COGNIT 1695 ARTICLE IN PRESS No. of Pages 52
30 June 2007 Disk Used

Available online at www.sciencedirect.com
ScienceDirect
Cognition xxx (2007) xxx–xxx
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, Toronto, 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

* 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
doi:10.1016/j.cognition.2007.06.002

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), 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

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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, 2007; 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, claim 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

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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 (2007) 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).

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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 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

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

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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.

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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.

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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.

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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:

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

> 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'.

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

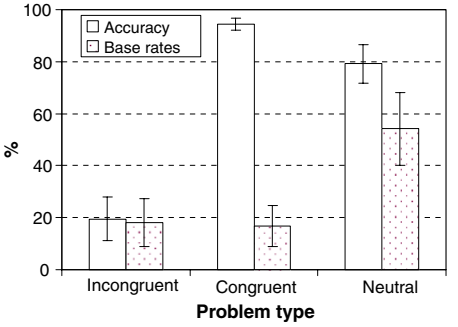

Fig. 1. Mean proportion correct responses and explicit base rate mentioning in verbal protocols. Error bars are standard errors.

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.

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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

* $p < .001$.

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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

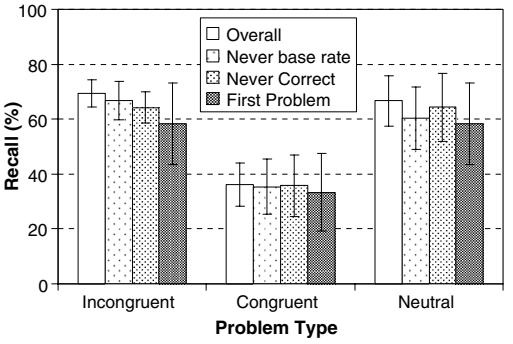

Fig. 2. Mean overall proportion of correct base rate recall. Recall performance for participants who never mentioned the base rates (Never base rate), always erred on the incongruent problems (Never correct), and the first presented problem of every type (First problem) are also presented. Error bars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

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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.

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

638 *3.1.2. Materials*

639 *3.1.2.1. Decision making task.* Participants were presented with Flemish versions of
640 the base rate problems. Problem content was slightly adapted for the Flemish test
641 context (e.g., we used the well known low-end European retailer *Aldi* instead of
642 *Wall-Mart*). As with the Canadian versions in Experiment 1 the stereotypical
643 descriptions were validated in a pilot rating study. As in Experiment 1, participants
644 solved a total of 18 problems (6 incongruent, 6 congruent, and 6 neutral ones) with
645 slightly varied base rates (e.g., 995/5, 996/4, 997/3).

646 The lack of explicit base rate mentioning for the Flemish versions was also vali-
647 dated in a short thinking aloud study with 14 Flemish undergraduates (these people
648 did not participate in Experiment 2). Participants solved four incongruent problems
649 aloud. Results replicated the thinking aloud findings with the Canadian students in
650 Experiment 1. Only 21% of the problems were solved correctly (i.e., $n = 12$ correct
651 responses out of a total of 56 trials) and base rates were mentioned in fewer than
652 20% of the cases (i.e., $n = 11$ out of 56). When the problem was solved erroneously
653 base rates were only mentioned 11% of the time (i.e., $n = 5$ out of 44). This estab-
654 lished that, as in Experiment 1, participants were typically not explicitly referring
655 to the base rates with the adapted material.

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

667

668 In a study 1000 people were tested. Among the participants there were 995 who
669 buy their clothes at high-end retailers and 5 who buy their clothes at Wal-Mart.
670 Karen is a randomly chosen participant of this study.

671

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

674

675 What is most likely?

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

678 Type down the letter reflecting your decision: __

679 Participants started the experiment by reading the following general instructions:

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

> 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.

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

718
719 You don't have to rush, just read all the information in a natural pace. Once
720 you've completely processed the information we ask you to press the 'a' key
721 immediately and then the next problem will be presented.

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

730 *3.1.2.3. Recall task.* The recall task used the same format as in Experiment 1. Since
731 Experiment 1 showed that recall was stable over the different items it was decided to
732 restrict the recall test to four problems of each item type (e.g., in Experiment 1 recall
733 of all 18 items was probed). We selected the 12 problems with the most diverse con-
734 tent. Despite the decent recall performance, a number of participants in Experiment
735 1 remarked that the task was quite lengthy and repetitive. It was hoped that the
736 shorter and more diverse version would result in a more engaging task and possibly
737 a more optimal measurement of the recall performance. The questions were printed
738 one to a page in a booklet. Recall questions were presented in one of eight randomly
739 determined orders. Except for the phrasing of the first sentence (e.g., 'One of the
740 problems you just solved/read...') booklets for the reading and decision making
741 group were completely similar. Of course, as in Experiment 1, recall was not
742 announced before the reading or decision making task was completed.

743 *3.1.3. Procedure*

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

749 *3.2. Results*

750 *3.2.1. Decision making task accuracy*

751 Participants' accuracy on the base rate problems very closely replicated the find-
752 ings of Experiment 1. On average, only 22% of the incongruent problems were solved
753 correctly but participants had far less trouble in selecting the correct response on the
754 congruent (97%) and neutral (80%) problems, $F(2, 86) = 184.19$, $MSE = 1.33$,
755 $p < .0001$. This establishes that the task adaptations did not fundamentally change
756 the nature of the task.

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

757 *3.2.2. Decision making task latencies*

758 More crucial is the time participants needed to solve the different problems. For
759 each participant we calculated the mean time needed to correctly solve the incongru-
760 ent, congruent, and neutral problems. Latencies for erroneously solved incongruent
761 problems were also entered in the analyses. As Fig. 3 shows, the decision making
762 time for the four types of problems clearly differed, $F(3,57) = 7.98$, $MSE = 45.77$,
763 $p < .001$. As expected, the congruent problems were solved fastest. People needed
764 more time to solve the neutral problems where the heuristic system does not cue a
765 response and correct responding requires analytic base rate reviewing. As one would
766 predict people needed even more time to override the erroneous heuristic response
767 and select the correct answer on the incongruent problems. The crucial finding is that
768 even when an incongruent problem was solved incorrectly, people spent more time
769 processing it than when solving the congruent problems. Newman–Keuls tests estab-
770 lished that the decision making time of incorrectly solved incongruent problems fell
771 in between that of correctly solved incongruent and congruent problems.

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

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

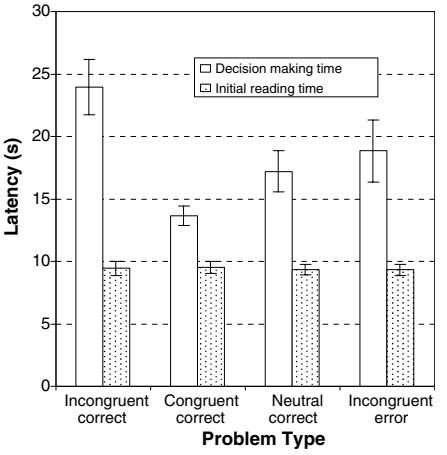

Fig. 3. Mean time (s) needed to make four crucial types of decisions. The time needed to read the preambles (initial reading time) in each of the four cases is also presented. Error bars are standard errors.

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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.

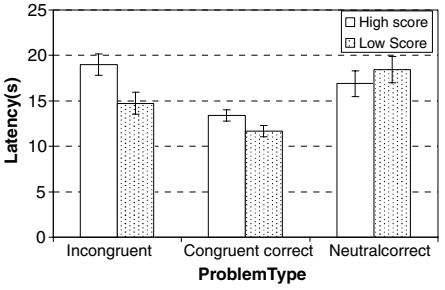

Fig. 4. Decision making time (s) as a function of the accuracy on the incongruent problems. The low score group are those people who failed to solve any of the incongruent problems correctly. Error bars are standard errors.

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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

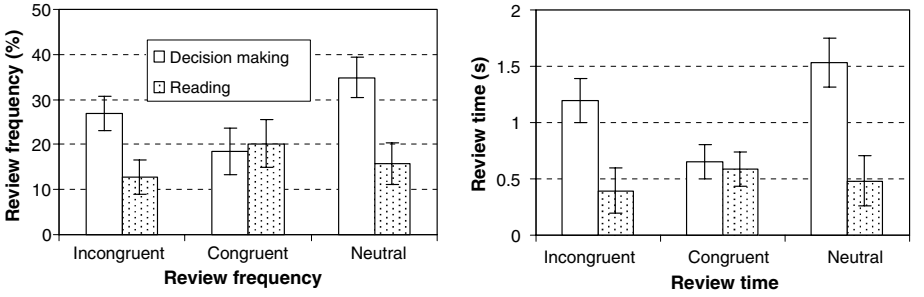

Fig. 5. Mean proportion of base rate reviewing and base rate reviewing time (s) for the three different problem types. Error bars are standard errors.

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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,

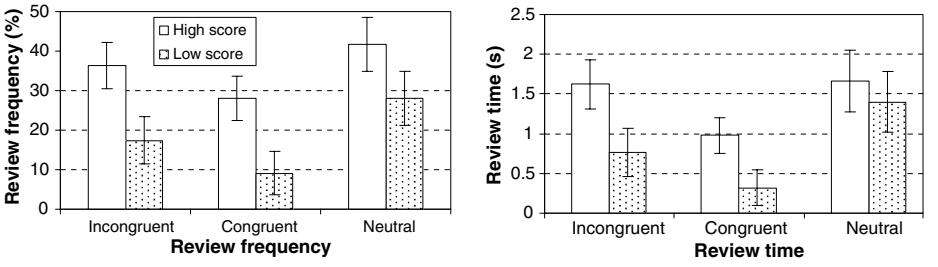

Fig. 6. Mean proportion of base rate reviewing and review time (s) for the three different problem types as a function of performance on the incongruent problems. The low score group are those people who erred on all incongruent problems. Error bars are standard errors.

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

873 $F(1,80) = 2.49$, $MSE = .05$. The increased recall performance for incongruent and
874 neutral problems in the decision making group did exceed the performance of the
875 reading group, $F(1,80) = 4.67$, $MSE = .05$, $p < .05$.
876 We also verified whether the recall results during decision making differed in terms
877 of the accuracy on the incongruent problems. As with the review findings, we com-
878 pared the performance of the high and low scoring group. Results are presented in
879 Fig. 8. Overall, recall performance of people in the high score group was better than
880 that of the people who always solved the incongruent problems incorrectly,
881 $F(1,40) = 24.41$, $MSE = .05$, $p < .0001$. There was also a main effect of problem
882 type, $F(2,80) = 6.97$, $MSE = .05$, $p < .005$, but as in the base rate review analysis,
883 both factors did not interact, $F(2,80) = 2.07$, $MSE = .05$. The two score groups
884 showed the same basic recall trend. Even people who solved all incongruent prob-
885 lems incorrectly managed to correctly recall the direction of the base rates on more
886 than 75% of the incongruent trials. As Fig. 8 suggests, if anything the superior recall
887 on incongruent and neutral vs. congruent problems even tended to be somewhat
888 more pronounced for the low score group.

889 *3.3. Conclusions*

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

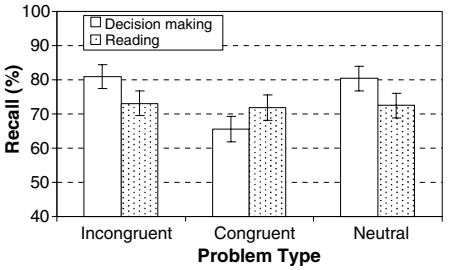

Fig. 7. Mean proportion correct base rate recall after decision making and mere reading. Error bars are standard errors.

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

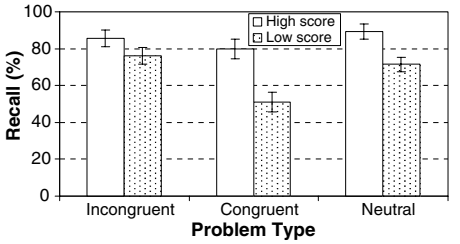

Fig. 8. Mean proportion correct base rate recall as of function of the accuracy on the incongruent problems. Error bars are standard errors.

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

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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.

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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 (2007) 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 (2007) 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

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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.

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

We noted that dual process theorists are not very explicit about the exact nature of the architecture they propose. Evans (2007) 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

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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.,

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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 (in press), 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.

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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.

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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.

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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.

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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.

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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.

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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.

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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*.

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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*.

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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.

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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.

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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.

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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*.

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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.

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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.

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

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. (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking and Reasoning*.

Evans, J. St. B.T. (in press). The heuristic-analytic theory of reasoning: Extension and Evaluation. *Psychonomic Bulletin and Review*.

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.

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002

2007 Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237–251.
2008 MacDonald, A. W., Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000). Dissociating the role of the
2009 dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science, 288*, 1835–1838.
2010 Markovits, H., & Doyon, C. (2004). Information processing and reasoning with premises that are
2011 empirically false: Interference, working memory, and processing speed. *Memory and Cognition, 32*,
2012 592–601.
2013 Moors, A., & de Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological*
2014 *Bulletin, 132*, 297–326.
2015 Moutier, S., & Houdé, O. (2003). Judgements under uncertainty and conjunction fallacy inhibition
2016 training. *Thinking and Reasoning, 9*, 185–201.
2017 Oaksford, M., & Chater, N. (1998). *Rationality in an uncertain world: Essays on the cognitive science of*
2018 *human reasoning*. Hove, UK: Psychology Press.
2019 Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin & Review,*
2020 *11*, 988–1010.
2021 Pashler, H., Johnston, J., & Ruthruff, E. (2001). Attention and performance. *Annual Review of Psychology,*
2022 *52*, 629–651.
2023 Payne, J. W. (1994). Thinking aloud: Insights into information processing. *Psychological Science, 5*,
2024 241–248.
2025 Reyna, V. F., Lloyd, F. J., & Brainerd, C. J. (2003). Memory, development, and rationality: An integrative
2026 theory of judgement and decision making. In S. Schneider & J. Shanteau (Eds.), *Emerging perspectives*
2027 *on judgment and decision research*. New York: Cambridge University Press.
2028 Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial
2029 frontal cortex in cognitive control. *Science, 306*, 443–447.
2030 Shiffrin, R. M. (1988). Attention. In R. C. Atkinson, R. J. Herrnstein, G. Lindsay, & R. D. Luce (Eds.).
2031 *Steven's handbook of experiment psychology* (Vol. 2, pp. 739–811). Oxford, England: Wiley.
2032 Simoneau, M., & Markovits, H. (2003). Reasoning with premises that are not empirically true: Evidence
2033 for the role of inhibition and retrieval. *Developmental Psychology, 39*, 964–975.
2034 Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*, 3–22.
2035 Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality
2036 debate. *Behavioral and Brain Sciences, 23*, 645–726.
2037 Stein, E. (1996). *Without good reason: The rationality debate in philosophy and cognitive science*. Oxford,
2038 England: Oxford University Press.
2039 Todd, P. M., & Gigerenzer, G. (2000). Precis of simple heuristics that make us smart. *Behavioral and Brain*
2040 *Sciences, 23*, 727–780.
2041 Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science, 185*,
2042 1124–1131.
2043

Please cite this article in press as: De Neys, W., & Glumicic, T., Conflict monitoring in dual process ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.06.002