whose intuitive gut feelings are cueing the correct logical response.

I have organized this article around three sections. I start with a brief overview of the conflict sensitivity studies that inspired my claim. In the second section, I discuss the nature and characteristics of the logical intuitions that I propose. Lastly, I point to some intriguing implications of this proposal for dual process theories and the debate on human rationality.

For clarity, the reader should bear some general points in mind with respect to the nomenclature and labels that I use in this article. When I refer to the "correct," "logical," or "normative" response, I simply refer to the response that has traditionally been considered as correct or normative according to standard logic or probability theory. As I describe in the last section, the appropriateness of these traditional norms has been questioned by a number of authors. Under this interpretation, the heuristic response should not be labeled as "incorrect" or "biased." I will discuss implications of the present proposal for this debate, but for the sake of simplicity I stick to the traditional labeling. In the same vein, I use the term "logical" as a general header to refer both to standard logic and probability theory. Hence, the term "logical intuition" refers to an intuitive grasping of the standard logical and probability theory principles (e.g., conjunction rule, proportionality principle, logical validity) that are evoked in the classic reasoning problems.

Looking for Conflict

My claims are based on recent work on conflict detection during thinking (e.g., Bonner & Newell, 2010; De Neys et al., 2010, 2011; Stupple & Ball, 2008). The question that this line of research tries to answer is whether people detect that they are biased. More specifically, the studies use a wide range of processing measures to examine whether people are sensitive to violations of the traditional logical and probabilistic normative principles. That is, when people give the heuristic answer to the classic problems, do they really totally disregard these principles or do they show some basic sensitivity to the fact that their answer is inconsistent with them? To address this question, researchers have used conflict studies to contrast people's processing of the classic problems with newly constructed control versions. Recall that the classic versions typically cue a strong heuristic response that conflicts with the traditional normative principles. In the control or no-conflict versions, this conflict is removed and the heuristic response is consistent with the normative principles. Box 1 also presents examples of these control versions. In sum, heuristic thinking will cue the correct response on the control no-conflict problems and the incorrect response on the classic conflict versions. Accuracy rates on the control versions are typically very high, whereas they are dramatically low on the conflict versions. However, researchers have used conflict detection studies to look under the accuracy surface and focus on more subtle measures that made it possible to test whether people processed the two types of problems any differently.

Response latencies

For example, one basic procedure has been to simply look at people's response latencies: A number of studies reported that people need typically more time to solve the conflict than the control versions (e.g., Bonner & Newell, 2010; De Neys & Glumicic, 2008; Stupple & Ball, 2008; Thompson, Striemer, Reikoff, Gunter, & Campbell, 2003; Villejoubert, 2009). Now, clearly, the only difference between the two versions is whether the cued heuristic response is consistent with the traditional normative principles or not. If people were mere heuristic thinkers that did not take these normative considerations into account, they should not process the two types of problems any differently. Hence, the latency findings support the idea that people are sensitive to the traditional normative status of their judgment.

Gaze and eye-tracking studies

Further support for this claim has come from gaze and eye-tracking studies that showed that the longer latencies are specifically accompanied by a longer inspection of normatively critical problem information. For example, it has been observed that after participants read the conclusion of a conflict syllogism in which the conclusion believability conflicts with its logical validity (e.g., a valid but unbelievable conclusion) they make saccades to the major and minor premises and start reinspecting this information (Ball, Philips, Wade, & Quayle, 2006). Such reviewing was found to be much less pronounced on the no-conflict problems.

A similar gaze trend has been observed with base-rate problems: When solving conflict versions, participants show an increased tendency to re-view the paragraph with the base-rate information after they have read the personality description (De Neys & Glumicic, 2008). A surprise recall test that followed showed that the increased base-rate inspection was accompanied by a better recall of the base-rate information for the conflict problems. Interestingly, a subsequent study showed that, in contrast to the normative information, information that was associated with the heuristic response was less accessible in memory after solving conflict problems (De Neys & Franssens, 2009). Participants in this study were given a lexical decision task in which they had to decide whether a string of letters formed a word or not after each reasoning problem. Results showed that lexical decisions about words that were linked to the cued heuristic response took longer after solving conflict problems, suggesting that participants had attempted to block this information during reasoning.

Neuropsychology

The behavioral conflict findings have also been validated with a brain-based approach. For example, in one study (De Neys, Vartanian, & Goel, 2008) fMRI was used to monitor the activation of a specific brain area, the anterior cingulate cortex (ACC), which is believed to mediate conflict detection during thinking (e.g., Botvinick, Cohen, & Carter, 2004). Participants were given classic conflict base-rate problems and the no-conflict control versions. In line with the behavioral findings, results showed that the ACC was much more activated when people solved the conflict versions than when they solved the control versions. In a subsequent study, participants' skin conductance was recorded to monitor autonomic nervous system activation while solving conflict and no-conflict syllogisms (De Neys et al., 2010). Results showed that solving the conflict problems resulted in a clear electrodermal activation spike. Hence, in addition to the ACC activation, solving conflict problems literally aroused participants. These neural conflict signals have also been shown to affect people's subjective response confidence: Participants typically indicate that they feel less confident about their answer after solving conflict problems than after solving the control problems (e.g., De Neys et al., 2011).

A Case for Logical Intuitions

The conflict detection studies established that despite the well-documented failure to give the correct answer on the classic problems, people do not simply disregard the traditional normative implication of their judgments; rather, they are sensitive to the fact that their heuristic answer conflicts with it. However, although the studies clarified that people might show some basic normative sensitivity, it is less clear how this sensitivity needs to be conceived. What is the exact nature of the normative knowledge that is needed to detect conflicts and where does it come from? In this section, I clarify my basic point that this knowledge is intuitive in nature. I validate my claim by demonstrating that the established normative sensitivity has two key characteristics of intuitive processes: That is, the necessary knowledge is activated automatically and it is implicit in nature. In an attempt to demystify the idea of intuitive logical thinking[1] I also point to the developmental origin of the postulated intuitions.

Automatic activation

In theory, one could argue that the documented normative sensitivity in the conflict detection studies results from effortful probabilistic or logical thinking. That is, people would detect that the cued heuristic response conflicts with the traditional normative response because they actively compute this normative or logical response by engaging in demanding logical or probabilistic analysis (e.g., some sort of hypothetical thinking, mental model construction or Bayesian computations). A number of influential authors have indeed argued that people would always simultaneously engage in intuitive-heuristic and demanding-logical thinking and consequently be sensitive to conflicts (e.g., Epstein, 1994; Sloman, 1996). However, in contrast with this view, I propose that the crucial normative considerations are activated automatically. Indeed, the idea is that people master the normative principles and that this knowledge is brought in a heightened activation state when faced with the reasoning problem. In other words, I suggest that in addition to the well-established heuristic response, the classic tasks also automatically evoke an intuitive logical response. The key point is that this activation is effortless and does not require any demanding or elaborate analytic thinking.

Cognitive load. Although the idea of an effortless logical sensitivity may sound somewhat counterintuitive, it is important to stress that there is direct empirical support for this assumption. For example, in one study, participants solved conflict and control base-rate problems while their cognitive resources were burdened with a secondary task (i.e., memorization of a dot pattern, see Franssens & De Neys, 2009). Solving conflict problems correctly is generally considered cognitively demanding because it requires, for example, the inhibition of the salient heuristic response, a process know to heavily tax our limited executive resources (e.g., Dempster & Corkill, 1999; De Neys & Van Gelder, 2008; Handley, Capon, Beveridge, Dennis, & Evans, 2004; Houdé, 1997, 2007; Morris, 2000; Moutier, Plagne-Cayeux, Melot, & Houdé, 2006; Perret, Paour, & Blaye, 2003; Reyna, Lloyd, & Brainerd, 2003; Simoneau & Markovits, 2003; Stanovich & West, 2000). As the heuristic response does not conflict with the normative considerations on the control problems, there is no need to engage in inhibitory processing and solving these problems is expected to be effortless (e.g., De Neys, 2006; Evans, 2009; Stanovich & West, 2000).

In line with these predictions, Franssens and De Neys indeed found that cognitive load did not affect accuracy on the control problems but decreased performance on the conflict problems. The crucial manipulation was that after the experiment was finished, participants took an unannounced, surprise memory test in which they were asked to recall the base rates of the problems that they just solved. As noted above, this recall index had been previously introduced as a measure of conflict detection efficiency: The extended reviewing that is associated with successful conflict detection was shown to boost recall of the base rates (see De Neys & Glumicic, 2008). In line with these findings, Franssens and De Neys indeed observed that in the no-load condition, base rates of the conflict problems were better recalled than the base rates of the no-conflict control problems. However, the critical finding was that although the reasoning accuracies on the conflict problems decreased under load, the load had no impact on the base-rate recall on these problems. Hence, the recall-conflict sensitivity index was not affected by cognitive load. This suggests that whatever the nature of the necessary knowledge that allows people to identify conflict problems as such might be, its activation is not cognitively demanding.

Cognitive capacity. Additional evidence for the automaticity of the normative sensitivity comes from the observation that the conflict detection findings did not depend on participants' cognitive capacities or response accuracy. Note that although most people are biased when solving the classic conflict problems, some participants do manage to solve the problems correctly. It has been shown that these participants are specifically those highest in executive resources (e.g., De Neys & Verschueren, 2006; Newstead, Handley, Harley, Wright, & Farrelly, 2004; Stanovich & West, 2000). One might argue that these cognitively gifted participants are driving the observed conflict sensitivity findings since they might have the potential to engage in demanding analytic computations. However, the detection studies clearly established that even the least gifted reasoners (i.e., the most biased reasoners with the lowest accuracy scores) showed the sensitivity effects (e.g., De Neys et al., 2010, 2011; De Neys & Glumicic, 2008). Hence, although solving conflict problems correctly might require abundant executive resources, detecting the conflict is successful even for the most biased reasoners. This lack of individual differences in conflict detection efficiency further suggests that the necessary normative knowledge activation is indeed effortless.

Repeated testing confound? Finally, a critic of the automatic activation idea might argue that the automaticity results from a repeated testing or training confound in the conflict detection studies. Note that these studies typically presented participants with multiple conflict and no-conflict problems. For example, in the fMRI study of De Neys et al. (2008) participants solved about 100 base-rate neglect problems. One might argue that this repeated presentation primed the activation of the necessary normative principles through some kind of learning process. That is, at the start of the experiment, conflict detection would only occur after successful completion of a demanding logical reasoning process. After repeated problem presentation, however, this process might become automated. Nevertheless, such a confound can be discarded since item analyses showed that the conflict sensitivity effects are present from the first problem presentation (e.g., De Neys et al., 2010, 2011; De Neys & Franssens, 2009; De Neys & Glumicic, 2008).

Taken together, these findings indicate that consistent with the idea of a logical intuition, the conceptual knowledge that is needed to detect heuristic and logical conflict is activated automatically and does not draw on demanding computations.

Implicit knowledge

A second issue that points to the intuitive nature of people's normative sensitivity is its implicitness. For example, when participants were asked to think aloud while they were solving base-rate problems, they hardly ever explicitly referred to the base-rate information when solving the classic conflict versions (see De Neys & Glumicic, 2008). Hence, although participants needed more time to solve these problems, made eye-movements to the base-rate information, showed increased ACC activation, increased autonomic arousal, and decreased response confidence when solving these very same problems, they did not verbally express that the base rates mattered. In general, this fits with the long established observation that people's online verbalizations during thinking and their retrospective response justifications do not typically indicate that they are taking any normative logical or probabilistic considerations into account (e.g., Evans & Over, 1996; Wason & Evans, 1975). Indeed, it is the lack of such explicit reference to traditional normative principles that initially contributed to the popular belief that people do not take these principles into account (De Neys & Glumicic, 2008): If people do not give the correct logical response and do not refer to any traditional logical or probabilistic principles or information, it is not surprising that researchers became convinced that these principles play little role in reasoning. Note that it is only by introducing new and more subtle processing measures that researchers using conflict detection studies managed to start cutting the ground under this view. However, the point is that the activated knowledge that allows people to detect the conflict is implicit knowledge. People will not manage to label the detected normative violations explicitly. Hence, the postulated logical intuition can be conceived as a "gut feeling" (e.g., Franssens & De Neys, 2009; Thompson, 2009): People will be aware that there is something fishy about their heuristic response, but they will not be able to put their finger on it and explain why their response is questionable. More precisely, the idea that I propose is that the conflict between implicitly activated normative knowledge and the cued heuristic response creates arousal. People experience this arousal, this makes them doubt their heuristic response, but they will not be able to justify why their response is questionable. Such explicit justification will require engaging in a proper, demanding logical or probabilistic analysis. However, the implicit knowledge suffices to signal that the heuristic response is not fully warranted.

Developmental basis

The automatic activation and implicitness of the demonstrated normative sensitivity in the conflict detection studies support the idea that the process is intuitive in nature and does not result from a demanding and explicit logical or probabilistic reasoning process. These characteristics help to validate the claim that people have indeed normative logical or probabilistic intuitions. Nevertheless, a critic might argue that the postulation of such intuitive logicality has a quite esoteric or mythical flavor. That is, the basis or origin of the hypothesized normative knowledge might be questioned: If the demonstrated normative sensitivity does not result from demanding computations, then where does it come from? More generally, one might wonder about independent evidence (i.e., independent from the conflict findings) that indicates that people do master the crucial normative principles. Therefore, I identify developmental findings that suggest that the core of the normative principles that are evoked in the classic problems are actually acquired quite early in life.

For example, with respect to the role of base rates, several studies have now clearly shown that even very young infants seem to grasp the importance of proportionality in random drawing (e.g., Kushnir, Xu, & Wellman, 2010; Téglás, Girotto, Gonzalez, & Bonatti, 2007; Xu & Garcia, 2008). Following the pioneering work of Téglás et al. (2007), one study showed 8-month old infants a woman taking four red balls and one white ball out of a box with her eyes closed (see Xu & Garcia, 2008). When the content of the box were revealed, infants looked longer at an unexpected population (a box full of mostly white balls with some red balls) than at an expected population (mostly red balls and some white balls). In a variation of this paradigm, 20-month old infants were shown a puppet that removed five toys of one and the same type (i.e., the target toy) from a box containing two types of toys (i.e., target toys and alternate toys). Next, they were presented with the two types of toys and were asked to give the puppet the one he liked most. The critical finding was that the infants' choices were affected by the base rates of the target and alternate toys: The smaller the number of target toys in the container, the more likely that children selected it as the preferred toy of the puppet (Kushnir et al., 2010). Kushnir et al. reasoned that the infants inferred that the puppet had a preference for that type of toy when there was a mismatch between the sampled toys and the population of toys in the box. Hence, these findings clearly indicate that even infants are sensitive to the role of base rates in probability judgments.

Similar observations, although with somewhat older children, have been made with respect to mastery of the conjunction rule and logical validity principles. Knowing the conjunction rule boils down to grasping the class inclusion principle that subsets will never be more numerous than superordinate sets (e.g., Reyna, 1991). Hence, there will always be more bank tellers (i.e., the superordinate set) than bank tellers that are also active in the feminist movement (i.e., the subset). However, Piaget and Szeminska's (1941/1967) seminal work established that children learn this principle between the age of 7 and 11. In a typical class inclusion task, children will be shown a number of objects—for example, five cows and two dogs. Children are then asked whether there are more cows (i.e., the more numerous subset) or more animals (i.e., the superordinate set). Although children younger than five typically pick the subset, 10-year olds already show quasiperfect performance (Brainerd & Reyna, 2001; Perret et al., 2003). It has been shown that in the same preadolescent age range, children also start to show good competence at discriminating classic valid (e.g., Modus Ponens) and invalid (e.g., Affirmation of the Consequent) logical arguments (Morris, 2000).

Possible misconception. The fact that even young children master the key normative principles to solve the classic problems underscores the point that there does not need to be anything esoteric about the claim that educated adults master these too. Indeed, given the developmental findings, one might wonder why the reasoning field ever started questioning adults' knowledge of these principles in the first place. However, it is important to stress an important theoretical point and misconception. Although some authors (e.g., Wason, 1968, 1983) have indeed claimed that people's failure to solve the classic tasks pointed to a genuine lack of normative knowledge (i.e., so-called "mindgaps"; see Stanovich & West, 2008), others, such as the founding fathers of the heuristic and biases field, Kahneman and Tversky (e.g., 1973), have refrained from drawing this conclusion. Kahneman and Tversky's point was not that adults did not master the traditional normative principles, but rather that this knowledge was not used or activated when faced with salient heuristics. Indeed, in their classic studies, Kahneman and Tversky often included abstract versions of the classic problems. In contrast with the conflict (or no-conflict) versions, these abstract problems did not cue a heuristic response. For example, in an abstract base-rate problem, people would be shown the base rates without accompanying personality description. In line with the developmental findings, Kahneman and Tversky observed that adults did an almost perfect job in solving these abstract problems, indicating that people must have basic knowledge of the role of these principles. The same point is illustrated by studies that show how small changes in the problem cover story, aimed to evoke consideration of the normative principles, can dramatically decrease heuristic responding (e.g., Nisbett, Krantz, Jepson, & Kunda, 1983). Of course, the fact that people might know these principles does not imply that they also use them. Hence, Kahneman and Tversky could still claim that people faced with salient heuristics in the standard tasks will not consider the normative insights and fail to detect the biased nature of their judgment. It is this critical issue that was tackled by the conflict detection studies. If these normative principles were not activated when people were biased by salient heuristics—if they were not taken into account—then reasoners should not process the conflict and no-conflict versions any differently.

In sum, the established normative sensitivity in the conflict detection studies invalidated the idea that people do not detect their bias. My point in this article is that the necessary normative knowledge enabling conflict detection is intuitive in nature (i.e., activated automatically and implicit). I have pointed to the developmental findings and findings with abstract problem versions to clarify that there does not need to be anything mystical about the origin of these intuitions. In and by itself, there is ample evidence that even children master the basic principles. What is critical about the present claim is that these principles are taken into account even when people are biased and that this results from intuitive processing.

Potential Implications

In this final section, I explore potential implications of the logical intuition proposal for ongoing debates in the reasoning and decision-making field. I focus on two critical issues concerning dual process theories and the role of traditional norms for thinking. I also discuss the boundary conditions of the implications.

Logical intuitions and dual process theories

The influential dual process theories have characterized human thinking as an interplay of an intuitive-heuristic and deliberate-analytic system (e.g., Epstein, 1994; Evans, 2003; Evans & Over, 1996; Sloman, 1996; Stanovich & West, 2000). The intuitive system is typically conceived as the system that cues the heuristic response on the classic problems by relying on prior knowledge and beliefs. The deliberate system is conceived as the system that enables the type of effortful hypothetical thinking that allows people to reason logically and probabilistically. Hence, it is assumed that the heuristic response to the classic problems is cued by the intuitive system, whereas the logical response (i.e., the response that is considered correct according to standard logic or probability theory) is computed by the deliberate system.

Note that this does not entail that deliberate processing always results in a correct, logical answer and intuitive processing in a biased answer. Dual process theorist have clarified that in some cases, people might be biased precisely because their cognitive resources are overburdened by too much deliberation (e.g., Evans, in press; Stanovich, 2010). Likewise, a person who is guessing might end up giving a logically correct response without engaging in any deliberate processing. However, the point is that, in the prototypical case, the dual process framework assumes that the logical response on the classic reasoning problems will be computed by the deliberate system. The concept of a logical intuition forces one to revise this idea. In dual process terms, the present claims imply that the intuitive system also cues a logical response. This proposal is puzzling from a standard dual process perspective (Evans, 2010; Handley, Newstead, & Trippas, 2011), but I believe it actually may help to understand how the intuitive and deliberate system can interact.

It has been noted previously that the nature of the relation between the two systems is not clear (e.g., Evans, 2007, 2009). In a nutshell, a serial and a parallel activation view can be distinguished (see Fig. 1). According to the parallel view (e.g., Epstein, 1994; Sloman, 1996), both systems are supposed to be simultaneously computing a problem solution from the start. According to the serial view (e.g., Kahneman & Frederick, 2005; Stanovich & West, 2000), a reasoner initially relies on the intuitive system and the deliberate system will only be recruited in case the intuitively cued response conflicts with the output of the deliberate system. However, a fundamental conceptual problem for the serial view is how the reasoner can ever detect a conflict between the output of the intuitive and

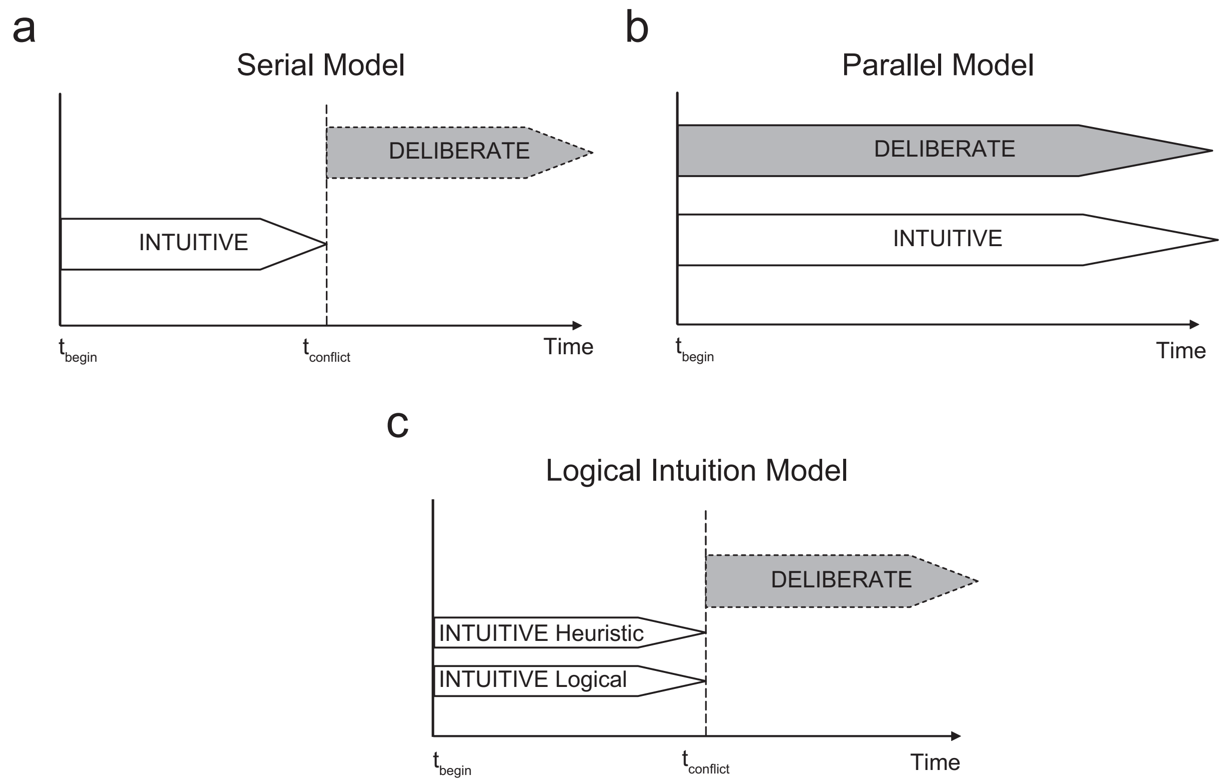

Fig. 1. Three different theoretical models of the relation between the intuitive and deliberate system. Deliberate processing is represented by gray bars and intuitive processing by white bars. The horizontal axis represent the time flow. In the serial model (A), the deliberate system is only activated after a conflict ($t_{conflict}$) with the intuitive system. In the parallel model (B), the intuitive and deliberate system are both activated from the start. In the logical intuition model (C), deliberate processing is triggered by conflict ($t_{conflict}$) between intuitive heuristic and intuitive logical processing. The dashed lines represent the optional nature of the triggered deliberate processing in the serial and logical intuition model.

deliberate system if the deliberate system is not yet engaged. The assumed simultaneous activation of the two systems in the parallel view sidesteps this problem. However, the parallel view faces its own problems. In the parallel model, the deliberate route is blindly engaged from the start. People always start the time-consuming and demanding deliberate computations. Thus, the parallel model basically throws away the benefits of the intuitive route. Clearly, intuitive and deliberate thinking do not always conflict. When there is no conflict, it is perfectly fine to rely on the intuitive route. Engaging in demanding deliberate operations is redundant in this case and would be a waste of scarce cognitive resources (De Neys & Glumicic, 2008). Hence, what dual process models need is a way to detect whether deliberate thinking is required without having to engage in deliberate thinking (e.g., Evans, 2009; Thompson, Prowse Turner, & Pennycook, 2011).

The cueing of an intuitive logical response can help to solve this conceptual puzzle. If the intuitive system cues both a logical and heuristic response, potential conflict can be detected without prior engagement of the deliberate system. Hence, the idea is that rather than parallel activation of the two systems, there would be parallel activation of two different types of intuitive responses: A heuristic intuitive response based on mere semantic and stereotypical associations, and a logical intuitive response based on the activation of traditional logical and probabilistic normative principles. If the two intuitive responses are consistent, people will select the cued response, and the reasoning process ends without further deliberate reflection. Any conflict between the two responses would signal the need to engage the deliberate system. Clearly, the fact that deliberate operations are called upon does not imply that they will be successfully recruited or completed. However, it does present the human reasoning engine with a clear switch rule to determine whether deliberate reflection is required without a need to postulate an inefficient, permanent activation of the deliberate system.

Further dual process considerations. As one reviewer suggested, it might be interesting to note that the idea of a logical intuition is not entirely in opposition to standard dual process theories. Dual process theories do allow for the possibility that a deliberate process becomes automated and intuitive in nature through repeated practice (Evans, 2003; Sloman, 1996; Stanovich & West, 2000). This point has been typically used to explain expert performance. For example, few scholars would contest that a professional logician might be able to solve logical reasoning problems in an entirely intuitive manner after years of extensive training. One could argue that the logical intuition proposal shares some common conceptual ground with the basic automatization idea. Given the developmental origin of the logical intuitions that I have sketched, one might want to conceive the critical mastering of the logical and probabilistic principles throughout a child's development as a kind of automatization process. Although I would not necessarily object to such an analogy, it should be clear that a key aspect of the logical intuitions proposal is precisely that these are maintained by all reasoners and not just by a small subgroup of highly trained experts. Note that another aspect in which the analogy falls short is that we would still expect an expert to be able to justify her response even after automatization.

A final issue with respect to the dual process implications of the logical intuition proposal concerns the status of the two intuitive responses. That is, if I am right and the intuitive system cues both a heuristic and logical response, one might wonder why the heuristic response nevertheless typically dominates in case of conflict. One straightforward explanation is that the activation levels of the two types of intuitive responses differ. That is, the heuristic response might be more strongly activated, salient, or appealing than the logical response. Hence, there is no need to assume that the two intuitive responses have the exact same strength or status. I do claim that conflict between a heuristic and logical intuition will result in doubt and a questioning of the heuristic response, but this does not imply that reasoners consider the logical response to be fully warranted. All that is needed is that conflict lowers the default activation or confidence level of the heuristic response. In absolute terms, the intuitive heuristic response might still be stronger than the intuitive logical response. Note that such differential activation level would also explain why a final selection of the logical response will still require a demanding inhibition of the heuristic response (e.g., Evans, 2003; Handley et al., 2004; Houdé, 1997, 2007; Stanovich & West, 2000).

Logical intuitions and normative debate

Over the decades, the apparent omnipresent failure of educated adults to select the response that is consistent with the traditional logical and probabilistic norms on the classic problems has led some researchers to question the validity of these norms (e.g., Gigerenzer, 1996; Hertwig & Gigerenzer, 1999; Mercier & Sperber, 2011; Oaksford & Chater, 2007). These scholars argued that humans are adhering to other norms than the traditional logical or probabilistic standards when solving classic reasoning tasks (Bonnefon, 2009). People would interpret tasks such as the base-rate or conjunction fallacy task as a type of social classification problem in which they try to determine to which social group a character belongs. Given this alternative task interpretation, people would consider the heuristic response perfectly valid, and additional standard logical or probabilistic normative considerations would play no role in their reasoning. These claims resulted in the view that, except for some highly trained logicians, standard logic or probability theory principles would be irrelevant for human reasoning.

My proposal argues against this view. Although people rarely give the traditional normative answer or explicitly refer to the traditional principles, the reviewed evidence suggests that they do activate these normative principles implicitly. The fact that a logical response is intuitively cued and affects a reasoners' task processing makes it very hard to argue that the traditional norms play no role in reasoning. At the very least, one needs to acknowledge that the intuitive activation questions the claim that reasoners interpret the classic tasks as mere social classification tasks. If this were the case, and normative considerations such as the conjunction rule, sample sizes, or logical validity were considered irrelevant, then it becomes hard to explain why the presence of a conflict between cued social intuitions and the very same normative principles decreases people's response confidence or makes them review the normative problem information, for example.

Clearly, the normative debate in the cognitive sciences is a complex and multilayered debate. To avoid confusion, it is probably worthwhile to stress explicitly that my claim with respect to the role of the traditional norms is situated at the psychological processing level. Obviously, the fact that people show sensitivity to violations of a certain norm does not entail that the norm is valid. From an epistemological point of view, it might still be that other norms are more appropriate. In other words, my claim is not that the traditional norms are ultimately correct, but rather that human reasoners at least seem to consider them to be correct (i.e., relevant for their inference making). Note that this does not imply that people need to be fully confident about their logical intuitions or consider them to be fully appropriate either. As I argued with respect to the possible differential status of the intuitive heuristic and intuitive logical response, people might still find the heuristic response more appealing than their logical intuition when solving the conflict problems. The point is that the logical intuition proposal implies that people are giving some weight to the traditional logical and probabilistic principles during their decision-making process. This argues against the view that reasoners consider these principles irrelevant and should give pause for thought before rejecting the role of traditional logic and probability theory principles in human reasoning.

Boundary condition

In closing, when considering the present proposal and its implications, it is important to keep an obvious but critical boundary condition in mind. As I clarified in the introduction, I use the logical intuition label to refer to the idea that people intuitively take into account the traditional logical and probabilistic normative principles that are evoked in the classic reasoning problems. Hence, my claims specifically apply to the classic tasks that have been the basis for most of the theorizing in the reasoning and decision-making field. To be clear, I do not argue that people have logical intuitions about each and every problem they may encounter in life. One of the main reasons for postulating that people intuitively consider the logical and probabilistic principles in the classic problems is precisely the fact that these principles are so elementary and acquired early in life. Note that it was the same elementary nature of these principles that gave the original bias studies such a wide impact. Indeed, few people would have been surprised if Tversky and Kahneman had shown that reasoners were biased when solving nuclear physics equations, for example. Clearly, one important part of the impact of the studies came from the suggestion that people are not even taking the most basic logical and probabilistic principles into account. It is this point that the conflict detection studies and logical intuition proposal argue against. However, the elementary nature of the principles involved presents an intrinsic boundary condition for the logical intuition claim. Logical intuitions are bound to arise in situations where the logical solution or principle is "simple" and easily (i.e., automatically) activated. Indeed, in a sense, one might state that what I have tried to clarify in this article is precisely that the traditional standard logic and probability theory principles in the classic reasoning problems fit this criterion.

Summary and Conclusion

Recent studies on conflict detection during biased reasoning indicate that people are especially sensitive to violations of traditional normative principles in the classic "fruit flies" tasks. I argued that these findings call for the postulation of logical intuitions. That is, I claim that despite the erroneous answer, people have implicit knowledge of the logical and probabilistic normative principles that are evoked in the classic problems and automatically activate this knowledge when faced with the reasoning problem. I presented evidence for the automatic activation and implicit nature of the postulated intuitions, pointed to their developmental origin, and sketched potential implications for dual process theories and the debate on the validity of the traditional norms.

As I stated in the introduction, the goal of this article was to sketch a new conceptual idea. Clearly, the present claims do not amount to a fully developed theoretical framework yet. Hence, I fully acknowledge that my suggestions will need to be tested further. However, I hope to have clarified that the proposal is supported by recent data, generates testable predictions, and may help to shine a fresh light on long lasting controversies in the field. I believe that this should convince the reasoning and decision-making community that the idea that people have logical intuitions is valuable and should become a primary area of future empirical and theoretical scrutinizing.

Acknowledgments

I would like to thank Jean-Francois Bonnefon and Barbara Spellman for valuable comments and suggestions.

Declaration of Conflicting Interests

The author declared that they had no conflicts of interest with respect to his authorship or the publication of this article.

Note

1. As I stated earlier, I use the term "logical" in this article as a general header to refer to both standard logic and probability theory.

References

Ball, L. J., Philips, P., Wade, C. N., & Quayle, J. D. (2006). Effects of belief and logic on syllogistic reasoning: Eye-movement evidence for selective processing models. *Experimental Psychology, 53*, 77–86.

Bonnefon, J. F. (2009). A theory of utility conditionals: Paralogical reasoning from decision-theoretic leakage. *Psychological Review, 116*, 888–907.

Bonnefon, J. F. (2011). *Le raisonneur et ses modèles*. [The reasoner and his models]. Grenoble, France: Presses Universitaires de Grenoble.

Bonner, C., & Newell, B. R. (2010). In conflict with ourselves? An investigation of heuristic and analytic processes in decision making. *Memory & Cognition, 38*, 186–196.

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

Brainerd, C. J., & Reyna, V. F. (2001). Fuzzy-trace theory: Dual Processes in memory, reasoning, and cognitive neuroscience. *Advances in Child Development and Behavior, 28*, 49–100.

Dempster, F. N., & Corkill, A. J. (1999). Interference and inhibition in cognition and behaviour: Unifying themes for educational psychology. *Educational Psychology Review, 11*, 1–88.

De Neys, W. (2006). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science, 17*, 428–433.

De Neys, W., Cromheeke, S., & Osman, M. (2011). Biased but in doubt: Conflict and decision confidence. *PLoS ONE, 6*, e15954.

De Neys, W., & Franssens, S. (2009). Belief inhibition during thinking: Not always winning but at least taking part. *Cognition, 113*, 45–61.

De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition, 106*, 1248–1299.

De Neys, W., Moyens, E., & Vansteenwegen, D. (2010). Feeling we're biased: Autonomic arousal and reasoning conflict. *Cognitive, Affective, & Behavioral Neuroscience, 10*, 208–216.

De Neys, W., & Van Gelder, E. (2008). Logic and belief across the life span: The rise and fall of belief inhibition during syllogistic reasoning. *Developmental Science, 12*, 123–130.

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.

Epstein, S. (1994). Integration of the cognitive and psychodynamic unconscious. *American Psychologists, 49*, 709–724.

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 & Reasoning, 13*, 321–339.

Evans, J. St. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59*, 255–278.

Evans, J. St. B. T. (2009). How many dual process theories do we need: One, two or many? In J. St. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 33–54). Oxford, UK: Oxford University Press.

Evans, J. St. B. T. (2010). Intuition and reasoning: A dual process perspective. *Psychological Inquiry, 21*, 313–326.

Evans, J. St. B. T. (in press). Dual-process theories of reasoning: Facts and fallacies. In K. Holyoak & R. G. Morrison (Eds.), *The Oxford handbook of thinking and reasoning*. New York, NY: Oxford University Press.

Evans, J. St. B. T., & Over, D. E. (1996). *Rationality and reasoning*. Hove, UK: Psychology Press.

Franssens, S., & De Neys, W. (2009). The effortless nature of conflict detection during thinking. *Thinking & Reasoning, 15*, 105–128.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). *Psychological Review, 103*, 592–596.

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

Handley, S. J., Newstead, S. E., & Trippas, D. (2011). Logic, beliefs, and instruction: A test of the default interventionist account of belief bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 28–34.

Hertwig, R., & Gigerenzer, G. (1999). The "conjunction fallacy" revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making, 12*, 275–305.

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 & Reasoning, 13*, 81–89.

Kahneman, D., & Frederick, S. (2005). A model of heuristic judgement. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 267–293). Cambridge, MA: Cambridge University Press.

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

Kushnir, T., Xu, F., & Wellman, H. M. (2010). Young children use statistical sampling to infer the preferences of other people. *Psychological Science, 21*, 1134–1140.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences, 34*, 57–111.

Morris, A. K. (2000). Development of logical reasoning: Children's ability to verbally explain the nature of the distinction between logical and nonlogical forms of argument. *Developmental Psychology, 36*, 741–758.

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

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.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review, 90*, 339–363.

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

Perret, P., Paour, J.-L., & Blaye, A. (2003). Respective contributions of inhibition and knowledge levels in class inclusion development: A negative priming study. *Developmental Science, 6*, 283–288.

Piaget, J., & Szeminska, A. (1967). *La genèse du nombre chez l'enfant* [Children's development of number concepts]. Neuchâtel, Switzerland: Delachaux et Niestlé. (Original work published 1941)

Reyna, V. F. (1991). Class inclusion, the conjunction fallacy, and other cognitive illusions. *Developmental Review, 11*, 317–336.

Reyna, V. F., Lloyd, F. J., & Brainerd, C. J. (2003). Memory, development, and rationality: An integrative theory of judgment and decision-making. In S. Schneider & J. Shanteau (Eds.), *Emerging perspectives on judgment and decision research* (pp. 201–245). New York, NY: Cambridge University Press.

Simoneau, M., & Markovits, H. (2003). Reasoning with premises that are not empirically true: Evidence for the role of inhibition and retrieval. *Developmental Psychology, 39*, 964–975.

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

Stanovich, K. E. (2010). *Rationality and the reflective mind*. New York, NY: Oxford University Press.

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.

Stupple, E. J. N., & Ball, L. J. (2008). Belief-logic conflict resolution in syllogistic reasoning: Inspection-time evidence for a parallel-process model. *Thinking & Reasoning, 14*, 168–181.

Téglás, E., Girotto, V., Gonzalez, M., & Bonatti, L. L. (2007). Intuitions of probabilities shape expectations about the future at 12 months and beyond. *Proceedings of the National Academy of Sciences of the United States of America, 104*, 19156–19159.

Thompson, V. A. (2009). Dual process theories: A metacognitive perspective. In J. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 171–196). Oxford, UK: Oxford University Press.

Thompson, V. A., Striemer, C. L., Reikoff, R., Gunter, R. W., & Campbell, J. I. D. (2003). Syllogistic reasoning time: Disconfirmation disconfirmed. *Psychonomic Bulletin & Review, 10*, 184–189.

Thompson, V. A., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology, 63*, 107–140.

Villejoubert, G. (2009). Are representativeness judgments automatic and rapid? The effect of time pressure on the conjunction fallacy. *Proceedings of the Annual Meeting of the Cognitive Science society, 30*, 2980–2985.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology, 20*, 273–281.

Wason, P. C. (1983). Realism and rationality in the selection task. In J. St. B. T. Evans (Ed.), *Thinking and reasoning* (pp. 45–75). London, England: Routledge.

Wason, P. C., & Evans, J. (1975). Dual processes in reasoning. *Cognition, 3*, 141–154.

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences of the United States of America, 105*, 5012–5015.