Cognition 158 (2017) 90–109

Contents lists available at ScienceDirect

Cognition

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

Original Articles

Fast logic?: Examining the time course assumption of dual process theory

Bence Bago [a,b,*], Wim De Neys [a,b,c]

[a] Paris Descartes University, Sorbonne Paris Cité, UMR 8240, LaPsyDÉ, France
[b] Caen Basse-Normandie University, UMR 8240, LaPsyDÉ, France
[c] CNRS, UMR 8240, LaPsyDÉ, France

ARTICLE INFO

Article history:
Received 4 May 2016
Revised 12 October 2016
Accepted 25 October 2016

Keywords:
Dual process theory
Reasoning
Conflict detection

ABSTRACT

Influential dual process models of human thinking posit that reasoners typically produce a fast, intuitive heuristic (i.e., Type-1) response which might subsequently be overridden and corrected by slower, deliberative processing (i.e., Type-2). In this study we directly tested this time course assumption. We used a two response paradigm in which participants have to give an immediate answer and afterwards are allowed extra time before giving a final response. In four experiments we used a range of procedures (e.g., challenging response deadline, concurrent load) to knock out Type 2 processing and make sure that the initial response was intuitive in nature. Our key finding is that we frequently observe correct, logical responses as the first, immediate response. Response confidence and latency analyses indicate that these initial correct responses are given fast, with high confidence, and in the face of conflicting heuristic responses. Findings suggest that fast and automatic Type 1 processing also cues a correct logical response from the start. We sketch a revised dual process model in which the relative strength of different types of intuitions determines reasoning performance.

1. Introduction

Decades of research on reasoning and decision-making have indicated that educated adult reasoners often violate elementary logical or probabilistic rules. As an example, consider that there is an event with 1000 people, you are told that most people at the event are I.T. technicians, but there are also 5 attendees who are professional boxers. Assume that you are searching for someone you do not know and you are only given one piece of information; the person is described to you as being 'strong'. What do you think is more likely? Is this person a boxer or an I.T. technician?

On the basis of the base rate probabilities, one might say that the person is an I.T. technician because there are much more I.T. technicians than boxers at the event. However, intuitively people will be tempted to conclude that the person is a boxer based on the stereotypical association (''I.T. technicians are weak") that the description cues. Many studies have shown that people tend to neglect the base rates in these situations (e.g., Pennycook, Trippas, Handley, & Thompson, 2014; Tversky & Kahneman, 1974). Hence, participants typically base their choice on the stereotypical association and conclude that that the person is a boxer. Such intuitive or ''heuristic" associations have been shown to bias people's judgment in a wide range of tasks and situations (Gilovich, Griffin, & Kahneman, 2002).

One of the possible explanations for the phenomenon is presented by dual process theories of thinking. According to the classic dual process view, there are two different types of thinking: Type 1 and Type 2 processes. Type 1 processing is fast, autonomous, does not require working memory, operates unconsciously and immediately triggers an answer. Type 2 processing puts a heavy load on working memory, operates consciously, controlled and relatively slow. The two types of processes are also often referred to as 'intuitive' or 'heuristic' vs. 'deliberate' or 'analytical' (Stanovich & Toplak, 2012). It is important to note that dual process theory is an umbrella term; several types of dual process theories exist (Stanovich & West, 2000). In this study, we focus on the influential, default-interventionist view of dual processes that has been advocated in the seminal work of Evans and Stanovich (2013) and Kahneman (2011).

The standard assumption in the default-interventionist dual process (DI) framework is that the automatic and fast Type 1 process first produces an intuitive heuristic answer. Generation of the heuristic answer might subsequently be followed by a deliberative, slow Type 2 process, which may result in a correction of the initial heuristic answer. Note that in cases - such as the introductory reasoning problem - in which the initial heuristic response conflicts

* Corresponding author at: Paris Descartes University, 46 Rue Saint-Jacques, FR-75005 Paris, France.

E-mail address: bencebagok@gmail.com (B. Bago).

http://dx.doi.org/10.1016/j.cognition.2016.10.014
0010-0277/

with the correct logical[1] response, the corrective Type 2 thinking is believed to be critical to arrive at the correct logical answer. In cases where the Type 2 processing fails, the heuristic response will not be corrected and the reasoner will end up giving the erroneous heuristic answer. Thus, the expected time course assumption is that reasoners will first generate a heuristic answer and, if needed, will after additional reflection correct this to arrive at the correct logical response.

To avoid confusion it is important to stress that the DI time-course prediction does not entail that Type 1 processing necessarily results in an incorrect response or that Type 2 processing necessarily results in a correct response. Normative correctness is not a defining feature of Type 1 or Type 2 processing (e.g., it is not because a response is correct that it resulted from Type 2 processing, and Type 2 processing does not necessarily result in a correct response; e.g., Evans, 2012; Evans & Stanovich, 2013; Stanovich & Toplak, 2012). For example, sometimes reasoners might err precisely because their cognitive resources are overburdened by too much deliberation (e.g., Evans, 2010; Stanovich, 2011). Likewise, it is not hard to see that a person who is guessing can end up giving a correct response without engaging in any deliberation. The DI time course prediction concerns the processing of the typical reasoner in the prototypical situation in which a cued heuristic response conflicts with the correct logical response such as it has been studies in numerous classic tasks from the reasoning and decision-making field since the 1960s. In this case the DI model clearly entails that the typical reasoner will need to recruit Type 2 thinking to correct the initial heuristic Type 1 response in order to arrive at a correct response. Indeed, it is precisely the failure to engage in Type 2 processing that DI theorists have put forward as the primary cause of the massive "bias" in these tasks (Evans, 2012; Kahneman, 2011; Stanovich & West, 2000). Nevertheless, it is important to keep in mind that dual process theories do not claim that one can universally equate Type 2 processing with normative correctness.

But unfortunately, and perhaps somewhat surprisingly, there is little evidence in the literature that allows us to directly validate the core DI time course assumption. For example, in one study De Neys (2006a) presented participants with a range of classic reasoning problems in which a cued heuristic response conflicted with the correct logical response and recorded response latencies. Results consistently showed that correct responses were given much slower than heuristic (i.e., incorrect) responses. One might argue that this finding is in agreement with the time course assumption. Giving a (correct) response that is assumed to result from slow Type 2 processing takes more time than giving an (incorrect) response that is assumed to result from fast Type 1 processing. However, although this fits with the claim that Type 2 processing is slower than Type 1 processing, it does not imply that someone who engaged in Type 2 reasoning first engaged in Type 1 reasoning. The latency data does not imply that correct reasoners generated the incorrect answer first, and then corrected it. Reasoners who complete Type 2 thinking might give the correct response without ever having considered the incorrect, heuristic response.

In another illustrative study, Evans and Curtis-Holmes (2005) used an experimental design in which people had to judge the logical validity of reasoning problems under time pressure; one group of reasoners were given only 2 s to answer, whereas a control group were allowed to take as much time as they wanted to give an answer. A higher percentage of incorrect answers was found in the time pressure group. Hence, this also indicates that giving the correct response requires time. However, this does not necessarily show that individuals who gave the correct response in the free time condition generated the heuristic response first and corrected this subsequently. As with the latency data of De Neys (2006a), it might be that reasoners engaged in Type 2 thinking right away, without any need to postulate an initial generation of a heuristic response.

One might note that there is also some incidental evidence for the DI time course assumption. For example, Frederick (2005) notes that when participants solve his Cognitive Reflection Test (which was designed to cue a strong heuristic response), correct responders often considered the incorrect, heuristic answer first "as is apparent from introspection, verbal reports, and scribbles in the margin" (Frederick, 2005, p. 27). But unfortunately, he gives no further information about the protocol analysis or the precise prevalence of these observations. Frederick also mentions that incorrect responders rate the problems as easier than correct responders and suggests that this presumably indicates that correct responders are more likely to consider both responses. But even when this assumption holds, it does clearly not imply that correct responders considered the heuristic response *before* the correct response.

Arguably, the most direct evidence to evaluate the dual process time course assumption comes from experiments using the two response paradigm (Newman, Gibb, & Thompson, submitted for publication; Pennycook & Thompson, 2012; Thompson & Johnson, 2014; Thompson, Prowse Turner, & Pennycook, 2011). In this paradigm, participants are presented with a reasoning problem and are instructed to respond as quickly as possible with the first, intuitive response that comes to mind. Afterwards, they are presented with the problem again, and they are given as much time as they want to think about it and give a final answer. A key observation for our present purposes was that Thompson and colleagues noted that people spent little time rethinking their answer in the second stage and hardly ever changed their initial response. Note that the fact that people do not change an initial heuristic response is not problematic for the dual process framework, of course. It just implies that people failed to engage the optional Type 2 processing. Indeed, since such failures to engage Type 2 are considered a key cause of incorrect responding, a dominant tendency to stick to incorrect initial responses is not surprising from the classic dual process stance. However, the lack of answer change tentatively suggests that in those cases where a correct logical response was given as final response, the very same response was generated from the start. Bluntly put, the logical response might have been generated fast and intuitively based on mere Type 1 processing (Pennycook & Thompson, 2012; Thompson & Johnson, 2014). This would pose a major challenge for standard dual process theory. However, it cannot be excluded that Thompson et al.'s participants engaged in Type 2 processing when they gave their first, initial response. Although Thompson et al. instructed participants to quickly give the first response that came to mind, participants might have simply failed to respect the instruction and ended up with a correct response precisely because they recruited Type 2 thinking.[2] Clearly, researchers have to make absolutely sure that only Type 1 processing is engaged at the initial

[1] Note that we will be using the label "correct" or "logical" response as a handy shortcut to refer to "the response that has traditionally been considered as correct or normative according to standard logic or probability theory". The appropriateness of these traditional norms has sometimes been questioned in the reasoning field (e.g., see Stanovich & West, 2000, for a review). Under this interpretation, the heuristic response should not be labeled as "incorrect" or "biased". For the sake of simplicity we stick to the traditional labeling. In the same vein, we use the term "logical" as a general header to refer both to standard logic and probability theory.

[2] Note that Thompson et al. obviously realized this and tried to control for it. For example, they always asked participants to verify that their first response was really the one that came to mind first, and they discarded the rare trials with negative verification answers. However, there is no way to be sure that participants' verification answer was true or not. The problem is not so much that people might be intentionally lying but simply that they might have little explicit insight into which thought was generated first. The point here is that a more stringent control is needed.

response stage.

In follow-up work Newman et al. (submitted for publication) have started to address this problem by giving participants a challenging response deadline to enter their initial response. One critical observation was that even in the initial response stage, people showed some sensitivity to the logical status of the problems. For example, participants were slightly more likely to accept valid than invalid inferences even when responding under deadline. Although this logical discrimination ability was more pronounced after additional reflection in the final response stage, the initial sensitivity suggests that to some degree they processed the logical status of the problems intuitively. Nevertheless, a critic can always argue that the deadline was not demanding enough to exclude all Type 2 processing.

There is also some indirect evidence that could make one suspicious of the central time course assumption of dual process theory. One source of evidence comes from recent studies on conflict detection during thinking that try to determine whether biased reasoners notice that their heuristic answer violates logical principles (see De Neys, 2014, 2015, for review; Pennycook, Fugelsang, & Koehler, 2015). Therefore these studies typically contrast reasoners' processing of conflict and control no-conflict problems. Conflict problems are constructed such that a heuristically cued intuitive response conflicts with the correct logical response. In the control no-conflict problem this conflict is not present and the cued heuristic response is also logically correct. For example, the introductory base rate neglect problem that we presented above was a conflict problem; the description will cue a heuristic response that conflicts with the response that is cued by considerations of the base-rates. A no-conflict version of this problem can be constructed by simply reversing the base rates (i.e., 995 boxers/5 I.T. technicians). In this case the answer cued by the base rates, and the heuristic answer cued by the description, are pointing to the same conclusion: the person we are looking for is a boxer. The conflict detection studies have shown that biased reasoners who fail to give the correct response to the conflict problems typically do show sensitivity to the presence of conflict. For example, when solving conflict (vs. no-conflict) problems even incorrect responders show elevated response times (e.g., Bonner & Newell, 2010; De Neys & Glumicic, 2008; Stupple, Ball, & Ellis, 2013; Villejoubert, 2009), decreased post-decision confidence (e.g., De Neys, Cromheeke, & Osman, 2011; De Neys, Rossi, & Houdé, 2013; Gangemi, Bourgeois-Gironde, & Mancini, 2015), and increased activation in brain regions believed to mediate conflict detection (De Neys, Vartanian, & Goel, 2008; Simon, Lubin, Houdé, & De Neys, 2015). The fact that heuristic responders are sensitive to the presence of conflict between their heuristic answer and logical considerations has led some authors to suggest that some elementary logical processing might be occurring from the start of the reasoning process (e.g., De Neys, 2012, 2014; see also Pennycook et al., 2015).

Related indirect evidence comes from work by Handley and colleagues (e.g., (Handley, Newstead, & Trippas, 2011; Handley & Trippas, 2015; Pennycook, Trippas, et al., 2014; Trippas, Handley, Verde, & Morsanyi, 2016). For example, in one of their studies participants were given syllogistic reasoning problems in which the believability of the conclusion could conflict with its logical validity (e.g., a believable but invalid conclusion such as "All flowers need water. Roses need water. Roses are flowers"). It has long been established that when asked to judge the logical validity of such problems, people will tend to be biased by intuitive, heuristic associations based on the believability of the conclusion. Hence, they will be tempted to accept the invalid conclusion simply because it fits with their prior beliefs. However, one of Handley et al.'s (2011) key manipulations was to explicitly instruct participants to give the heuristically cued response. That is, participants were asked to quickly judge whether the conclusion was believable or not without making any reference to logical reasoning. Interestingly, Handley et al. observed that the logical status of the conclusion nevertheless affected people's believability judgments. People had more difficulty (i.e., took longer and made more errors) judging the believability of the conclusion when it conflicted with its logical validity. Bluntly put, although there was no reason to engage in logical reasoning, people couldn't help to do so. These findings led Handley and colleagues. to suggest that the logical response might be generated simultaneously with the heuristic, belief-based response by fast and automatic Type 1 processing.

Taken together, previous literature has not provided sufficient evidence for the critical time course assumption of the DI dual process model, and indirect evidence has led some authors to challenge this presumption as well. For completeness, one might note that the indirect evidence is not uncontroversial either (e.g., Aczel, Szollosi, & Bago, 2016; Klauer & Singmann, 2013; Mata, Schubert, & Ferreira, 2014; Pennycook, Fugelsang, & Koehler, 2012; Singmann, Klauer, & Kellen, 2014; Travers, Rolison, & Feeney, 2016). However, the point is that there is a strong need for the field to validate the time course assumption directly. As Newman et al. (submitted for publication), the present study focuses on this key issue. For this purpose, we adopted a two response paradigm. Participants were asked to give an immediate first answer, and then they were allowed to take as much time as they needed to give a final answer. We were specifically interested in the correctness of the initially generated answers and used a range of methodological procedures to make sure that the initial response was truly intuitive in nature.

Default-interventionist (DI) dual process theory would predict that people typically give the heuristic answer for the first response, which is the incorrect answer in the case of conflict problems. Afterwards, when sufficient time is allotted for Type 2 processing to occur, they might be able to correct their initial response and arrive at the correct answer. In sum, in principle there should be only two main answer types according to standard DI theory: either incorrect for first response – incorrect for second response or incorrect for first response – correct for second response. Our key question is whether generation of a correct final response is indeed preceded by generation of an initial incorrect response or whether people can generate the correct logical answer for the first answer as well. This latter pattern would provide direct evidence for the existence of fast, logical Type 1 reasoning.

Critically, we wanted to make sure and validate that the first response that participants gave only reflected the output of Type 1 processing. For this reason, in four experiments we used a combination of techniques that allowed us to minimize or control the impact of Type 2 processing. Experiment 1 (instructions only), served as a baseline condition in which we merely instructed participants to give their very first intuitive response and answer as quickly as possible. In Experiment 2 (response deadline), we made sure to avoid that (some) participants might take too long to give their first response by enforcing a strict and challenging response deadline. In Experiment 3 (cognitive load) we knocked out Type 2 processing experimentally by imposing a cognitive load task that burdened participants' executive resources. In Experiment 4 (deadline + load) we combined both the response deadline and a cognitive load.

Finally, to check the generality of the findings, two different reasoning tasks were used; a syllogistic reasoning and a base rate task. These were selected because of two reasons: first, these tasks are highly popular in the research community and have inspired much of the theorizing in the field. Second, the tasks are different in the sense that different normative systems are required to solve them correctly (standard logic for syllogistic reasoning, and probability theory for base rate task). The differences or similarities between the tasks can start to give us an indication of the generality of the findings.

2. Method

As we outlined above, we ran four studies that used a range of methodological procedures to make sure that the initial response was truly intuitive in nature. The rationale is that if the initial response results from automatic, intuitive processing, results should not be affected by any of these manipulations and findings should be similar across the four studies. To pre-empt the results, we indeed observed that the findings hardly varied across our studies. For ease of presentation we will present a single results section in which the study factor is included as a between-subject factor in the analyses. Here we present an overview of the method sections of the four studies. Every participant was allowed to take part in only one experiment.

2.1. Experiment 1 – instructions only

2.1.1. Participants

A total of 101 participants were tested (61 female, Mean age = 38.95, SD = 12.69). The participants were recruited via the Crowdflower platform, and received $0.30 for their participation. Only native English speakers from the USA or Canada were allowed to participate in the study. A total of 48% of the participants reported high school as highest completed educational level, while 51% reported having a post-secondary education degree (1% did not answer).

2.1.2. Materials

2.1.2.1. Base rate task. Participants solved a total of eight base-rate problems. All problems were taken from Pennycook, Cheyne, Barr, Koehler, and Fugelsang (2014). Participants always receive a description of the composition of a sample (e.g., "This study contained I.T. engineers and professional boxers"), base rate information (e.g., "There were 995 engineers and 5 professional boxers") and a description that was designed to cue a stereotypical association (e.g., "This person is strong"). Participants' task was to indicate to which group the person most likely belonged.

The problem presentation format we used in this research was based on Pennycook et al.'s (2014) rapid-response paradigm. In this paradigm, the base rates and descriptive information are presented serially and the amount of text that is presented on screen is minimized. Pennycook et al. introduced the paradigm to minimize the influence of reading times and get a purer and less noisy measure of reasoning time per se. First, participants received the names of the two groups in the sample (e.g., "This study contains clowns and accountants"). Next, under the first sentence (which stayed on the screen) we presented the descriptive information (e.g., Person 'L' is funny). The descriptive information specified a neutral name ('Person L') and a single word personality trait (e.g., "strong" or "funny") that was designed to trigger the stereotypical association. Finally, participants received the base rate probabilities. The following illustrates the full problem format:

This study contains clowns and accountants.
Person 'L' is funny.
There are 995 clowns and 5 accountants.
Is Person 'L' more likely to be:
∘ A clown
∘ An accountant

Half of the presented problems were conflict items and the other half were no-conflict items. In no-conflict items the base rate probabilities and the stereotypic information cued the same response. In conflict items the stereotypic information and the base rate probabilities cued different responses. Three kinds of base rates were used: 997/3, 996/4, 995/5.

Note that the material that was selected for the present study was extensively pretested. Pennycook et al. (2014) made sure that words that were selected to cue a stereotypical association consistently did so but avoided extremely diagnostic cues. The importance of such a non-extreme, moderate association is not trivial. Note that we label the response that is in line with the base rates as the correct response. Critics of the base rate task (e.g., Gigerenzer, Hell, & Blank, 1988; see also Barbey & Sloman, 2007) have long pointed out that if reasoners adopt a Bayesian approach and combine the base rate probabilities with the stereotypic description, this can lead to interpretational complications when the description is extremely diagnostic. For example, imagine that we have an item with males and females as the two groups and give the description that Person 'A' is 'pregnant'. Now, in this case, one would always need to conclude that Person 'A' is a woman, regardless of the base rates. The more moderate descriptions (such as 'kind' or 'funny') help to avoid this potential problem. In addition, the extreme base rates (997/3, 996/4, or 995/5) that were used in the current study further help to guarantee that even a very approximate Bayesian reasoner would need to pick the response cued by the base-rates (see De Neys, 2014).

Each problem started with the presentation of a fixation cross for 1000 ms. After the fixation cross disappeared, the sentence which specified the two groups appeared for 2000 ms. Then the stereotypic information appeared, for another 2000 ms, while the first sentence remained on the screen. Finally, the last sentence specifying the base rates appeared together with the question and two response alternatives. Note that we presented the base-rates and question together (rather than presenting the base-rate for 2000 ms first) to minimize the possibility that some participants would start solving the problem during presentation of the base-rate information. Once the base-rates and question were presented participants were able to select their answer by clicking on it. The position of the correct answer alternative (i.e., first or second response option) was randomly determined for each item. The eight items were presented in random order. An overview of the full item set can be found in the Supplementary material.

2.1.2.2. Syllogistic reasoning task. Participants were given eight syllogistic reasoning problems taken from De Neys, Moyens, and Vansteenwegen (2010). Each problem included a major premise (e.g., "All dogs have four legs"), a minor premise (e.g., "Puppies are dogs"), and a conclusion (e.g., "Puppies have four legs"). The task was to evaluate whether the conclusion follows logically from the premises. In four of the items the believability and the validity of the conclusion conflicted (conflict items, two problems with an unbelievable–valid conclusion, and two problems with a believable–invalid conclusion). For the other four items the logical validity of the conclusion was in accordance with its believability (no-conflict items, two problems with a believable–valid conclusion, and two problems with an unbelievable–invalid conclusion). We used the following format:

All dogs have four legs
Puppies are dogs
Puppies have four legs
Does the conclusion follow logically?
∘ Yes
∘ No

The premises and conclusion were presented serially. Each trial started with the presentation of a fixation cross for 1000 ms. After the fixation cross disappeared, the first sentence (i.e., the major premise) was presented for 2000 ms. Next, the second sentence (i.e., minor premise) was presented under the first premise for 2000 ms. After this interval was over, the conclusion together with the question "Does the conclusion follow logically?" and two response options (yes/no) was presented right under the premises. Once the conclusion and question were presented, participants could give their answer by clicking on the corresponding bullet point. The eight items were presented in a randomised order. An overview of the full item set can be found in the Supplementary material.

2.1.3. Procedure

The experiment was run online. People were clearly instructed that we were interest in their first, initial response to the problem. Instruction stressed that it was important to give the initial response as fast as possible and that participants could afterwards take additional time to reflect on their answer. The literal instructions that were used, stated the following:

> "Welcome to the experiment! Please read these instructions carefully!
> This experiment is composed of 16 questions and a couple of practice questions. It will take about 20 minutes to complete and it demands your full attention. You can only do this experiment once.
> In this task we'll present you with a set of reasoning problems. We want to know what your initial, intuitive response to these problems is and how you respond after you have thought about the problem for some more time. Hence, as soon as the problem is presented, we will ask you to enter your initial response. We want you to respond with the very first answer that comes to mind. You don't need to think about it. Just give the first answer that intuitively comes to mind as quickly as possible. Next, the problem will be presented again and you can take all the time you want to actively reflect on it. Once you have made up your mind you enter your final response. You will have as much time as you need to indicate your second response.
> After you have entered your first and final answer we will also ask you to indicate your confidence in the correctness of your response. In sum, keep in mind that it is really crucial that you give your first, initial response as fast as possible. Afterwards, you can take as much time as you want to reflect on the problem and select your final response. You will receive $0.30 for completing this experiment. Please confirm below that you read these instructions carefully and then press the "Next" button."

All participants were presented with both the syllogistic reasoning and base-rate task in a randomly determined order. After the general instructions were presented the specific instructions for the upcoming task (base-rates or syllogisms) were presented. The following specific instructions were used for the syllogistic reasoning task:

> "In this part of this experiment you will need to solve a number of reasoning problems. At the beginning you are going to get two premises, which you have to assume being true. Then a conclusion will be presented. You have to indicate whether the conclusion follows logically from the premises or not. You have to assume that the premises are all true. This is very important.
> Below you can see an example of the problems.
> Premise 1: All dogs have four legs
> Premise 2: Puppies are dogs
> Conclusion: Puppies have four legs
> *Does the conclusion follow logically?*
> ○ Yes
> ○ No
> The two premises and the conclusion will be presented on the screen one by one. Once the conclusion is presented you can enter your response.
> As we told you we are interested in your initial, intuitive response. First, we want you to respond with the very first answer that comes to mind. You don't need to think about it. Just give the first answer that intuitively comes to mind as quickly as possible. Next, the problem will be presented again and you can take all the time you want to actively reflect on it. Once you have made up your mind you enter your final response. After you made your choice and clicked on it, you will be automatically taken to the next page. After you have entered your first and final answer we will also ask you to indicate your confidence in the correctness of your response. Press "Next" if you are ready to start the practice session!"

For the base-rate task these instructions were presented:

> "In a big research project a large number of studies were carried out where a psychologist made short personality descriptions of the participants. 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 one personality trait 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. As we told you we are interested in your initial, intuitive response. First, we want you to respond with the very first answer that comes to mind. You don't need to think about it. Just give the first answer that intuitively comes to mind as quickly as possible. Next, the problem will be presented again and you can take all the time you want to actively reflect on it. Once you have made up your mind you enter your final response. After you made your choice and clicked on it, you will be automatically taken to the next page. After you have entered your first and final answer we will also ask you to indicate your confidence in the correctness of your response.
> Press "Next" if you are ready to start the practice session!"

After the task specific instructions, participants solved two practice (no-conflict) problems to familiarize them with the task. Then they were able to start the experiment. For the first response people were instructed to give a quick, intuitive response. After they clicked on the answer, they were asked to give their confidence in their answer, on a scale from 0% to 100%, with the following question: "How confident are you in your answer? Please type a number from 0 (absolutely not confident) to 100 (absolutely confident)". Next, they were presented with the problem again, and they were told that they could take as much time as they needed to give a final answer. As a last step, they were asked to give the confidence in their final answer (the same question format as for first answer confidence was used).

The colour of the actual question and answer options were green during the first response, and they were blue during the second response phase, to visually remind participants which question they were answering at the moment. For this purpose, right under the question a reminder sentence was placed: "Please indicate your very first, intuitive answer!" and "Please give your final answer." respectively.

The presentation order of the base rate and syllogistic reasoning tasks was randomized. After participants finished the first task they could briefly pause, were presented with the instructions and practice problems of the second task, and started the second task. For both the base-rate and syllogistic reasoning task two different problem sets were used. The conflict items in one set were the no-conflict items in the other, and vice versa. This was done by reversing the base-rates (base-rate task) or by switching the conclusion and minor premise (syllogisms, see De Neys et al., 2010). Each of the two sets was used for half of the participants. Supplementary material section A and B gives an overview of all problems in each of the sets. This counterbalancing minimized the possibility that mere content or wording differences between conflict and no-conflict items could influence the results. At the end of the study participants were asked to answer demographic questions.

2.2. Experiment 2 – response deadline

2.2.1. Participants

In the actual experiment (response deadline), 104 participants were recruited (63 female, $M = 39.9$ years, $SD = 13.31$ years). An additional 52 participants (31 female, $M = 44.13$, $SD = 13.2$) were recruited for a reading pretest (i.e., reading pretest; see further). Participants received $0.11 for their participation in the reading pretest and $0.50 for participation in the actual experiment. The same recruitment procedure was used as in Experiment 1. In the response deadline condition, 35% of the participants reported high school as highest educational level, 63% reported having a post-secondary education degree, while 2% reported less than high school educational level. In the reading pre-test 40% of the participants reported high school as highest educational level, while 60% of them reported having a post-secondary education degree.

2.2.2. Materials & procedure

2.2.2.1. Reading pre-test. In the reading pretest participants were asked to simply read each one of the base-rate and syllogistic reasoning problems that were used in Experiment 1. The basic goal of this reading condition was to define the response deadline for the actual reasoning study. Our rationale was to base the deadline on the average reading times for the syllogistic reasoning and base-rate problems (see further). Note that as many critics have argued, dual process theories are massively underspecified (Kruglanski, 2013). The theory only posits that Type 1 processes are relatively faster than Type 2 processes. However, no criterion is available that would allow us to a priori characterize a response as a Type 1 response in an absolute sense (i.e., faster than x seconds = Type 1). Our reading baseline provides a practical criterion to define a response deadline. The rationale is simple, if we allot participants only as much time as it takes to read the problems, we can be reasonable sure that reasoning related Type 2 processing will be minimal. Obviously, making a strict distinction between reading and reasoning is not possible but the point here is that the reading pretest will give us a practical criterion that should serve as a reasonable and universally applicable proxy.

As a side-note, note that our reasoning task format was especially selected with an optimization of the deadline in mind. As we clarified, we aimed to minimize the amount of text that was presented on screen. This is again not a trivial issue. In two-response studies with traditional base-rate problems Pennycook and Thompson (Pennycook & Thompson, 2012; Thompson et al., 2011; see also Newman et al., submitted for publication) already tried to introduce a deadline for the first response. However, because of the lengthy nature of the problems, pilot testing indicated that the deadline needed to be set at 12 s. Arguably, such a lengthy response window leaves ample room for a possible impact of Type 2 processing. This is one of the complications that can be sidestepped by adopting Pennycook et al.'s (2014) fast response base-rate format that we used in the present studies. Hence, by minimizing the to-be read text we hoped to minimize reading time and set a much stricter deadline in absolute terms.

Participants were instructed that the goal of the pretest was to determine how long people needed to read item materials. They were instructed that there was no need for them to try to solve the problems and simply needed to read the items in the way they typically would. When they were finished reading, they were asked to randomly click on one of the presented response options to advance to the next problem. Presentation format was the same as in Experiment 1. The only difference was that the problem was not presented a second time and participants were not asked for a confidence rating. To make sure that participants would be motivated to actually read the material we told them that we would present them with two (for both tasks, four in sum) very easy verification questions at the end of the study to check whether they read the material. The literal instructions were as follows:

> "Welcome to the experiment! Please read these instructions carefully! This experiment is composed of 16 questions and 4 practice questions. It will take 5 min to complete and it demands your full attention. You can only do this experiment once. In this task we'll present you with a set of problems we are planning to use in future studies. Your task in the current study is pretty simple: you just need to read these problems. We want to know how long people need on average to read the material. In each problem you will be presented with two answer alternatives. You don't need to try to solve the problems or start thinking about them. Just read the problem and the answer alternatives and when you are finished reading you randomly click on one of the answers to advance to the next problem. The only thing we ask of you is that you stay focused and read the problems in the way you typically would. Since we want to get an accurate reading time estimate please avoid whipping your nose, taking a phone call, sipping from your coffee, etc. before you finished reading. At the end of the study we will present you with some easy verification questions to check whether you actually read the problems. This is simply to make sure that participants are complying with the instructions and actually read the problems (instead of clicking through them without paying attention). No worries, when you simply read the problems, you will have no trouble at all at answering the verification questions.
> You will receive $0.11 for completing this experiment. Please confirm below that you read these instructions carefully and then press the "Next" button."

> Specific instructions before the syllogistic items started:

> "In the first part of this experiment you will need to read a specific type of reasoning problems. At the beginning you are going to get two premises, which you have to assume being true. Then a conclusion, question and answer alternatives will be presented. We want you to read this information and click on any one of the two answers when you are finished. Again, no need to try to solve the problem. Just read it. Below you can see an example of the problems.
> Premise 1: All dogs have four legs
> Premise 2: Puppies are dogs
> Conclusion: Puppies have four legs
> *Does the conclusion follow logically?*
> ○ Yes
> ○ No
> The two premises and the conclusion will be presented on the screen one by one. Once the conclusion is presented, you simply click on one of the answer alternatives when you finished reading and the next problem will be presented. Press "Next" if you are ready to start a brief practice session!"

Specific instructions before the base rate items started:

> "In a big research project a large number of studies were carried out where a psychologist made short personality descriptions of the participants. 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 one personality trait of this randomly chosen participant. You'll also get information about the composition of the population groups tested in the study in question. Then, a question to indicate to which population group the participant most likely belongs will appear. We simply want you to read this question and the two answer alternatives. Once you finished reading this, you simply click on either one of the answer alternatives and then the next problem will be presented. Again no need to try to solve the problem, just read the question and simply click on either one of the answers when you are finished. Press "Next" if you are ready to start a brief practice session!"

The verification questions were constructed such that a very coarse reading of the problems would suffice to recognize the correct answer. The following are examples of the verification questions for the syllogistic and base-rate problems:

> *"We asked you to read the conclusions of a number of problems. Which one of the following conclusions was NOT presented during the task:*
> ◦ *Whales can walk*
> ◦ *Boats have wheels*
> ◦ *Roses are flowers*
> ◦ *Waiters are tired*"

And an example of the verification question for the base rate task:

> *"We asked you to read problems about a number of population groups. Which one of the following combination of two groups was NOT presented during the task:*
> ◦ *Nurses and artists*
> ◦ *Man and woman*
> ◦ *Scientists and assistants*
> ◦ *Cowboys and Indians*"

The correct answer was blatantly unrelated to any of the presented material content. Note that 94% of the verification questions were solved correctly, which indicates that by and large, participants were at least minimally engaged in the reading task. Only those participants were analysed, who correctly solved both verification questions regarding each task.

In sum, the reading condition should give us a baseline against which the reasoning response times for the initial response can be evaluated. Any Type 1 response during reasoning also minimally requires that (a) the question and response alternatives are read, and (b) participants move the mouse to select a response. The reading condition allows us to partial out the time needed for these two components. In other words, the reading condition will gives us a raw indication of how much time a Type 1 response should (minimally) take.

Results of the reading pretest indicated that participants needed on average 2.92 s (SD = 1.95) for base rate problems, and 2.62 s (SD = 1.89) for the syllogistic reasoning problems to read the problems and click on a response option.[3] We rounded this value to the nearest integer (3 s) to give participants some minimal leeway. Hence, we set a universal response deadline of 3 s for the reasoning experiment.

2.2.2.2. Response deadline experiment. The same two reasoning tasks and problems were used as in Experiment 1. The only difference was that a response deadline was introduced to minimize the possibility that participants would engage in time-consuming Type 2 processing when giving their first response. Once the question was presented, participants had 3000 ms to click on one of the answer alternatives and after 2000 ms the background colour turned yellow to remind them to pick an answer immediately. If participants did not select an answer within 3000 ms they got feedback to remind them that they had not answered within the deadline and they were told to make sure to respond faster on subsequent trials. Obviously, there was no response deadline for the second response.

Participants were given 3 (no-conflict) practice problems before starting each task to familiarize them with the deadline procedure. During the actual reasoning task, participants failed to provide a first response within the deadline on 12% of the trials. These missed trials were discarded and were not included in the reported data.

2.3. Experiment 3 – load

2.3.1. Participants

A total of 99 participants were recruited (44 female, M = 39.28, SD = 13.28). The same recruitment procedure was used as in Experiment 1. Participants received $0.50 for their participation. A total of 48% of the participants reported high school as highest educational level, while 43% of them reported having a post-secondary degree. Six percent of the participants did not provide education level information.

2.3.2. Materials & procedure

In Experiment 3 we used a concurrent load task - the dot memorization task (Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001) - to burden participants' executive cognitive resources while they were solving the reasoning tasks. The idea behind the load manipulation is straightforward. One of the defining features of Type 2 processing is that it requires executive (working memory) resources (e.g., Evans and Stanovich, 2013; Kahneman, 2011). Hence, if we burden participants' cognitive resources with a secondary load task while they are solving the reasoning problems, we reduce the possibility that they can engage in Type 2 thinking (De Neys, 2006a; De Neys & Schaeken, 2007; Franssens & De Neys, 2009).

The same two reasoning tasks and problems were use as in Experiment 1. In every trial, after the fixation cross disappeared, participants were shown a matrix in which 4 dots were presented in a complex interspersed pattern in a 3 × 3 grid (see Fig. 1) for 2000 ms. Participants were instructed to memorize the pattern. Previous studies established that this demanding secondary task successfully burdens executive resources during reasoning (De Neys, 2006b; Franssens & De Neys, 2009; Miyake et al., 2001). After the matrix disappeared, the reasoning problem was presented as in Experiment 1 and participants had to give their first response and their response confidence. After this, they were shown four matrices with different dot patterns and they had to select the correct, to-be-memorized matrix (see Fig. 1). Participants were given feedback as to whether they recalled the correct matrix or not. Subsequently, the problem was presented again and participants selected their final response and response confidence. Hence, no load was imposed during the second, final response stage. There was no time limit for either one of the responses. All trials on which an incorrect matrix was selected (11% of trials) were removed from the analysis.

Before the actual experiment participants were familiarized with the task procedure. First, they received two reasoning

[3] Note that reading time averages were calculated on the logarithmically transformed data, but they were transformed back to seconds.

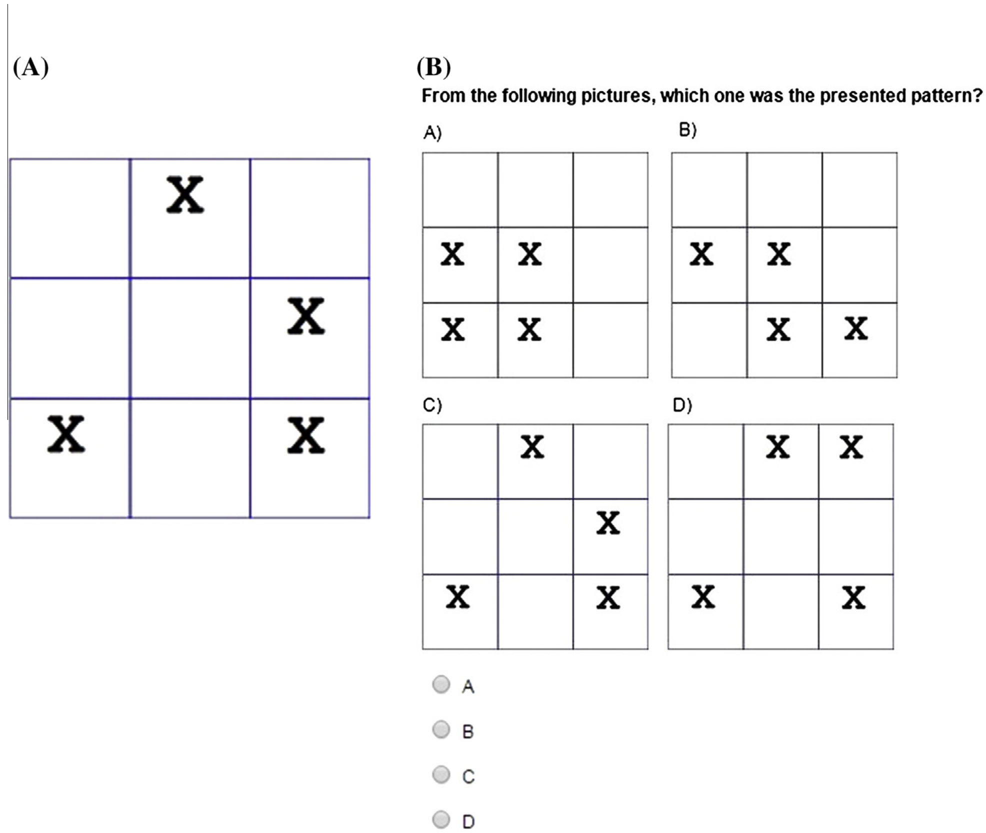

Fig. 1. Example of a dot matrix (left panel, A) and a pattern recall question with the possible answer options (right panel, B).

problems which were identical to the practice questions used in Experiment 1. After, participants were presented with a dot matrix practice question – they were simply shown a dot pattern for 2000 ms and after it disappeared they were asked to identify the pattern from four presented options. As a last step, they were presented two more practice reasoning problems which they needed to solve under load following the procedure outlined above.

2.4. Experiment 4 – response deadline and load

2.4.1. Participants

In Experiment 4 (deadline and load), 115 participants were recruited (53 female, M = 38.85 years, SD = 12.12 years). The same recruitment procedure was used as in Experiment 1. Participants received $0.50 for their participation. A total of 40% of the participants reported high school as highest educational level, 57% of them reported having a post-secondary education degree, and 3% reported less than high school educational level.

2.4.2. Materials & procedure

In Experiment 4, the same load task was used as in Experiment 3, the only exception was that participants had to give both their first and final response under load.[4] Hence, participants had to memorize the dot pattern until they had given their final response and confidence rating. In addition, the same response deadline (3 s) procedure as in Experiment 2 was used to further limit the possibility of Type 2 engagement in the first response stage. Participants in Experiment 4 failed to provide a first response before the deadline on 6.3% of the trials. These missed trials were discarded from the analysis. Trials on which an incorrect matrix was selected (11% of trials) were also removed. Taken together, this resulted in a total number of 16.7% of the trials that were excluded from the analysis.

3. Results

3.1. Accuracy of final response

For consistency with previous work we first present the response accuracies for the final response. Table 1 gives an overview of the results. As the table indicates, accuracies across the four experiments are generally in line with previous studies that adopted a single response paradigm. On the majority of trials, people are typically biased when solving the conflict problems. Average final accuracy across the four experiments is only 40% for the base rate task and 48% for the syllogistic reasoning task (differences between experiments are considered in the next section). However, on the no-conflict problems where heuristic Type 1 processing cues the correct response, participants perform significantly better with average final accuracies reaching 94% for the base-rate task, $b = -4.77$, $Z = -25.06$, $p < 0.001$, and 77% for the syllogistic reasoning task, $b = -1.51$, $Z = -17.07$, $p < 0.001$. Hence, final response accuracies are generally consistent with what can be expected based on the literature that adopted a classic one-response paradigm with these task (e.g., De Neys et al., 2010; Pennycook et al., 2015).

3.2. Direction of change

Our primary interest in the present study (inspired by Pennycook & Thompson, 2012) is what we will refer to as the "Direction of Change" analysis for the conflict items. By direction

[4] The idea was to test whether the additional load during the final response stage would affect the likelihood that people changed from an initial incorrect to final correct answer. However, because of the floored number of such change responses (see further) in the no load conditions, this sub-hypothesis could not be verified.

Table 1
Percentage of correct final responses (SD) on conflict and no-conflict problems in the base-rate and syllogistic reasoning tasks across the four experiments.

Experiment	Base rate		Syllogisms	
	Conflict	No-conflict	Conflict	No-conflict
Experiment 1: Instructions only	36.3% (48.1)	95.4% (22.8)	41.8% (49.4)	77.4% (41.8)
Experiment 2: Response deadline	37% (48.3)	93.7% (24.2)	42.8% (49.6)	77.9% (41.5)
Experiment 3: Load	45.2% (49.8)	95.3% (21.2)	53.6% (49.9)	76.3% (42.6)
Experiment 4: Deadline + Load	42.8% (49.5)	93% (25.4)	54.5% (49.9)	76% (42.7)
Average	40.1% (49)	94.1% (23.5)	47.9% (50)	76.9% (42.1)

Table 2
Percentage of trials within every direction of change categories for conflict items. The raw number of trials in each category is presented between brackets.

Task	Experiment	Direction of change			
		11	00	10	01
Base rate	Experiment 1: Instructions only	27% (108)	61% (244)	2.75% (11)	9.25% (37)
	Experiment 2: Response deadline	25.1% (92)	59.9% (217)	4.4% (16)	11.2% (41)
	Experiment 3: Load	38.8% (122)	52.5% (179)	2.3% (8)	9.4% (32)
	Experiment 4: Deadline + Load	32.7% (127)	53.4% (207)	3.1% (13)	10.7% (41)
	Average	30%	56.7%	3.2%	10.1%
Syllogism	Experiment 1: Instructions only	35.1% (142)	54.7% (221)	3.5% (14)	6.7% (27)
	Experiment 2: Response deadline	38.9% (139)	50.4% (180)	5.3% (19)	5.3% (19)
	Experiment 3: Load	46.2% (156)	43.5% (147)	3% (10)	7.4% (25)
	Experiment 4: Deadline + Load	49.7% (187)	43.1% (162)	2.7% (10)	4.5% (17)
	Average	42.3%	48.1%	3.6%	6%

of change, we mean the way or direction in which a given person in a specific trial changed (or didn't change) her initial answer during the rethinking phase. More specifically, people can give a correct or incorrect response in each of the two response stages. Hence, in theory this can result in four different types of answer change patterns: (1) a person could either give the incorrect (heuristic) answer as the first response, and then change to the correct (logical) answer as the final response (we will use the label "01" to refer to this type of change pattern), (2) one can give the incorrect answer as the first response and final response (we use the label "00" for this type of pattern), (3) one can give the correct answer as the first response and change to the incorrect response as the final response (we use the label "10" for this type of pattern), and (4) one can give the correct answer for the first and final response (we use the label "11" for this pattern). To recap, we will use the following labels to refer to four possible types of answer change patterns: "01" (i.e., response 1 incorrect, response 2 correct), "00" (i.e., response 0 incorrect, response 2 incorrect), "10" (i.e., response 1 correct, response 2 incorrect), and "11" (i.e., response 1 correct, response 2 correct).

Table 2 shows how frequent each of the four types of directions of change were for the critical conflict problems. First thing to note is that for both the base-rate and syllogistic reasoning tasks there are two general trends that support the DI dual process view. First, there is a high prevalence of 00 responses. For both reasoning tasks this was clearly the dominant response category. The dominance of the 00 category in the conflict problems supports DI theory; people will typically tend to stick to the heuristic response which results in an erroneous first response that is subsequently not corrected. Second, we also observe a small number of trials in the 01 category. In line with standard DI theory, sometimes an initial erroneous response will be corrected after additional reflection, but these cases are quite rare. By and large, these trends fit the standard DI predictions. However, a key challenge for the standard DI model is the high frequency of "11" responses (as Table 2 shows, 31% and 42% of responses for base-rate and syllogisms, respectively). Indeed, both for the base-rate and syllogistic reasoning task it was the case that for the *majority* of trials in which the final response was correct, this correct response was already given as the initial response (i.e., 74.8% and 87.6% of the final correct response trials in the base-rate and syllogistic reasoning task, respectively). Hence, in these cases the correct logical response was given immediately.

Moreover, the high prevalence of 11 responses is observed across experiments; by and large all experiments show similar results. A high proportion of 11 responses in Experiment 1 (instructions only) solely could have been attributed to the fact that some participants might simply not respect the instructions. However, the 11 responses remained stable in Experiments 2–4 that minimized the possibility to engage in slow and demanding Type 2 thinking. Indeed, if anything it seems that the proportion of 11 responses is slightly elevated in the two conditions where cognitive load was applied (Experiment 3: Load, and Experiment 4: Load + deadline). Mixed-effect multilevel logistic regression models[5] showed that this trend reached significance in the syllogistic reasoning task, χ^2 (5) = 12.07, p = 0.007, but not in the base-rate task, χ^2 (5) = 6.67, p = 0.083. However, the key point here is that none of the experimental procedures *decreased* the frequency of 11 responses. This indicates that the correct initial responses did not result from additional Type 2 processing.

Another potential explanation for the high prevalence of 11 responses is that they simply result from random guessing. Indeed, the experimental design is challenging for participants; they were asked to produce a very quick answer, and could even be faced with a strict deadline and/or secondary task load. One might argue that the task is simply too demanding and participants are consequently answering randomly when asked to enter an initial response. Obviously, if people guess, they would give a correct

[5] Rationale for this analysis can be found under the section "Confidence and response time analysis". As the dependent variable here was categorical, we used multilevel logistic regression models, instead of regular regression. Note that one might also argue that the analysis should include 10 responses. However, results are not affected. The trend towards more initial correct responses under load reached significance in the syllogistic reasoning task, χ^2 (5) = 10.02, p = 0.02; but not in the base-rate tasks, χ^2 (5) = 5.08, p = 0.17.

initial response at about 50%. Hence, in theory, a high prevalence of 11 responses might result from such random guessing. However, direct evidence against the guessing account comes from the no-conflict problems. As Table 3 shows, responses here are almost exclusively of the 11 type. Across experiments they accounted for 90% (base-rates) and 73% (syllogism) of responses. Mixed-effect multilevel logistic regression models showed that there were no significant differences with respect to the frequency of 11 no-conflict responses across the four experiments, neither in the base rate task, χ^2 (5) = 1.07, $p = 0.78$, nor in the syllogistic reasoning task, χ^2 (5) = 1.73, $p = 0.63$. However, both for the base, $b = -4.49$, $Z = -25.64$, $p < 0.0001$, and syllogistic reasoning task, $b = -1.49$, $Z = -16.97$, $p < 0.0001$, the frequency of 11 responses was clearly higher for the no-conflict than for the conflict problems. Note that this dominance of 11 responses on no-conflict problems is as predicted by DI theory given that the heuristic Type 1 processing is expected to cue the correct response on no-conflict problems. However, the point here is that the pattern directly argues against the guessing account: If our experimental demands were too challenging and participants were simply guessing when giving their initial response, the initial responses should not have differed for conflict and no-conflict problems either. Our stability and confidence analyses below will provide further evidence against the guessing account.

Finally, we would like to note that the consistent low prevalence of 10 and 01 responses that we observed on the conflict trials across tasks and experiments support Thompson et al.'s (2011) and Pennycook and Thompson's (2012) earlier observations that people mostly stick to their initial response and rarely change their answer regardless whether it was correct or not. On average the 10 and 01 categories accounted for less than 11.4% of responses on the conflict trials across tasks and experiments.

In sum, the key challenge for the time course assumption of DI dual process theory of the present direction of change analysis is the high prevalence of "11" response. Although we observed the predicted dominance of 00 responses, we also found that in the majority of cases in which the correct response was given as final answer, this response was already selected as initial answer. This tentatively suggests that in those cases where people arrive at a correct final response, the correct response was already generated intuitively.

3.2.1. Stability index analysis

Our direction of change analysis was computed across items and participants (using a mixed model approach). One might wonder whether participants are stable in their preference for one or the other type of change category. That is, does an individual who produces a correct vs. incorrect response on one conflict problem does so consistently for the other items or are people more variable in their response preferences across problems? To answer this question we calculated for every participant on how many (out of the number of trials they answered) conflict problems they displayed the same direction of change category. We refer to this measure as the *stability index*. For example, if an individual shows the same type of direction of change on all four conflict problems, the stability index would be 100%. If the same direction of change is only observed on two trials, the stability index would be 50%, etc. Table 4 presents an overview of the findings. Note that due to our methodological restrictions (discarding of no response trials under deadline and load trials in which the memorization was not successful) for a small number of participants only 3 responses were available. Here the stability index is calculated over the available items. The table shows the percentage of participants who displayed the same direction of change type on 100%, 75%, 66%, 50%,

Table 3
Percentage of trials within every direction of change categories for no-conflict items. The raw number of trials in each category is presented between brackets.

Task	Experiment	Direction of change			
		11	00	10	01
Base rate	Experiment 1: Instructions only	90.3% (361)	3.5% (14)	2% (8)	4.3% (17)
	Experiment 2: Response deadline	90.6% (338)	3.2% (12)	2.1% (8)	4% (15)
	Experiment 3: Load	91.3% (273)	3.3% (10)	1.3% (4)	4% (12)
	Experiment 4: Deadline + Load	88.2% (290)	4.3% (14)	2.1% (7)	5.5% (18)
	Average	90.9%	3.6%	1.9%	4.5%
Syllogism	Experiment 1: Instructions only	73.8% (298)	18.3% (74)	4.2% (17)	3.7% (15)
	Experiment 2: Response deadline	74.7% (272)	15.4% (56)	4.1% (15)	5.8% (21)
	Experiment 3: Load	72.7% (245)	17.8% (60)	5.9% (20)	3.6% (12)
	Experiment 4: Deadline + Load	70.6% (272)	20% (77)	3.1% (12)	6.2% (24)
	Average	73%	17.9%	4.3%	4.8%

Table 4
Total frequency of stability indexes for each direction of change category. The raw number of participants in each category is presented between brackets.

Task	Experiment	Stability				
		<33%	50%	66%	75%	100%
Base rate	Experiment 1: Instructions only	2% (2)	13% (14)	0	21% (21)	63% (63)
	Experiment 2: Response deadline	2.97% (3)	15.84% (16)	8.91% (9)	14.85% (15)	57.43% (58)
	Experiment 3: Load	4.21% (4)	10.52% (10)	5.26% (5)	22.11% (21)	57.89% (55)
	Experiment 4: Deadline + Load	3.51% (4)	12.28% (14)	10.52% (12)	13.16% (15)	60.53% (69)
	Average	3.17%	12.91%	8.23%	17.78%	59.71%
Syllogism	Experiment 1: Instructions only	0	32.76% (33)	0	23.76% (24)	43.56% (44)
	Experiment 2: Response deadline	1.01% (1)	27.27% (27)	10.1% (10)	26.26% (26)	35.35% (35)
	Experiment 3: Load	2.13% (2)	39.36% (37)	15.96% (15)	14.89% (14)	27.66% (26)
	Experiment 4: Deadline + Load	0.88% (1)	36.84% (42)	16.67% (19)	14.04% (16)	31.57% (36)
	Average	1.34%	34.06%	14.24%	19.74%	34.54%

or <33% of trials. As the table shows, for both reasoning tasks, in all four experiments, the vast majority of participants displayed the exact same type of change on at least two out of 3 or three out of 4 conflict problems. This pattern held across experiments. The average stability index for the base-rate task was 84.9% (*SD* = 20.5) and 73.9% (*SD* = 21.2) for the syllogistic reasoning task. This indicates that the type of change is highly stable at the individual level. If people show a specific direction of change pattern on one problem, they tend to show it on all problems. Note that the stability index analysis also argues further against the guessing account. If people were guessing randomly, they should not tend to pick the same response consistently.

3.2.2. Confidence ratings and response time analysis

Examining response latencies and confidence for the four different types of direction of change categories can help to get some further insight in the reasons behind people's answer change (or lack thereof). We focus our analysis here on the critical conflict items (the contrast with no-conflict items can be found in the next section). Results are presented in Figs. 2 and 3. These figures present a rich data set. We are looking at results from two different reasoning tasks (base-rate; syllogisms), four experiments (Experiments 1–4), two response stages (initial and final response), and two dependent measures (confidence and latencies). However, as with the direction of change analysis above, by and large the findings are fairly consistent across experiments and tasks. For ease of interpretation, Fig. 4 presents the findings averaged across experiments and tasks. This will allow us to identify and discuss the main trends first. Subsequently, we will present more detailed statistical tests of the findings.

As Fig. 4 (top panels) indicates, the key pattern with respect to the confidence ratings is that the 00 and 11 direction of change categories show very high confidence both for the initial and final response. Confidence for the 01 and 10 cases – in which participants changed their initial response - is considerably lower at both response stages. Hence, cases in which participants change their initial response tend to be characterized by a lower response confidence. The latency findings for the final response (bottom panel D) further indicate that the lower confidence for the 01 and 10 cases is accompanied by elevated final response times. Participants take more time to give their final answer in the 01 and 11 cases than in the 00 and 11 ones. In other words, the few cases in which people do change their initial response are characterized by a longer rethinking in the final response stage. Note that this pattern is consistent with Thompson et al.'s (2011) original two-response findings and supports the DI dual process prediction that answer change results from time-consuming Type 2 thinking.

With respect to the initial response times Fig. 4 (bottom panel C) indicates that all initial responses are given fairly fast. The vertical line in Fig. 4 (and 2) denotes the 3 s response deadline that was set based on our reading pretest. Average response times for each of the four direction of change categories are all below this threshold. Obviously, as Fig. 2 (left hand panels) indicates, the initial response times do show more variability across experiments. Not surprisingly, in the two experiments in which a deadline was set (and responses above 3 s were discarded), average initial response times values are smaller than in the experiments without deadline. However, even without the deadline participants generally tend to give their initial response within reasonable limits. The only apparent deviancy is the 10 case in which the initial response seems to take slightly longer than in the other direction of change categories.

To analyse the results statistically we used the nlme statistical package in R to create mixed effect multi-level models (Pinheiro,

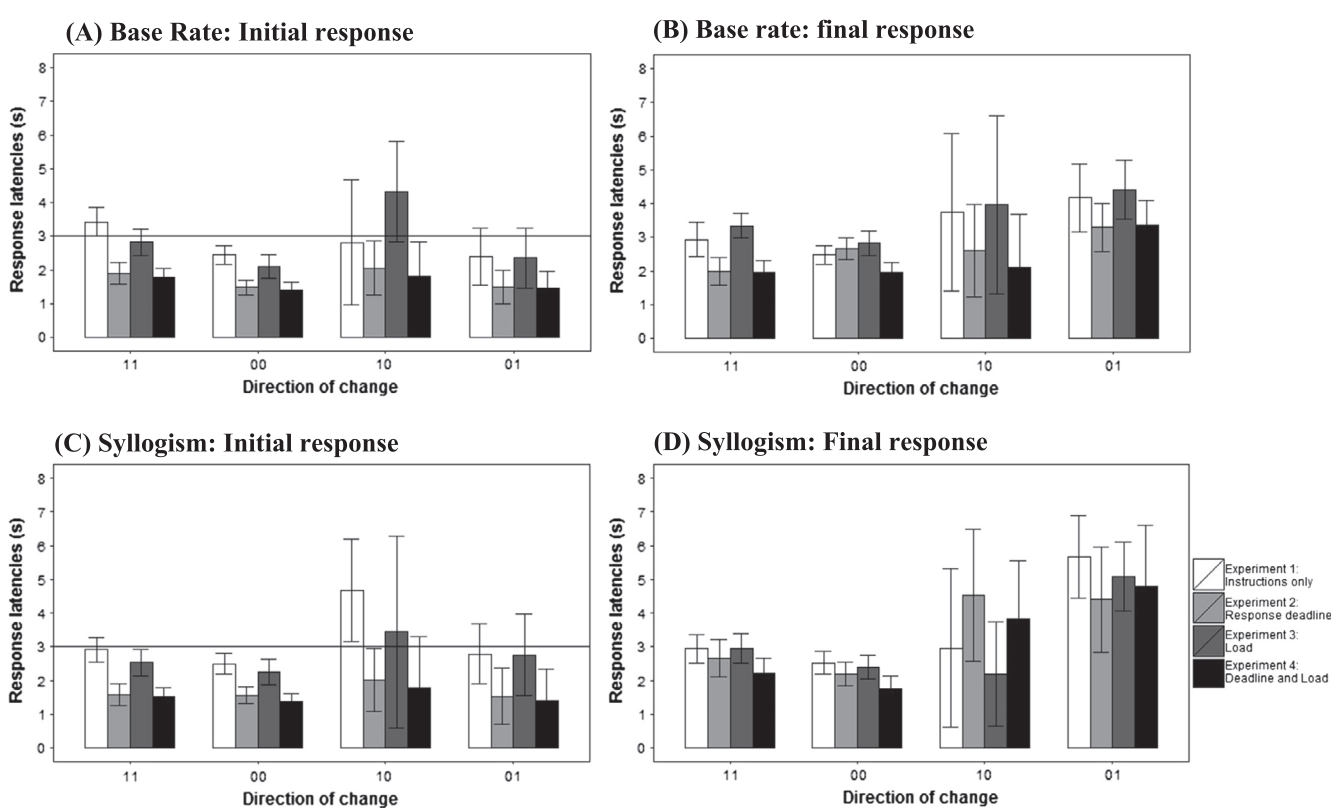

Fig. 2. Mean conflict problems response latencies of the initial and final responses in the base-rate and syllogistic reasoning task for each of the direction of change categories. Error bars are 95% confidence intervals. Note that averages and confidence intervals were calculated on log-transformed latencies. The figure shows the back-transformed (anti-logged) latencies.

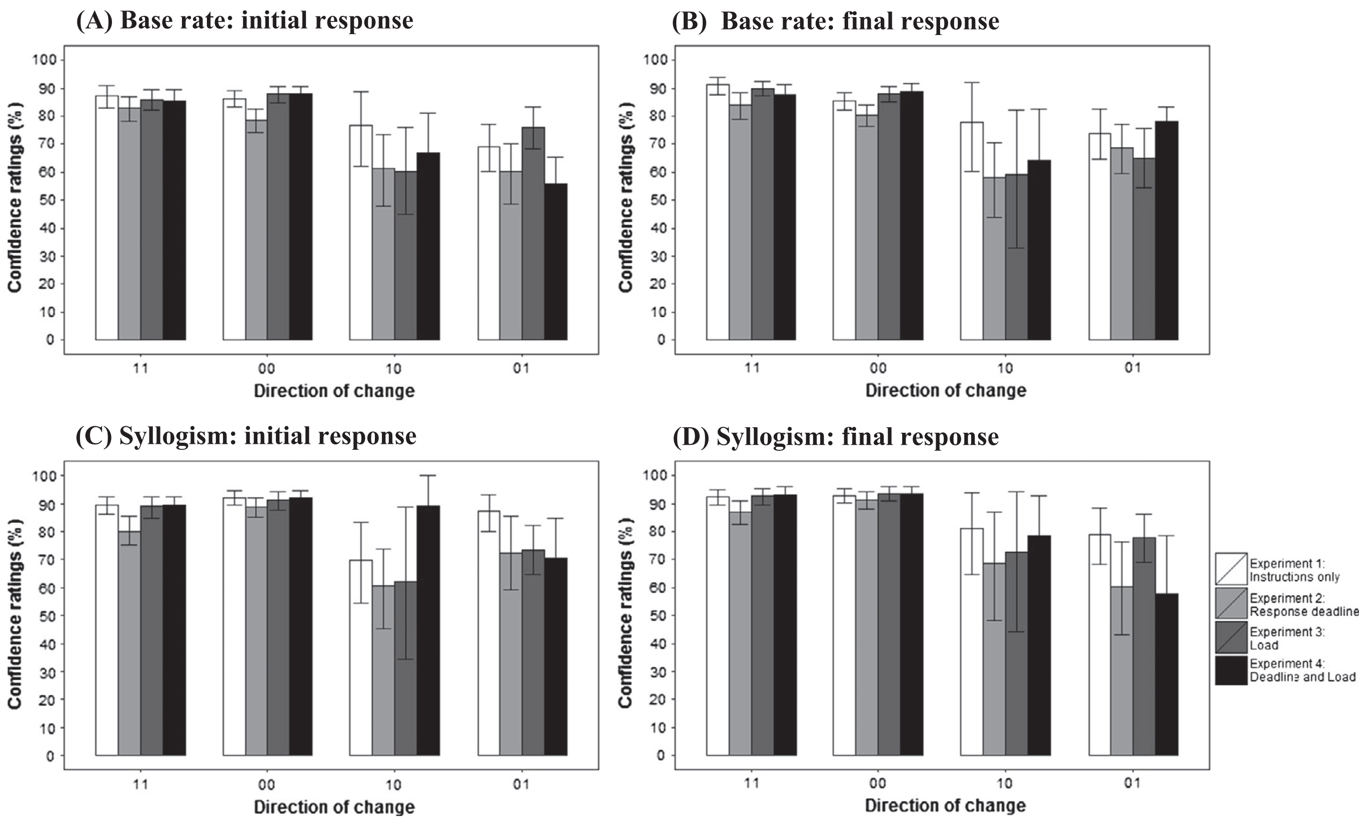

Fig. 3. Mean conflict problem confidence ratings for initial and final responses in the base-rate and syllogistic reasoning tasks for each of the direction of change categories. Error bars are 95% confidence intervals.

Bates, Debroy, & Sarkar, 2015). This allows us to analyse the data in a trial-by-trial basis, while accounting for the random effect of subjects (Baayen, Davidson, & Bates, 2008). Mixed effect models have increased statistical power due to the inclusion of random effects, and the ability to handle data which violates the assumption of homoscedasticity (Baayen et al., 2008). The direction of change category (11, 00, 10, and 01) and the experimental condition (Experiments 1–4) were entered to the model as fixed effect factors, and participants were entered as a random factor. We ran separate analyses for each of the two reasoning tasks. In the few cases in which we found a significant interaction between direction of change and experimental condition we also analysed each experiment separately.

For all response time analyses reported in the present paper, the response times were transformed logarithmically prior to analysis. Note that given the positively skewed nature of the logarithmically transformed reaction time data we further excluded trials whose log-transformed value was over 1.5 (these trials were over one and the half times of the interquartile range, amounting to a deviation of >3.5 SD from the mean) to get a more normal-like distribution. This implied that 3.3% of trials were excluded from the reaction time analysis (initial and final response combined). In the confidence analyses, trials where participants entered a confidence rating higher than 100 were also excluded. This amounted to 2.7% of trials (initial and final response combined).

3.2.3. Response times – initial response

Results for the initial response time analysis in the base rate task showed that the main effect of direction of change category significantly improved model fit, χ^2 (7) = 36.25, $p < 0.0001$, as well as the main effect of experiment, χ^2 (10) = 116.23, $p < 0.0001$. The interaction factor did not improve model fit significantly, χ^2 (19) = 5.03, $p = 0.83$. Similarly, for the syllogistic reasoning task the main effect of direction of change significantly improved model fit, χ^2 (7) = 31.04, $p < 0.0001$, as did the main effect of experiment, χ^2 (10) = 161.13, $p < 0.0001$. The interaction factor did not improve model fit further, χ^2 (19) = 8, $p = 0.53$.

With respect to the main effect of experiment we ran a follow-up contrast test to verify whether the visually identified trend towards longer initial response latencies in the experiments without response deadline was significant. The contrast test indeed indicated that this was the case for both reasoning tasks (base-rate: $b = 0.21$, t (408) = 11.266, $p < 0.0001$, $r = 0.49$; syllogisms: $b = 0.23$, t (406) = 13.608, $p < 0.0001$, $r = 0.56$). The main effect of direction of change on the other hand, seemed to be driven by overall longer initial response latencies in the 10 case. Follow-up contrast analyses also established that this specific trend was significant for the syllogisms, $b = -0.16$, t (39) = −4.488, $p = 0.0001$, $r = 0.58$, but not for base rate problems: $b = -0.06$, t (41) = −1.966, $p = 0.056$, $r = 0.29$. It is important to stress here that there was no interaction between the direction of change and experiment factors. Hence, the increased initial 10 latencies are observed across experiments. If the longer initial 10 responses were only observed in the instruction only experiment, for example, they could have been attributed to Type 2 thinking. Participants would not respect the instruction to respond intuitively, take extra time to deliberate and consequently manage to give the correct response. However, the fact that the longer initial 10 response times are also observed under time pressure and load argues against this explanation. This tentatively suggests that the Type 1 processing in the 10 case (i.e., selection of a correct initial response that is afterwards changed), might be genuinely slower than the Type 1 processing that is occurring in the other direction of change categories.

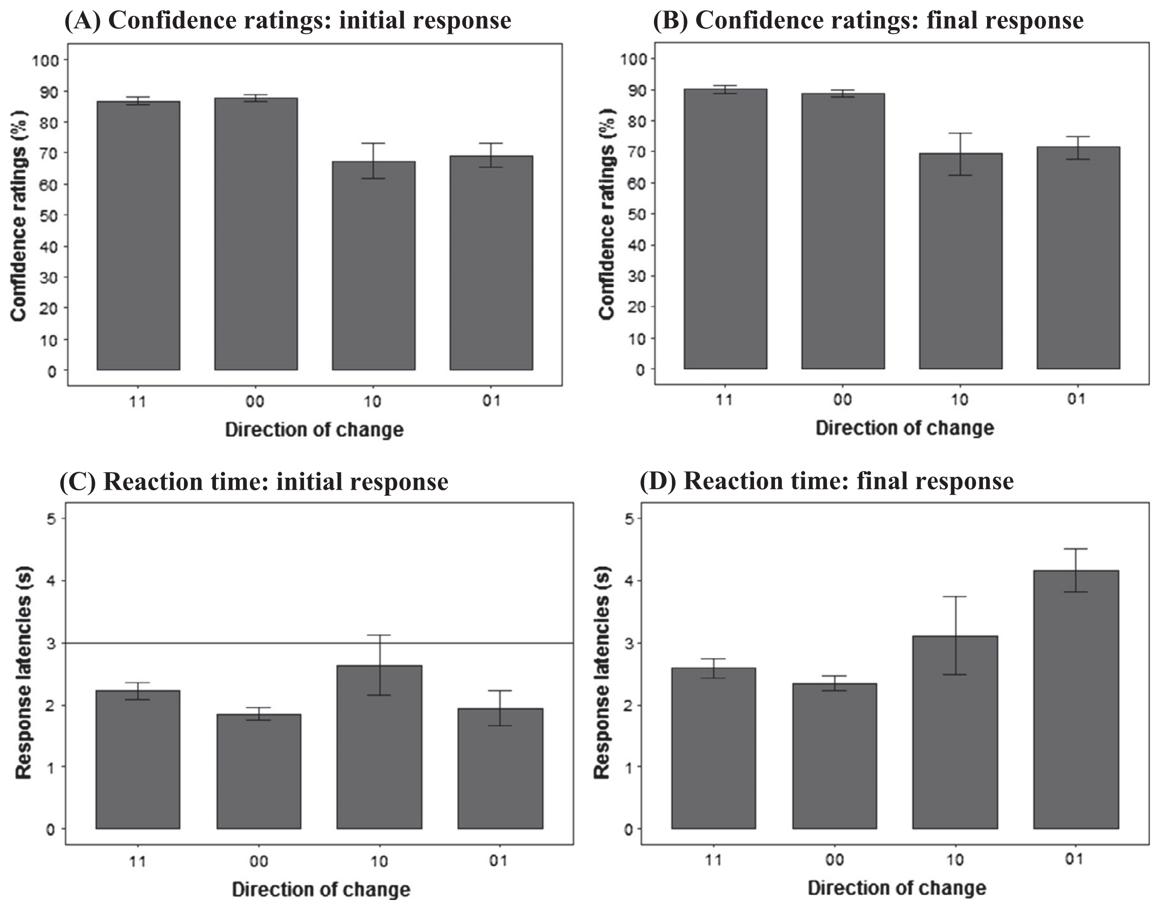

Fig. 4. Mean initial and final conflict problem response latencies and confidence ratings averaged across reasoning tasks and experiments. Error bars are 95% confidence intervals. The figure shows the back-transformed (anti-logged) latencies.

3.2.4. Response time – final response

For the base-rate problems results of the final response time analysis showed that there was a significant main effect of direction of change, χ^2 (7) = 26.37, $p < 0.0001$, and experimental condition, χ^2 (10) = 31.45, $p < 0.0001$. The interaction between both factors did not improve model fit, χ^2 (19) = 8.71, $p = 0.46$. The same pattern was observed for the syllogistic reasoning task; there was a main effect of direction of change, χ^2 (7) = 80.12, $p < 0.0001$, and experimental condition, χ^2 (10) = 14.8, $p = 0.002$, and the interaction did not improve model fit, χ^2 (19) = 11.88, $p = 0.22$.

Follow-up tests for the main effect of direction of change indicated that, as the visual inspection indicated, final response times were longer in the cases where participants changed their initial response (10 and 01) than in cases (11 and 00) where the initial response was not changed (base rate task, $b = -0.1$, t (115) = −4.91, $p < 0.0001$, $r = 0.42$; syllogistic reasoning, $b = -0.24$, t (98) = −8.059, $p < 0.0001$, $r = 0.63$). For completeness, we also note that an exploratory follow-up test for the main effect of experiment indicated that somewhat surprisingly, the final response times in the two conditions that set an initial deadline (Experiment 2: deadline & Experiment 4: deadline + load) also tended to be slightly faster than in the conditions without initial deadline (base rate task, $b = 0.11$, t (408) = 4.659, $p < 0.0001$, $r = 0.22$; syllogistic reasoning, $b = 0.08$, t (405) = 2.953, $p = 0.0033$, $r = 0.15$). Hence, although there was never a deadline for the second response, the fact that participants had previously been faced with one tended to make them speed-up for the final response too.

3.2.5. Confidence – initial response

The confidence analysis for the initial response for the syllogistic reasoning task showed that the direction of change factor improved model fit significantly, χ^2 (7) = 79.88, $p < 0.0001$, but the experiment factor, χ^2 (10) = 6.69, $p = 0.08$, and interaction did not, χ^2 (19) = 16.33, $p = 0.06$. Follow-up contrast tests for the main effect of direction of change indicated that the visually identified trend towards lower initial confidence ratings for the two change categories (10 and 01) was also significant, $b = -15.42$, t (110) = −7.449, $p < 0.0001$, $r = 0.58$. For the base-rate problems we found a main effect of direction of change category, χ^2 (7) = 41.7, $p < 0.0001$, as well as a main effect of experimental condition, χ^2 (10) = 9.47, $p = 0.024$, and a significant interaction, χ^2 (19) = 17.39, $p = 0.043$. With respect to the main effect of direction of change, follow-up tests established that as in the syllogistic reasoning task, initial confidence in the two categories in which the initial response was changed, was lower than in the no-change categories, b (116) = −14.96, t (116) = −7.14, $p < 0.0001$, $r = 0.55$. Because the experiment and direction of change factors interacted we also tested whether this effect was present in all experiments. Results showed that this was the case for Experiment 1 (instructions only), $b = -10.94$, t (23) = −2.787, $p = 0.01$, $r = 0.5$, Experiment 2 (response deadline), $b = -13.14$, t (32) = −3.158, $p = 0.004$, $r = 0.49$, and Experiment 4 (deadline and load), $b = -26.79$, t (30) = −6.224, $p < 0.0001$, $r = 0.75$. However, the trend towards lower initial response confidence in the change vs. no-change categories was less pronounced and failed to reach significance in Experiment 3 (load), $b = -5$, t (28) = −1.427, $p = 0.16$, $r = 0.26$.

3.2.6. Confidence – final response

The confidence analysis for the final response in the syllogistic reasoning task showed that the direction of change effect was significant, χ^2 (7) = 100.58, $p < 0.0001$, but neither the main effect of experiment, χ^2 (10) = 5.4, $p = 0.145$, nor the interaction, χ^2 (19) = 15.95, $p = 0.068$, improved model fit significantly. A follow-up contrast test for the main effect of direction of change indicated that the visually identified trend towards lower final confidence rating for the two change categories (10 and 01) was significant, $b = -20.26$, t (108) = −9.523, $p < 0.0001$, $r = 0.68$. In the base-rate task both the main effect of direction of change, χ^2 (7) = 49.79, $p < 0.0001$, and the main effect of experiment were significant, χ^2 (10) = 10.84, $p = 0.013$. The interaction did not improve model fit, χ^2 (19) = 7.28, $p = 0.608$. With respect to the main effect of direction of change, the follow-up test established that as in the syllogistic reasoning task and initial confidence analysis, final confidence in the 10 and 01 change categories was lower than the final confidence in the 00 and 11 categories, $b = -15.34$, t (117) = −7.522, $p < 0.0001$, $r = 0.57$.

In sum, taken together, these analyses support the major trends that were visually identified. Both in the syllogistic reasoning and base-rate tasks we consistently observe across our experiments that answer change is associated with lowered response confidence and longer final rethinking times.

3.3. Conflict detection analysis

Our key observation so far has been that in the cases where people end up giving a correct final response, they already selected this response as their initial, intuitive response. This suggests that correct logical responses can be generated by fast and automatic Type 1 processes. In this final section, we want to examine whether reasoners are faced with two competing intuitions at the first response stage. That is, one reason for why people in the 11 category manage to give a correct initial response might be that the problem simply does not generate an intuitive heuristic response for them. Hence, they would only generate a correct, logical intuition and would not be faced with an interfering heuristic one. Likewise, one might question whether Type 1 processes for reasoners in the 00 direction of change category also generate a logical intuition in addition to the heuristic intuition that led them to select the incorrect response. In other words, so far our findings indicate that there are conflict trials on which some people generate correct, logical intuitions and there are conflict trials on which some people generate an incorrect, heuristic intuition. What we want to test here is whether both intuitions are also generated concurrently within the same trial.

We can address this question by looking at the contrast between conflict and no-conflict problems. If conflict problems cue two conflicting initial intuitive responses, people should process the problems differently than the no-conflict problems (in which such conflict is absent) in the initial response stage. As we noted in the introduction, conflict detection studies that used a classic single response paradigm have shown that processing conflict problems typically results in lower confidence and longer response latencies, for example. Interestingly, Thompson et al. (2011) found that the lowered confidence ratings for conflict vs. no-conflict were also observed for the initial response. Note that Thompson et al. did not find an impact on response latencies but this might be accounted for by the design characteristics of the two response paradigm (i.e., forcing people to give an explicit response as fast as possible might prevent the slowing effect from showing up). Nevertheless, Thompson et al.'s confidence findings suggest that - averaged over possible change types - there is some evidence for the hypothesis that reasoners are faced with conflicting intuitions when giving their initial responses. The critical question that we want to answer here is whether this is the case for each of the four direction of change categories. Therefore, we contrasted the confidence ratings for the first response on the conflict problems for each direction of change category with the first response confidence on the no-conflict problems. Note that we used only the dominant no-conflict 11 category for this contrast (which we will refer to as "baseline"), as responses in the other no-conflict direction of change categories cannot be interpreted unequivocally.

Fig. 5 shows the results. Visual inspection of Fig. 5 indicates that there is a general trend across tasks and experiments towards a decreased initial confidence when solving conflict problems for all direction of change categories. However, this effect is much larger for the 01 and 10 cases in which reasoners subsequently changed their initial response. This suggests that although reasoners might be experiencing some conflict between competing intuitions in all cases, this conflict is much more pronounced in the 10 and 01 case. For completeness, Fig. 6 also presents an overview of the conflict vs. no conflict contrast response time findings. As the figure indicates, the data were noisier here and there is no clearly consistent pattern that seems stable across tasks and experiments. To avoid spurious conclusions we refrained from analysing these response time data further (Simmons, Nelson, & Simonsohn, 2011).

To analyse the confidence results statistically we again created mixed effect multi-level models (Pinheiro et al., 2015). We ran a separate analysis for each of the four direction of change conflict problem categories in each of the two reasoning tasks. In the analysis the confidence for the first response in the direction of change category in question was contrasted with the first response confidence for 11 no-conflict problems which served as our baseline. We will refer to this contrast as the Conflict factor. The conflict factor was entered as fixed factor, and participants were entered as random factor. We also entered the experimental condition (Experiments 1–4) as fixed factor in the model to see whether it interacted with the conflict factor and the findings were stable across our experiments. As before, in the cases in which we found a significant interaction we also analysed each experiment separately.

3.3.1. 11 Category

Results for the 11 category indicated that for the base rate problems, the main effect of conflict was significant, χ^2 (5) = 41.94, $p < 0.0001$. There was no significant effect of experiment, χ^2 (8) = 7.24, $p = 0.065$, and interaction, χ^2 (11) = 1.97, $p = 0.58$. Hence, people were less confident in the 11 category, $b = -6.567$, t (179) = −6.789, $p < 0.0001$, $r = 0.45$, than in the baseline condition. Similar results were found with regard to the syllogistic reasoning problems, where we observed a significant main effect of conflict, χ^2 (5) = 32.43, $p < 0.0001$, whereas the main effect of condition, χ^2 (8) = 3.96, $p = 0.2663$, and interaction, χ^2 (11) = 7.34, $p = 0.062$ were not significant. People were less confident in the 11 category, $b = -4.804$, t (285) = −5.825, $p < 0.0001$, $r = 0.33$, compared to the baseline condition.

3.3.2. 00 Category

In the base rate task, the effect of conflict was significant, χ^2 (5) = 31.17, $p < 0.0001$, as well as the effect of experiment, χ^2 (8) = 11.74, $p = 0.008$, but not the interaction, χ^2 (11) = 6.12, $p = 0.11$. People were less confident in the 00 responses, b (279) = −5.598, t (279) = −5.657, $p < 0.0001$, $r = 0.32$, than the baseline. In the syllogistic reasoning problems the effect of conflict did not reach significance, χ^2 (5) = 3.35, $p = 0.067$, although there was a trend in the expected direction. The effect of condition, χ^2 (8) = 2.53, $p = 0.47$, and the interaction were not significant, χ^2 (11) = 2.67, $p = 0.45$.

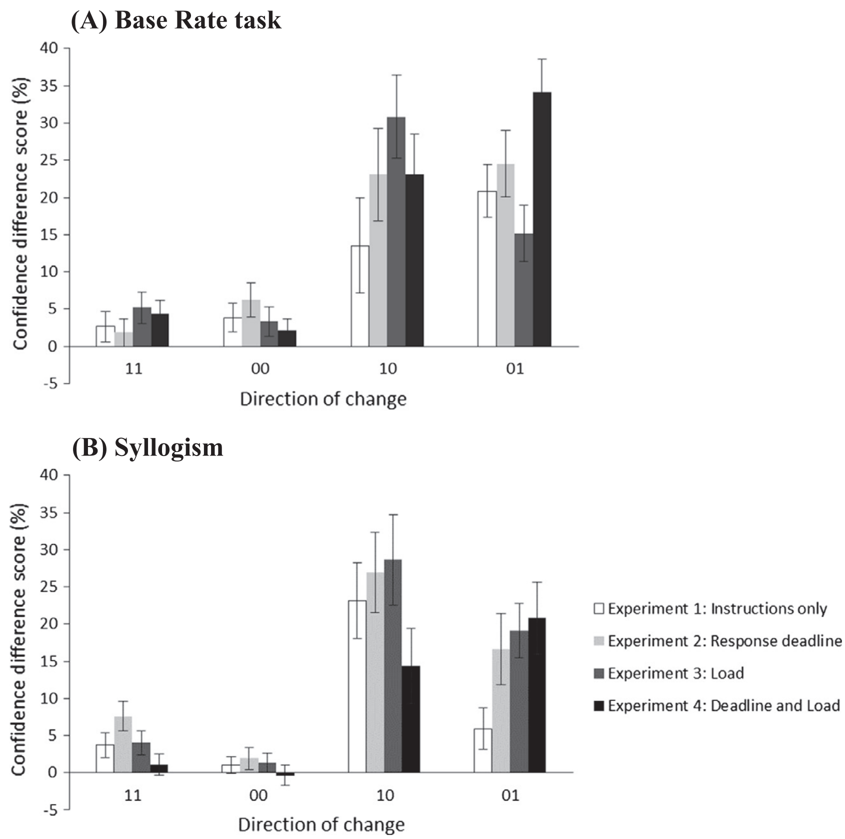

Fig. 5. Confidence rating differences between each direction of change category and the baseline (11 no-conflict) for the initial responses in the base rate (A) and syllogistic reasoning task (B). Positive values mean that people were less confident in a given direction of change category than in the baseline. Error bars are standard errors of the corresponding t-values plotted here for illustrative purposes.

3.3.3. 10 Category

For the base rate problems we found a significant main effect of conflict, χ^2 (5) = 42.85, $p < 0.0001$. The experiment factor, χ^2 (8) = 7.72, $p = 0.052$, and interaction did not reach significance, χ^2 (11) = 1.48, $p = 0.686$. The 10 answers yielded a lower confidence level, $b = -20.58$, t (37) = -6.667, $p < 0.0001$, $r = 0.74$, than the baseline condition. Similar results were found for the syllogistic reasoning problems; there was a significant effect of conflict, χ^2 (5) = 72.81, $p < 0.0001$, but no significant effect of condition, χ^2 (8) = 2.3, $p = 0.51$, or interaction, χ^2 (11) = 5.72, $p = 0.13$. The initial confidence in the 10 category was lower than in the baseline condition, $b = -24.54$, t (44) = -8.834, $p < 0.0001$, $r = 0.8$.

3.3.4. 01 Category

In the base rate task, conflict was found to be significant, χ^2 (5) = 93.62, $p < 0.0001$, as well as condition, χ^2 (8) = 10.65, $p = 0.01$, and the interaction factor, χ^2 (11) = 13.26, $p = 0.0041$. With regard to syllogistic reasoning problems, we found that conflict, χ^2 (5) = 46.82, $p < 0.0001$, and the interaction, χ^2 (11) = 7.84, $p = 0.049$ improved model fit significantly, but not the main effect of experimental condition, χ^2 (8) = 4.36, $p = 0.23$. Because of the interactions we also ran the conflict contrast for each of the experiments separately. As Fig. 5 indicates, although the confidence decrease was especially pronounced in Experiment 4 (deadline + load) for the base-rate task, our analyses indicated that it reached significance in each of the four experiments (in every condition $r > 0.57$, and $p < 0.01$). A similar pattern was observed in the syllogistic reasoning task. Although the conflict factor failed to reach significance in the instructions only condition, $b = -3.8$, t (18) = -1.364, $p = 0.19$, $r = 0.31$, the effect was present in every other category (all $r > 0.61$, and $p < 0.01$).

Taken together, the conflict detection analysis on the confidence ratings for the first, intuitive answer indicates that by and large participants showed decreased response confidence (in contrast with the no-conflict baseline) after having given their first, intuitive response on the conflict problems in all direction of change categories. This supports the hypothesis that participants were always being faced with two conflicting responses when solving the conflict problems. In other words, results imply that 11 responders also activate a heuristic intuition in addition to the logical response they selected. Likewise, 00 responders also activate a logical intuition despite their selection of the incorrect, heuristic response. But visual inspection also clearly shows that the decreased confidence effect is much larger for the 10 and 01 cases than for the 11 and 00 ones. A contrast analysis[6] that tested this trend directly indicated that it was indeed significant, both for the base rate, $Z = -16.57$, $p < 0.0001$ ($r = 0.33$ for the no-change group, while $r = 0.71$ for the change group), and syllogistic reasoning problems, $Z = -17.49$, $p < 0.0001$, ($r = 0.23$ for no-change and $r = 0.69$ for change group). This indicates that although reasoners might be generating two intuitive responses and are being affected by conflict between them in all cases, this conflict is much more pronounced in cases where people subsequently change their answer. This tentatively suggests that it is this more pronounced conflict experience that makes them subsequently change their answer. As we will explain in the General discussion section, we believe that this more

[6] For this contrast analysis, we first calculated the r effect sizes in the same way we did in previous sections. As a next step we used Fisher r-to-z transformation to assess the statistical difference between the two independent r-values. We used the following calculator for the transformation and p-value calculation: http://vassarstats.net/rdiff.html.

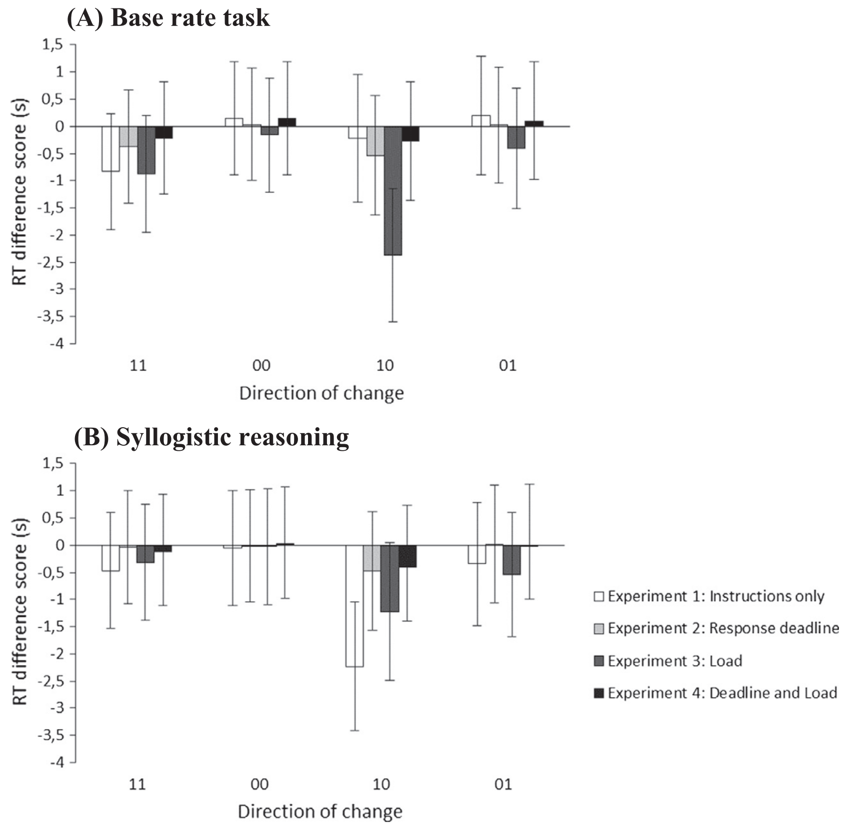

Fig. 6. Response latency differences between each direction of change category and the baseline (11 no-conflict) for the initial response in the base rate (A) and syllogistic reasoning (B) task. Negative values mean that responses took participants longer in a given direction of change category than in the baseline. Error bars are standard errors of the corresponding t-values plotted here for illustrative purposes. The figure shows the back-transformed (anti-logged) latencies.

pronounced conflict in the change categories points to relative differences in the strength of the logical and intuitive intuition in the different answer categories.

4. General discussion

In this study we aimed to examine the time course assumption of classic default-interventionist dual process theory (DI). DI theory suggests that reasoners typically produce a fast, intuitive (Type 1) response by default, and that subsequently this response might be overridden by further, more deliberative processes (Type 2). The quick, initial Type 1 answer is believed to be driven by heuristics based on stereotypes or common beliefs, thus in "conflict" situations (where the stereotype or common belief cues a response that differs from what the logical answer would be) this initial intuition is producing an erroneous answer. Type 2 thinking is expected to override this response, but sometimes Type 2 processing fails too, which results in biased reasoning. Hence, in principle correct responding for conflict items must originate from slow, deliberative Type 2 processing according to DI theory. In the present study we used the two response paradigm to test this assumption. In this paradigm people are instructed to give a very quick initial response and are afterwards allotted as much time as they want to indicate their final response. We also used different research designs to minimize the possibility that participants engaged in Type 2 processing when giving their initial response so as to make sure that the initial answer provided by the participant was really intuitive in nature.

Our analyses focused on four possible direction of change categories: initial response correct, but final response incorrect (10), initial response incorrect and final response correct (01), both responses correct (11) and both incorrect (00). DI theory predicts that reasoners either give a 00 response when they cannot override their first, erroneous heuristic answer, or a 01 response, when Type 2 processing overrides and corrects the initial incorrect response. In line with this hypothesis, we got a high prevalence of 00 responses (about 50% across our studies and reasoning tasks) which basically means that people were typically biased and failed to answer the problems correctly. Less frequently - in about 10% of the cases - we also observed responses in the 01 category. This suggests that correction of an initial erroneous response by Type 2 processing is rare which is also in line with DI theory. However, contrary to DI predictions, our key finding was a relatively high prevalence (+30% throughout) of 11 answers, which suggests that people were giving the correct response intuitively.

Confidence and latency analyses indicated that both in the initial and final response stage answers in the 11 and 00 categories were given quickly and with high confidence. For the rare 01 and 10 responses in which reasoners changed their initial answer, we observed - in line with previous observations (e.g., Thompson & Johnson, 2014; Thompson et al., 2011) lower confidence ratings and longer final response times.

As a final step, we examined whether people were facing two competing intuitions at the initial response stage. Therefore, we contrasted the initial confidence levels and latencies of conflict and no-conflict problems. Initial response latencies did not differ but the confidence ratings did indicate that participants were generally experiencing some conflict in every direction of change category: reasoners were less confident in the correctness of their answer in conflict than in no-conflict trials. This suggests that people typically generated both a logical and heuristic intuition when faced with the conflict problems. If reasoners generated only one type of intuition any intrinsic conflict should obviously not impact their processing. However, the size of the experienced conflict (i.e., the difference between confidence levels of conflict and no-conflict problems) was quite different across categories. In categories where people changed their answer (10 and 01), people experienced more conflict than in the 11 and 00 cases where they did not change their initial answer. Hence, although reasoners in all direction of change categories might be experiencing conflict between competing intuitions, this conflict seems much more pronounced in the cases in which an initial answer is changed. In line with some recent suggestions this might indicate that one factor that determines whether or not a first intuitive answer is changed, is the level of experienced conflict (e.g., Thompson & Johnson, 2014).

With few exceptions our findings were consistent across the two reasoning tasks and four different studies that we ran. This stability indicates that the findings are quite robust and minimize the possibility that they result from any idiosyncratic task or experimental feature. Nevertheless, the study only started to adopt the current direction of change analysis and experimental control. It will obviously be important to generalize and validate the findings further in future studies. With this consideration in mind the findings do indicate that there is substantial ground to question the traditional DI theory time-course characterization. But it is important to avoid confusion with respect to our precise claim. Our core finding is that people are able to give the logical answer to conflict problems intuitively. It is this phenomenon that we are referring to as *fast* or *Type 1* logic to contrast it with slow and demanding logical reasoning based on Type 2 processing. However, it should be stressed that although the outcome of the two types of logical responding might be the same (i.e., selection of the correct response) this does obviously not entail that the underlying processing is also similar. That is, we do not claim that people are going through the exact same complex calculations as they would in the Type 2 case but simply perform these faster in the Type 1 case. Clearly, the point is that both types of logical responding are based on different types of processes (i.e., Type 1 and Type 2) and will consequently have different characteristics. For example, we believe it is quite unlikely that people will manage to justify their initial logical response and explain why it is correct without engaging in additional Type 2 processing (e.g., De Neys & Glumicic, 2008; Mercier & Sperber, 2011; Trouche, Sander, & Mercier, 2014). To further clarify this point let us draw an analogy between our account and the recall and recognition distinction in memory research (e.g., Haist, Shimamura, & Squire, 1992). Imagine you are given a list of ten names to study and your memorization performance is being tested. Recall memory will allow you to explicitly retrieve (some of) the names on the list (e.g., you might manage to jot down that "Dave", "Tom", and "Lauren" were among the presented names). Recognition memory will allow you to merely decide whether or not a certain item was on the list (e.g., you might manage to say "yes" when asked whether "Dave" was among the presented names or not). Sometimes you might not be able to recall the name, but you could still recognize whether you saw it on the list or not. Recall and recognition can both allow us to retrieve a memory trace but they differ (among other things) in the processes involved in memory retrieval (e.g., Anderson & Bower, 1972; Ben-Yakov, Dudai, & Mayford, 2015; Buratti & Allwood, 2012). This recall/recognition dichotomy is akin to what we are alluding to here. In our view, fast Type 1 logical responding can be conceived as a more superficial, recognition memory-like process that activates a stored logical principle and allows us to recognize that a competing heuristic intuition is questionable, without us being able to explicitly label or justify the principle. Although the present study does not allow us to pinpoint how the Type 1 and Type 2 logical responses differ precisely, the key point we want to make is that our theorizing does not entail that fast, Type 1 logical responses are similar – let alone superior – to Type 2 logical responses.

Note that the differential nature of Type 1 and Type 2 logical responding that we are alluding to here might receive some support from the recent work of Trippas et al. (2016). Inspired by initial studies of Morsanyi and Handley (2012a, 2012b), Trippas et al. presented participants with logical arguments and simply asked them to indicate how much they liked the conclusion or how bright (i.e., "luminous") they judged the conclusion to be. Although these tasks made no reference whatsoever to logical reasoning, the authors observed that people gave higher liking and brightness ratings to logically valid than to logically invalid conclusions. As the authors noted, these findings (i.e., sensitivity to logical structure outside of an explicit reasoning context) lend credence to the idea that logical structure might be processed automatically and intuitively. But more critically, Trippas et al. explained (and predicted) the results within a fluency misattribution framework (e.g., Topolinski & Strack, 2009). The rationale is that if logical structure is processed intuitively, valid conclusions will give rise to feelings of conceptual fluency. However, because of the implicit nature of the process people will have no explicit insight into the nature of this feeling. As Trippas et al. argue, the enhanced processing fluency of logically valid conclusions will consequently be (mis)attributed to positive affect and will lead to the judgment that the conclusion is brighter and more likeable (see also Thompson & Morsanyi, 2012, for a related suggestion). This fluency account might help to characterize the precise origin of Type 1 logical sensitivity and underscores the point that fast (Type 1) and slow (Type 2) logical responses likely result from a qualitatively different type of processing.

Before discussing more theoretical implications of our findings we would like to highlight some important methodological and practical considerations. One issue concerns the validity of the two-response paradigm. As we noted, a potential critique of previous studies that adopted a two-response paradigm is that we cannot be certain that participants respect the instructions to give the first intuitive response and did not engage in Type 2 processing during the initial response stage. In the present study we used a combination of methods (e.g., response deadline, cognitive load) to minimize the amount of Type 2 thinking at the initial answer stage. By and large, we found that none of these experimental manipulations critically affected the results. This implies that participants in the standard/instruction-only paradigm are in fact doing what they are instructed to do and refrain from engaging in Type 2 processing during the initial response stage. In this sense, the present study presents a methodological validation of the standard two-response paradigm that relies on instructions only. When adopting a two-response paradigm it is reasonable for scholars to assume that participants will stick to purely intuitive responding in the initial response phase (see also Thompson et al., 2011, for a further discussion of the validity of the two-response paradigm).

Another methodological point is that from the outset our task design was optimized to identify potential Type 1 logical responding in case it existed. For example, we used a fast response version of the base-rate task (Pennycook, Cheyne, et al., 2014) that did not require reading through a lengthy description and minimized response time variance. In addition, it is also the case that both of our reasoning tasks used a simplified binary-response format in which participants selected one of two presented response options by clicking on them. Consequently, participants did not have to explicitly generate the conclusion themselves when giving their initial response. One might wonder whether these adaptations or simplifications invalidate the results. For example, a critic might object that with a harder task version in which reasoners had to generate their own conclusions, there would be no evidence for fast logical responding. There are a number of points to make here. First, although our tasks were modified, the dominant response category was still of the 00 type. In the majority of cases, participants failed to solve the problems correctly even after they were allowed additional processing time in the final response stage. If our tasks would have been too easy, our sample of educated adults should obviously not have struggled to solve them correctly. Second, if we were to ask participants to type down or verbalize a conclusion, the typing or verbalization itself might need controlled processing and prevent a proper measurement of pure, intuitive processing. By definition, if we want to study intuitive processing, we need to use the proper methodological tools to measure it. Third, it is not necessary, nor claimed that Type 1 and Type 2 logical responding have the same characteristics. We concur, for example, that it is quite likely that the explicit conclusion generation in syllogistic reasoning cannot be achieved by Type 1 processes. But the point that there are differences between a fast/intuitive and slower/deliberative type of logical responding should not be held against the postulation of the existence of logical intuitions per se. This would be as nonsensical as arguing against the postulation of recognition memory because the recognized memory item cannot be explicitly recalled.

From a more theoretical point of view, it will be clear that it is hard for the standard DI theory to explain the current findings. But if the DI model is not adequate, are there possible alternative conceptualisations that allow us to make theoretical sense of the results? One might be tempted here to consider so-called parallel dual process models (e.g., Epstein, 1994; Sloman, 1996). These parallel models are a classic alternative to the popular DI model. The DI model posits that Type 1 and Type 2 processes interact in a serial fashion. Reasoners initially start with Type 1 processing and Type 2 processing is only engaged in a later stage. In the parallel model both Type 1 and Type 2 processing are engaged simultaneously from the start. Hence, Type 1 and Type 2 processing operate in parallel rather than serially. One might wonder whether such parallel Type 2 processing might explain correct immediate responses. Note, however, that - as all dual process models – the parallel model still defines Type 2 processing as being slow and demanding of cognitive resources (Epstein, 1994; Sloman, 1996). Now, our key observation was not that people generated correct responses, but that they did so intuitively. Correct initial responding was observed even when Type 2 processing was knocked out by a challenging deadline and concurrent load task. Hence, even if the parallel's model assumption with respect to the simultaneous activation of Type 1 and Type 2 processing were to be correct, it cannot explain the occurrence of fast and intuitive logical responding in the present study. Moreover, ideally we do not only need to account for the occurrence of initial correct responses, we also need to explain the direction of change results. That is, how can we explain why one reasoner is ending up in the 00 category and another one in the 11 or 01 category, for example?

We believe that a more promising explanation can be offered by recent hybrid dual process models of reasoning (De Neys, 2012; Handley & Trippas, 2015; Pennycook et al., 2015; see also Macchi & Bagassi, 2015, for a related view). Hybrid models assume that more than one Type 1 answer can be generated as a result of parallel intuitive processes, which might be followed by the more demanding Type 2 processing. Bluntly put, the hybrid model combines key features of the serial and parallel model: Just like the serial model it assumes that Type 2 processing is optional and starts later than Type 1 processing. And just like the parallel model, it assumes that there is parallel logical and heuristic processing. However, unlike the parallel model it is claimed that this logical processing results from Type 1 processing. For example, the hybrid "logical intuition model" of De Neys (2012) suggests that people intuitively detect the conflict between heuristic responses and logical principles. The basic idea is that conflict is caused by two simultaneously activated Type 1 responses, one is cueing the logical response based on stored knowledge of elementary logical principles, another is cueing the heuristic response based on belief-based semantic associations. Critically, De Neys (2012, 2014) indicated that this does not entail that the two Type 1 responses are similar in strength (see also Pennycook et al., 2015; Pennycook, Trippas, et al., 2014). More specifically, the idea is that most people are typically biased when solving traditional reasoning problems, for example, precisely because their heuristic intuition is more salient or stronger (i.e., has a higher activation level) than their logical intuition. Building on this differential strength suggestion can help us to make sense of the direction of change findings and explain why one ends up in a specific change category.

Note that what we refer to here as the differential strength of different intuitions is also a key feature of Pennycook et al.'s (2015) three-stage dual process model. This model proposes that initially multiple Type 1 responses will be cued by a stimulus (Stage 1), leading to the potential for conflict detection between different Type 1 responses (Stage 2). If successful, conflict detection will lead to Type 2 processing (Stage 3). What is critical for our present purposes is that one central feature of the model is that the multiple, potentially competing, Type 1 responses that are initially cued by a problem (e.g., a "logical" and "heuristic" intuition) are envisaged to differ in the ease and speed in which they come to mind. This idea nicely fits with our differential strength suggestion and allow us to make sense of the present findings.

More specifically, what we propose is that we need to consider both absolute (which one of the two intuitions is strongest?) and relative (how pronounced is the activation difference between both intuitions?) strength differences between the logical and heuristic intuition. The initial response will be determined by the absolute strength level. Whichever intuition is strongest will be selected as initial response. Whether or not the initial response gets subsequently changed will be determined by the relative difference between both intuitions. The smaller the difference, the less confident one will be, and the more likely that the initial response will be changed. Fig. 7 illustrates this idea. In the figure we have plotted the strength of the logical and heuristic intuition for each of the four direction of change categories in (imaginary) activation strength "units" for illustrative purposes. For example, in the 11 case, the logical intuition might be 4 units strong whereas the heuristic intuition might be only 1 unit strong. In the 00 case, we would have the opposite situation with a 4 unit strong heuristic intuition and a much weaker, 1 unit logical intuition. In the two change categories, one of the two intuitions will also dominate the other but the relative difference will be less pronounced. For example, in the 01 case the heuristic intuition might have strength level 3 whereas the logical intuition has strength level 2. Because the relative difference is less pronounced, there will be more doubt and this will be associated with longer final rethinking and answer change. In other words, in each of the four direction of change categories there will be differences in which intuition is the dominant one and how dominant the intuition is. The more dominant an intuition is, the more likely that it will be selected as initial response, and the less likely that it will be changed. Obviously, this "activation strength" proposal will need to be further tested but we believe it presents a coherent and parsimonious account to explain direction of change findings and re-conceptualize the time-course assumptions in the classic DI model.

To avoid confusion, it should be stressed that the hybrid model we are advocating does not question that people rely by default on Type 1 processing and switch to Type 2 processing in a later stage.

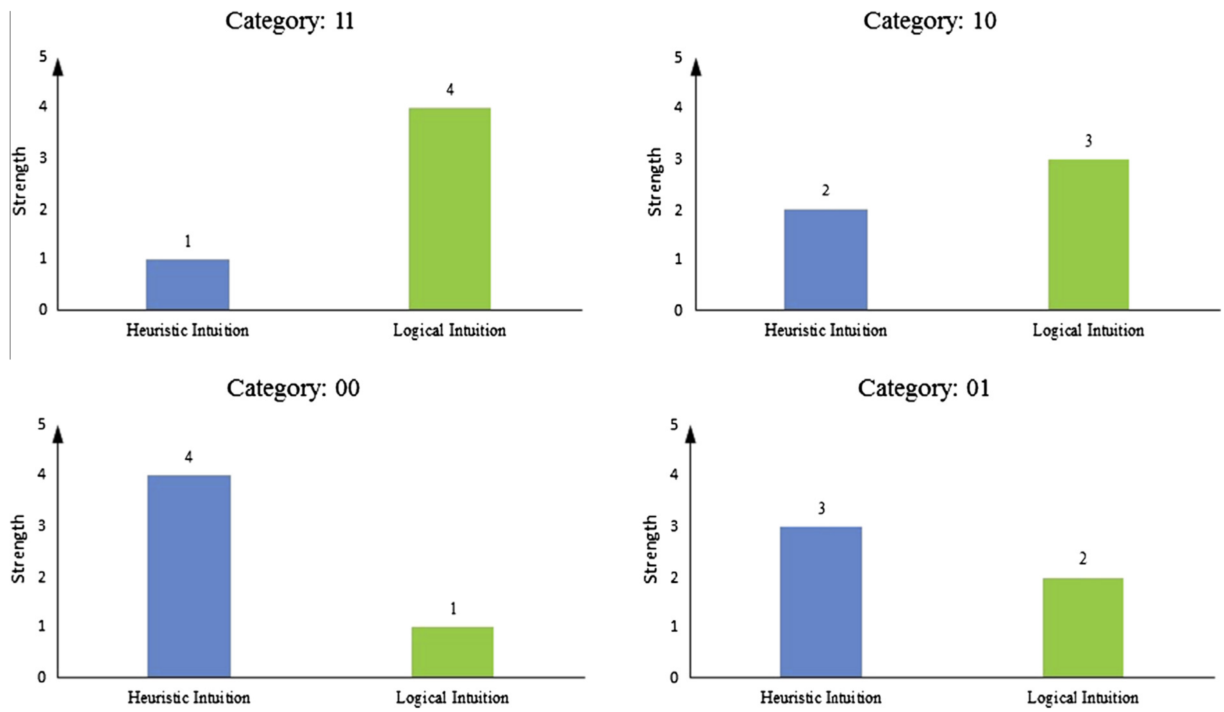

Fig. 7. Illustration of possible absolute (which one of the two intuitions is strongest?) and relative (how pronounced is the activation difference between both intuitions?) strength differences between the logical and heuristic intuition in the different direction of change categories. The figure shows the strength of the logical and heuristic intuition for each of the in (imaginary) activation strength "units" for illustrative purposes.

As we noted, the hybrid model still maintains the DI feature that default Type 1 processing precedes Type 2 processing. The key point is that the default Type 1 activation needs to include some elementary logical processing. If classic DI models allow for the postulation of logical intuitions as characterized here, they are of course fully coherent with the hybrid view (e.g., see De Neys, 2014). As one reviewer noted, at least at a high level of conceptualization classic DI theory might be open towards this possibility. If this is the case, the development or revision of DI theory we call for here would not be inconsistent with the spirit of classic DI theorists' ideas.

To conclude, the present studies indicate that fast and automatic Type 1 processing can cue a correct logical response from the start of the reasoning process. This pattern of results lends credence to a model in which the relative strength of different types of intuitions determines reasoning performance.

Acknowledgments

Bence Bago was supported by a doctoral fellowship from Ecole des Neurosciences de Paris Ile-de-France. Parts of this paper were presented at the 2015 International Convention of Psychological Science (ICPS, Amsterdam), the 2015 Subjective Probability, Utility, & Decision Making Conference (SPUDM24, Budapest), and the 2015 Annual meeting of the Society for Judgment and Decision Making (SJDM, Chicago). We would like to thank the various audiences for valuable feedback and discussion which helped us to improve the present manuscript.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2016.10.014.

References

Aczel, B., Szollosi, A., & Bago, B. (2016). Lax monitoring versus logical intuition: The determinants of confidence in conjunction fallacy. *Thinking & Reasoning, 22*(1), 99–117. http://dx.doi.org/10.1080/13546783.2015.1062801.

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review, 79*(2), 97–123.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412.

Barbey, A. K., & Sloman, S. A. (2007). Base-rate respect: From ecological rationality to dual processes. *Behavioral and Brain Sciences, 30*(3), 241–254.

Ben-Yakov, A., Dudai, Y., & Mayford, M. R. (2015). Memory retrieval in mice and men. *Cold Spring Harbor Perspectives in Biology, 7*(12), a021790.

Bonner, C., & Newell, B. R. (2010). In conflict with ourselves? An investigation of heuristic and analytic processes in decision making. *Memory & Cognition, 38*(2), 186–196.

Buratti, S., & Allwood, C. M. (2012). The accuracy of meta-metacognitive judgments: Regulating the realism of confidence. *Cognitive Processing, 13*(3), 243–253.

De Neys, W. (2006a). Automatic–heuristic and executive–analytic processing during reasoning: Chronometric and dual-task considerations. *The Quarterly Journal of Experimental Psychology, 59*(6), 1070–1100.

De Neys, W. (2006b). Dual processing in reasoning two systems but one reasoner. *Psychological Science, 17*(5), 428–433.

De Neys, W. (2012). Bias and conflict a case for logical intuitions. *Perspectives on Psychological Science, 7*(1), 28–38.

De Neys, W. (2014). Conflict detection, dual processes, and logical intuitions: Some clarifications. *Thinking & Reasoning, 20*(2), 169–187.

De Neys, W. (2015). Heuristic bias and conflict detection during thinking. *Psychology of Learning and Motivation, 62*, 1–32.

De Neys, W., Cromheeke, S., & Osman, M. (2011). Biased but in doubt: Conflict and decision confidence. *PLoS ONE, 6*(1), e15954.

De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition, 106*(3), 1248–1299.

De Neys, W., Moyens, E., & Vansteenwegen, D. (2010). Feeling we're biased: Autonomic arousal and reasoning conflict. *Cognitive, Affective, & Behavioral Neuroscience, 10*(2), 208–216.

De Neys, W., Rossi, S., & Houdé, O. (2013). Bats, balls, and substitution sensitivity: Cognitive misers are no happy fools. *Psychonomic Bulletin & Review, 20*(2), 269–273.

De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load. *Experimental Psychology (Formerly Zeitschrift Für Experimentelle Psychologie), 54*(2), 128–133.

De Neys, W., Vartanian, O., & Goel, V. (2008). Smarter than we think when our brains detect that we are biased. *Psychological Science, 19*(5), 483–489.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist, 49*(8), 709–724.

Evans, J. (2010). *Thinking twice: Two minds in one brain*. Oxford: Oxford University Press.

Evans, J. S. B. (2012). Spot the difference: Distinguishing between two kinds of processing. *Mind & Society, 11*(1), 121–131.

Evans, J. S. B., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning, 11*(4), 382–389.

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition advancing the debate. *Perspectives on Psychological Science, 8*(3), 223–241.

Franssens, S., & De Neys, W. (2009). The effortless nature of conflict detection during thinking. *Thinking & Reasoning, 15*(2), 105–128.

Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives, 19*(4), 25–42.

Gangemi, A., Bourgeois-Gironde, S., & Mancini, F. (2015). Feelings of error in reasoning—in search of a phenomenon. *Thinking & Reasoning, 21*(4), 383–396. http://dx.doi.org/10.1080/13546783.2014.980755.

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*(3), 513–525.

Gilovich, T., Griffin, D. W., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgement*. Cambridge: Cambridge University Press.

Haist, F., Shimamura, A. P., & Squire, L. R. (1992). On the relationship between recall and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(4), 691–702.

Handley, S. J., Newstead, S. E., & Trippas, D. (2011). Logic, beliefs, and instruction: A test of the default interventionist account of belief bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(1), 28–43.

Handley, S. J., & Trippas, D. (2015). Chapter Two-Dual processes and the interplay between knowledge and structure: A new parallel processing model. *Psychology of Learning and Motivation, 62*, 33–58.

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.

Klauer, K. C., & Singmann, H. (2013). Does logic feel good? Testing for intuitive detection of logicality in syllogistic reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(4), 1265–1273.

Kruglanski, A. W. (2013). Only one? The default interventionist perspective as a unimodel—Commentary on Evans & Stanovich (2013). *Perspectives on Psychological Science, 8*(3), 242–247.

Macchi, L., & Bagassi, M. (2015). When analytic thought is challenged by a misunderstanding. *Thinking & Reasoning, 21*(1), 147–164.

Mata, A., Schubert, A.-L., & Ferreira, M. B. (2014). The role of language comprehension in reasoning: How "good-enough" representations induce biases. *Cognition, 133*(2), 457–463.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences, 34*(2), 57–74.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General, 130*(4), 621–640.

Morsanyi, K., & Handley, S. (2012a). Does thinking make you biased? The case of the engineers and lawyer problem. *Proceedings of the annual meeting of the cognitive science society* (Vol. 34, pp. 2049–2054). .

Morsanyi, K., & Handley, S. J. (2012b). Logic feels so good—I like it! Evidence for intuitive detection of logicality in syllogistic reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(3), 596–616.

Newman, I., Gibb, M., & Thompson, V. A. (2016). Rule-based reasoning is fast and belief-based reasoning can be slow: Challenging current explanations of belief - bias and base-rate neglect (submitted for publication).

Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2014). Cognitive style and religiosity: The role of conflict detection. *Memory & Cognition, 42*(1), 1–10.

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2012). Are we good at detecting conflict during reasoning? *Cognition, 124*(1), 101–106.

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). What makes us think? A three-stage dual-process model of analytic engagement. *Cognitive Psychology, 80*, 34–72.

Pennycook, G., & Thompson, V. A. (2012). Reasoning with base rates is routine, relatively effortless, and context dependent. *Psychonomic Bulletin & Review, 19*(3), 528–534.

Pennycook, G., Trippas, D., Handley, S. J., & Thompson, V. A. (2014). Base rates: Both neglected and intuitive. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(2), 544–554.

Pinheiro, J., Bates, D., Debroy, S., & Sarkar, D. (2015). nlme: Linear and nonlinear mixed effects models.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359–1366.

Simon, G., Lubin, A., Houdé, O., & De Neys, W. (2015). Anterior cingulate cortex and intuitive bias detection during number conservation. *Cognitive Neuroscience, 6*(4), 158–168. http://dx.doi.org/10.1080/17588928.2015.1036847.

Singmann, H., Klauer, K. C., & Kellen, D. (2014). Intuitive logic revisited: New data and a Bayesian mixed model meta-analysis. *PLoS ONE, 9*(4), e94223.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*(1), 3–22.

Stanovich, K. (2011). *Rationality and the reflective mind*. Oxford: Oxford University Press.

Stanovich, K. E., & Toplak, M. E. (2012). Defining features versus incidental correlates of Type 1 and Type 2 processing. *Mind & Society, 11*(1), 3–13.

Stanovich, K. E., & West, R. F. (2000). Advancing the rationality debate. *Behavioral and Brain Sciences, 23*(5), 701–717.

Stupple, E. J., Ball, L. J., & Ellis, D. (2013). Matching bias in syllogistic reasoning: Evidence for a dual-process account from response times and confidence ratings. *Thinking & Reasoning, 19*(1), 54–77.

Thompson, V. A., & Johnson, S. C. (2014). Conflict, metacognition, and analytic thinking. *Thinking & Reasoning, 20*(2), 215–244.

Thompson, V. A., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology, 63*(3), 107–140.

Thompson, V., & Morsanyi, K. (2012). Analytic thinking: Do you feel like it? *Mind & Society, 11*(1), 93–105.

Topolinski, S., & Strack, F. (2009). The architecture of intuition: Fluency and affect determine intuitive judgments of semantic and visual coherence and judgments of grammaticality in artificial grammar learning. *Journal of Experimental Psychology: General, 138*(1), 39–63.

Travers, E., Rolison, J. J., & Feeney, A. (2016). The time course of conflict on the Cognitive Reflection Test. *Cognition, 150*, 109–118.

Trippas, D., Handley, S. J., Verde, M. F., & Morsanyi, K. (2016). Logic brightens my day: Evidence for implicit sensitivity to logical validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 1448–1457. http://dx.doi.org/10.1037/xlm0000248.

Trouche, E., Sander, E., & Mercier, H. (2014). Arguments, more than confidence, explain the good performance of reasoning groups. *Journal of Experimental Psychology: General, 143*(5), 1958–1971.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131.

Villejoubert, G. (2009). Are representativeness judgments automatic and rapid? The effect of time pressure on the conjunction fallacy. *Proceedings of the annual meeting of the cognitive science society* (Vol. 30, pp. 2980–2985). Cognitive Science Society.