The influence of conflict monitoring on meta-reasoning and response times in a base rate task

Marin Dujmović*, Pavle Valerjev

Department of Psychology, University of Zadar, Zadar, Croatia

*Corresponding author: Marin Dujmović

Department of Psychology, University of Zadar, Obala kralja Petra Krešimira IV, 2, 23000 Zadar

Email: marin.dujmovic@gmail.com

Phone number: +385919575536

This research was supported by Grant 4139 from the Croatian Science Foundation, and the University of Rijeka Research grant 13.04.1.3.11.

The influence of conflict monitoring on meta-reasoning and response times in a base rate task

Abstract

We examined the role of conflict monitoring processes in forming metacognitive judgments of confidence while performing base rate tasks. Recently proposed models of dual process reasoning, as well as research, have shown conflict detection might represent a link between Type 1 and Type 2 processing. Conflict detection has also been shown to affect metacognitive processes in reasoning tasks. By varying base rate probability and congruence we generated base rate tasks of four distinct levels of congruence. The results of two experiments showed participants were slower and less confident in conflict conditions regardless of their response. However, there were two distinct subsets of participants with different levels of sensitivity to conflict which resulted in different patterns of results when using low base rate ratios. In-depth analyses showed that the impact of base rate information in the formation of metacognitive judgments depended on congruence and response type. Base rate information was a more salient cue for metacognitive processes when responding according to base rates compared to responding according to belief. There is evidence that base rate information may serve as a direct cue for metacognition, independent of fluency.

Keywords: dual process reasoning, metacognition, confidence judgment, conflict monitoring, base rate neglect

Introduction

The dual process approach

For the past few decades psychology of reasoning has been developing under the umbrella of dual process theories of thinking. The simplified view of dual process theories states there are two systems or types of thinking. System 1 is fast, intuitive, and requires little cognitive resources. System 2, on the other hand, is slower, analytical with high cognitive costs (Evans, 2007). In his reviews Evans (2012a; 2012b) tackles typical fallacies concerning the dual process framework, e.g. all dual process theories are the same. Many misconceptions remain, even for researches within this field, so it is important to emphasize the complexity and abundance of findings concerning different dual process theories (for a detailed review see Evans, 2008; 2012a; 2012b; Osman, 2004).

Most of the proposed models of dual process reasoning can be categorised into one of the two main approaches. The default-interventionist view proposes a serial nature of dual processing. Type 1 processes are automatic and fast, while Type 2 processes may or may not interfere and provide a more in-depth analysis and possibly override the Type 1 response (Evans, 2012b). The parallel approach hypothesizes that both types of reasoning are engaged simultaneously, with a possibility of the analytic process overriding the results of the Type 1 process in case of conflict (Evans, 2012b). On one hand, there is a question about what would initiate Type 2 processing in the serial approach. On the other hand, assuming the two distinct systems are engaged in solving the same problem, parallel processing seems wasteful. Recently, there has been growing interest in redefining the nature of Type 1 and Type 2 processing. Some differences, like the speed asymmetry which dominates descriptions of the two types of processes, have been called into question by a number of researchers (De Neys, 2014;

Newman, Gibb, & Thompson, 2017; Pennycook & Thompson, 2012; Trippas, Thompson, & Handley, 2017).

One of the recent proposals has been laid out by De Neys in empirical and theoretical papers (Bago & De Neys, 2017a; De Neys, 2012; De Neys, 2014; De Neys & Glumicic, 2008). He proposes what could be described as a semi-parallel model of reasoning. Type 1 processes run in parallel with what he describes as a *shallow analytic process* (De Neys & Glumicic, 2008). This shallow analytic process is fast enough to produce a result which may, or may not, be in conflict with the output of the Type 1 process. If the conflict is successfully detected then what is traditionally thought of as Type 2 processing is engaged. The mere engagement of Type 2 processes does not necessarily lead to an override of the Type 1 response. De Neys (2012) and, more recently, Bago and De Neys (2017a) proposed multiple Type 1 processes which run in parallel. These Type 1 processes include logical, or probabilistic intuitions as well as belief based Type 1 processing. This approach elegantly fits in with the main findings which indicate people are sensitive to conflict even when the belief based response is not overridden. Research has shown conflict decreases participant confidence and prolongs response times in various reasoning tasks (Pennycook, Fugelsang, & Koehler, 2015; Pennycook, Trippas, Handley, & Thompson, 2014; Thompson, & Johnson, 2014).

Another novel approach has been proposed by Pennycook, Fugelsang, and Koehler (2015). They suggest a three stage dual process model of reasoning as shown in Figure 1.

[Figure 1 about here]

In the first stage, Type 1 processes generate cued intuitive responses (IR_1-IR_n) depending on the task. There is potential for more than one answer to be cued, as in the next example of a simple base rate task:

> Person A is well organized.
> The group from which person A is chosen at random consists of 900 artists and 100 lawyers.
> It is more probable that person A is:
> 1. An artist
> 2. A lawyer

One intuitive answer is that this particular person is a lawyer, because it fits the stereotype of lawyers being well organized. However, the mathematical probability is overwhelmingly in favour of person A being an artist. Following the proposed model two intuitive answers may be generated, one based on stereotypes, the other based on base rate information. These intuitive responses may not have the same level of importance, and processes that give rise to these responses may not have identical timelines. Research shows most participants respond according to belief in similar tasks (De Neys, Vartanian, & Goel, 2008; Obrecht & Chesney, 2016; Pennycook et al., 2015; Pennycook et al., 2014). The relative importance or dominance of a response may be linked with the fluency at which it was generated, with the more fluent response being labelled as IR_1. Stage two is thus the key to linking Type 1 and Type 2 thinking. This conflict monitoring process potentially detects conflicts between generated intuitive answers. Adhering to the previous example, if the conflict between two potential answers is detected one of two things may happen. The process of rationalization preserves the dominance of the IR_1 response. The decoupling process, on the other hand, results with an override of the initial dominant response in favour of the alternative response. One of the other initial responses (in this example a response based

on base rate information) may be accepted, or further processing may be applied to produce a new alternative. Pennycook et al. (2015) showed participants were slower when reacting to incongruent trials, regardless of the response they gave. The key finding here was that participants were slower when responding in accordance with belief in incongruent compared to congruent trials of the base rate task. This indicated participants are sensitive to the conflict even if the dominant response has not been overridden.

Both proposals (Bago & De Neys, 2017a; Pennycook et al., 2015) presume parallel processing in the initial stage of reasoning. There is no definite consolidation of the terminology used: multiple Type 1 processes (Glockner & Witteman, 2010; Pennycook et al., 2015; Bago & De Neys, 2017a), analytical intuitions (De Neys, 2012; 2014), Type 3 processes (De Neys, 2014), and shallow analytic processes (De Neys & Glumicic, 2008). While keeping in mind that, so far, research findings fit well regardless of these terminological differences, there is a question of whether these terms can be clearly distinguished in order to enable the empirical testing of different parallel approaches. The recent views of dual process reasoning all agree that early stage processes run in parallel and that outputs from these processes may or may not be in conflict. The detection of this conflict seems to play a vital role in the engagement of Type 2 processes.

Conflict monitoring and metacognition

Conflict detection has been highlighted as an important meta-process while performing various tasks, particularly reasoning tasks (De Neys, 2014; De Neys & Glumicic, 2008; De Neys, Moyens, & Vansteenwegen, 2010; Evans, 2007; Howarth, Handley, & Walsh, 2016). Conflict detection seems to play an important role not only for performance, but for metacognitive judgments as well. Metacognitive research has mainly concerned

meta-memory and learning, but has spread rapidly during the past decade (Ackerman & Thompson, 2015; 2017). Metacognitive processes primarily concern evaluations and monitoring of task demands, solvability, and performance. Ackerman and Thompson (2015) lay out the basic framework of different meta-reasoning judgments (which can be seen in Table 1) based on work by Nelson and Narens (1990).

[Table 1 about here]

Previous research has shown participants have little insight into objective performance for syllogistic reasoning (Bajšanski, Močibob, & Valerjev, 2014; Markovits, Thompson, & Brisson, 2015; Thompson & Johnson, 2014; Thompson & Morsanyi, 2012), the Wason selection task (Thompson, Evans, & Campbell, 2013), as well as base-rate type tasks (Pennycook et al., 2014), with a tendency to overestimate their performance. It seems judgments of confidence are made under the influence of other factors. Thompson and her colleagues (Thompson, Prowse Turner, & Pennycook 2011; Thompson et al., 2013; Thompson & Johnson, 2014) found answer fluency (speed and ease of generating answers) to be a significant indicator for both feeling of rightness (FOR) and retrospective confidence judgments in reasoning tasks. Outcomes of other metacognitive processes may be used as indicators of final confidence judgments: e.g. Markovits et al. (2015) found a positive correlation between initial judgments of solvability and final confidence judgments.

Levels of induced conflict, and conflict detection also influence confidence judgments. Pennycook et al. (2014) concluded higher levels of conflict lowered confidence judgments using base-rate tasks. Conflict detection has been found to induce an autonomic response in fMRI (functional magnetic resonance imaging) and SCR (skin

conductance response) research (De Neys et al., 2008; 2010). Results showed the anterior cingulate cortex is activated to a larger degree when participants are solving incongruent compared to congruent trials of a base rate task. This study also showed the RLPFC (right lateral prefrontal cortex) was activated when base rate responses were given in incognruent trials indicating its importance for conflict resolution, specifically, decoupling. Research using SCR has shown an increase of autonomic arousal when participant solved incongruent trials in a logical reasoning task (De Neys et al., 2010). This activity, which accompanies conflict detection, may influence metacognitive judgments directly as well as indirectly by decreasing fluency. The effect of conflict on fluency (by measuring response times) has been well documented (Howarth et al. 2016; Thompson & Johnson, 2014) thus the indirect influence on confidence judgments seems obvious.

Aims and rationale for the current study

While most research into meta-reasoning processes is conducted under the broad dual process approach, the focus thus far has been on the main determinants of metacognitive judgments such as: fluency, conflict detection and believability. The goal of the current study was to combine the meta-reasoning framework and the recent approaches to dual processing in order to better understand differences in confidence, depending on specific reasoning processes with a focus on stages 2 and 3 of the model proposed by Pennycook et al. (2015). At the same time, confidence judgments, as a function of induced conflict and response type, may provide further insight into the proposed model, beyond response time analysis.

The base rate task was chosen for this purpose because previous research has shown that it reliably induces conflict through the manipulation of congruence of the description and base rates. Recent research has been based on the adaptation of the

original task from Kahneman and Tversky (1973) by De Neys and Glumicic (2008), and a review of this research has shown there have been significant variations in the experimental procedures depending on research goals. These variations include: presentation of the task, manipulating base rate ratio extremity, and type of response. First, most of the studies presented participants with the entire text of the task after which they were required to provide their response (Newman et al., 2017; Pennycook, Fugelsang, & Koehler, 2012; Pennycook & Thompson, 2012; Thompson & Johnson, 2014; Thompson et al., 2011). This type of presentation results in response times (when measured) which include reading time, and are in general quite long (e.g. Thompson et al., 2011, average response time, congruent items 13.46 s). In order to better control these non-systematic variable factors we opt for a procedure similar to Pennycook et al. (2015). Secondly, most research using base rate tasks has involved extreme base rate ratios (e.g. 995/5) in order to produce a clear effect when the two sources of information are in conflict (Bialek, 2017; De Neys & Glumicic, 2008; Franssens & De Neys, 2009; Pennycook & Thompson, 2012; Pennycook et al., 2014; Stanovich & West, 2008). Some researchers (Bago & De Neys, 2017a; 2017b; Newman et al., 2017; Pennycook et al., 2012; 2014) also explored the effect of conflict when using moderate ratios (e.g. 700/300) which has prompted a separate discussion about the threshold for conflict detection (Pennycook et al., 2012; De Neys, 2014) in the base rate task. For the purposes of this study we used two levels of base rate extremity both to compare the different levels of induced conflict and to possibly provide some insight into this discussion. Finally, participants can be asked to decide from which group the described person was most likely chosen, or to estimate the probability of that person being from a particular group. The first type of response is a categorisation, while the other is a continuous estimate. We chose the former because categorisation is a clearer indicator

of which process occurred (e.g. decoupling or rationalization), while continuous estimates are better suited to other research. By opting for these methodological choices, alongside the use of confidence judgments, we used a novel combination (similar to Bago & De Neys, 2017a) with which we hoped to provide further insights into metacognitive processing during reasoning.

On the basis of previous research we hypothesized confidence judgments would be decreased for incongruent when compared to congruent trials. More specifically, we assumed confidence would be lower for high compared to low base rates when responding stereotypically in the incongruent conditions. On the other hand, we expected confidence to be higher in high base rate conflict trials compared to low base rate trials after the process of decoupling. In other words, the influence of base rate information on confidence was expected to depend on which process from stage 3 in Figure 1 took place.

EXPERIMENT 1

Method

Participants and materials

The sample (N=30) was recruited among undergraduate psychology students with no prior knowledge of the specific research area or the base rate task. Base rate neglect task items were constructed by varying congruence and base rate probabilities. The single presented trait was either congruent or incongruent with base-rate probabilities. Two levels of base-rate probabilities were produced; high and low base rates. To accomplish this we set the following restrictions:

(1) High base rates could not exceed a 90-10 ratio (900-100)
(2) High base rates could not go below an 85-15 ratio (850-150)

(3) Lower base rates could not exceed a 60-40 ratio (600-400)
(4) Lower base rates could not go below a 55-45 ratio (550-450)
(5) Ratios were generated randomly

Previous research has mainly been conducted by using extreme base rates with ratios 995/5 or higher (De Neys & Glumicic, 2008; De Neys et al., 2008; Obrecht & Chesney, 2016; Pennycook et al, 2014), and moderate base rates with ratios of 700/300, not lower (Pennycook et al., 2015). Our goal was to construct items which felt random and less extreme to participants. Items were constructed with traits stereotypical to one group, but not exclusive to it, and still plausible for members of the other group. In this way 80 items were constructed, then rated by the researchers on a five point scale. Items well-adjusted to the Croatian language and culture were rated as most appropriate. A total of 43 items were selected based on the ratings: three practice items, and forty for the main measurement. In total, there were ten items for each level of theoretical conflict based on congruence and base rate probability.

(1) Congruent high base rates – no conflict
(2) Congruent low base rates – no conflict
(3) Incongruent low base rates – low conflict between stereotypical and probability answers
(4) Incongruent high base rates – high conflict situation

Examples of the 4 levels can be seen in Table 2.

[Table 2 about here]

Items were randomly assigned to one of the four situations. Order of group presentation and order of response presentation were counterbalanced with respect to congruence

and group size. All of the items were presented randomly for each participant. The experiment was designed using E-Prime v.2.0.10.356., and conducted on five identical PCs located in the Department's Laboratory for Experimental Psychology.

Procedure

Prior to the main measurement, participants completed a short choice reaction time task to familiarize themselves with the way in which they were supposed to respond. In a total of twenty trials either the number one or number two was presented on a screen and participants had to respond as fast as possible by pressing the appropriate key on the keypad. The purpose was to, as much as possible, reduce error of measurement due to psychomotor variability and non-intended responses. Following this, participants completed three practice items.

For the main measurement, participants were told a single trait describing a person would be presented for a few seconds, after which they would receive information about the group from which the person was randomly selected. Finally, a question about whether it was more likely the person was chosen from group one or two was to be presented. They were instructed to choose which answer they considered more probable by pressing the corresponding numerical key on the keyboard as fast as possible. Response times were measured only for the decision phase to better control variability in processes such as reading, similar to Pennycook et al. (2015). After making a decision they had to make a judgment of confidence on a six point scale (with 10% increments) from 50% (pure guessing, no confidence) to 100% (complete confidence). For a detailed description of the single trial procedure see Figure 2.

[Figure 2 about here]

In half of the items answers according to base-rate probabilities appeared as the first option, while in the other half they appeared as the second option. Responses, response times and judgments of confidence were collected for analysis.

Results

Total data set analyses

Before analysis, response time data was processed to eliminate outliers: responses outside the +/- 3 SD range were removed, a total of 2.25% of all responses. To make sure no outlier items were present within each version an item analysis was performed. One-way repeated measures analyses of variance were calculated for response times at each theoretical conflict level. None of the analyses showed significant effects (all $F(9, 261) < 1.63$, $p > .1$) which means items at every congruence level were balanced.

Mean response times and confidence judgments across ten items at each level of congruence were calculated for each participant and made up the final data for analysis (Table 3).

[Table 3 about here]

As expected, participants responded stereotypically most of the time (87% of all cases) with a noticeable drop for incongruent items. To test for congruence and base rate extremity effects on response choice, a 2x2 repeated measures analysis of variance was calculated. Results showed significant main effects of congruence ($F(1, 29) = 13.38$, $p < .01$, $\eta_p^2 = .32$), and base rate extremity ($F(1, 29) = 16.77$, $p < .01$, $\eta_p^2 = .37$), as well as a significant interaction ($F(1, 29) = 8.39$, $p < .01$, $\eta_p^2 = .22$). Participants generally responded stereotypically more often for congruent items, and for lower base-rates. Post-hoc analyses (Tukey HSD) showed significantly less stereotypical choices

for the incongruent high base rate situation, which was the highest conflict situation. Stereotypical choices were significantly less common for this than any other situation.

A 2x2 analysis of variance was calculated for response times and judgments of confidence. Results showed a significant congruence effect ($F(1, 29) = 21.50$, $p < .01$, $\eta_p^2 = .43$), and congruence by base rate extremity interaction ($F(1, 29) = 8.47$, $p < .01$ $\eta_p^2 = .23$). Response times were significantly shorter for congruent items, and post-hoc analysis showed the high incongruence situation resulted in the slowest responses.

Judgments of confidence yielded a similar pattern: a significant congruence effect ($F(1, 29) = 18.13$, $p < .01$, $\eta_p^2 = .38$), and congruence by base rate interaction ($F(1, 29) = 7.74$, $p < .01$, $\eta_p^2 = .21$). While participants were generally very confident in their choices (an average of 85.85% across all items) they were significantly more confident for congruent compared to incongruent items. As expected the highest incongruence level situation resulted with significantly lower levels of confidence than the other situations.

Analyses of stereotype based responses

The same 2x2 analyses of variance were conducted on response times and confidence ratings only for stereotypical choices. Analysis of response times showed a significant congruence effect ($F(1, 28) = 16.25$, $p < .01$, $\eta_p^2 = .37$) with faster responses for congruent compared to incongruent items. Confidence rating analysis, similarly, resulted in a significant congruence effect ($F(1, 28) = 17.71$, $p < .01$, $\eta_p^2 = .39$), and a marginal congruence/base-rate interaction ($F(1, 28) = 4.09$, $p = .051$, $\eta_p^2 = .13$). Participants made significantly higher confidence ratings for congruent compared to incongruent items, with the lowest confidence rating for the highest conflict situation. Response times and confidence judgments for stereotypical responses can be seen in Figure 3.

[Figure 3 about here]

Correlation analysis

In addition, a correlational analysis was conducted on two levels. Participant level correlations indicate whether participants who were faster were also more confident. Pearson correlation coefficients were calculated between response times and judgments of confidence for each of the four experimental conditions. The correlation was not significant for the two congruent conditions ($r(28) < .21$, $p > .05$). For the conflict conditions the correlation was marginally significant in the high base rate condition ($r(28) = -.34$, $p = .06$), and significant in the low base rate condition ($r(28) = -.51$, $p < .01$). Generally speaking faster participants were more confident in the conflict conditions. Finally, we calculated an item-level correlation between response times and confidence judgments. Response times were negatively correlated with confidence judgments ($r(38) = -.59$, $p < .01$). Participants were more confident for trials which they solved faster.

Discussion

As expected, participants responded stereotypically in the vast majority of trials, even in incongruent situations with high base rates. Higher levels of conflict reduced response speeds and decreased confidence ratings. When analysing only the stereotypical answers the same trend was observed, so even when participants ignored the base rate information they were aware of it on some level. Conflict monitoring processes successfully detected the conflict between the description and base rates on a portion of trials thus pushing participants into Type 2 processing in order to resolve the conflict. This detection and resolution of conflict regardless of the final response increased

response times and decreased confidence ratings. Interpreting these results within the framework proposed by Pennycook and his colleagues (2015), it would seem participants rationalized their initial, intuitive response based on stereotypes, even after detecting conflict with base rate information. Confidence ratings were high, even though most responses were not normatively correct according to mathematical probability, but higher levels of conflict had a significant impact, lowering confidence judgments. As in previous research, confidence and objective performance were not aligned (Pennycook et al., 2014; Thompson & Johnson, 2014), but correlational analysis showed response times (an index of fluency) were a significant metacognitive cue. Participants who were faster were also more confident. More importantly, item-level analysis showed faster responses were highly correlated with confidence ratings. We can conclude that conflict monitoring is likely a meta-process mediating between different types of processing and is one possible incentive for Type 2 processing. Conflict detection and resolution leads to both prolonged responses and decreased confidence, however, it is not clear whether the experience of conflict is a direct metacognitive cue. High base rates in conflict trials lead to successful conflict detection more often than the low base rate trials.

Unexpectedly, one of the key findings from Pennycook et al. (2012; 2015), which concerns the effect of base rate extremity on response times was not replicated in our experiment. Their research showed extreme base rates (e.g. 995/5) lead to a greater degree of conflict detection compared to moderate base rates (e.g. 700/300) when participants responded according to belief in conflict conditions. Pennycook et al. (2012) reported that responses in the moderate base rate condition were not significantly slower for incongruent compared to congruent trials. A later study (Pennycook et al., 2015) showed moderate base rates can lead to conflict detection by using a rapid

response paradigm. Our results showed participants were significantly slower for incongruent compared to congruent trials in the low base rate condition (e.g. 575/425). There was a slight, but non-significant difference in response times for incongruent trials between the two base rate conditions. Newman et al. (2017) showed base rate extremity had an impact on probability judgments using a different paradigm. Our own analysis of confidence ratings resulted in a marginally significant congruence by base rate extremity interaction. Higher base rates had a greater impact on confidence ratings in incongruent trials than the lower base rates. There is an open question as to the cause of our unique results on the response time data. One possible reason might be a lack of statistical power due to small sample size. Another possibility is that, due to individual differences, there are individuals who show more sensitivity to induced conflict and that our sample consisted of a larger proportion of these participants compared to samples from Pennycook et al. (2012; 2015). Previous studies have shown there are participants who fail to detect conflict (Frey, Johnson, & De Neys, 2017; Mevel et al., 2015; Pennycook et al., 2014; 2015). Individual differences imply there may also be highly conflict sensitive individuals as well. Those individuals are better balanced out by less sensitive participants in larger sample sizes. There is also a need for more detailed analyses to better understand the different processes at the three stages proposed by the model in Figure 1. A more in-depth analysis of response times and confidence ratings in the incongruent conditions may also provide more insight about the importance of base rate information as a metacognitive cue. These analyses are not possible from the available data due to a small sample size and low statistical power.

EXPERIMENT 2

In order to verify the main findings of Experiment 1, and increase statistical strength for more in-depth analyses we replicated the procedure on a larger, independent sample.

Method

Participants and materials

A new sample (N = 51) was recruited among undergraduate psychology students, also with no prior knowledge of the base rate task. None of the participants in this sample took part in the first experiment. The same materials from Experiment 1 were used in Experiment 2.

Procedure

Participants completed the same familiarization task and main measurement as in Experiment 1.

Results

Total data set analyses

Outliers (+/- 3 standard deviations) were again eliminated before the main analyses, which made up 2.60% of all responses. Mean response times, confidence ratings and the proportion of stereotypical answers were calculated for each participant for each of the experimental situations (Table 4).

[Table 4 about here]

As in Experiment 1, participants responded stereotypically in the majority of trials (86.52%) with a noticeable drop for conflict trials. A 2x2 ANOVA showed participants responded in accordance with the description significantly less for the incongruent conditions ($F(1, 50) = 21.64$), $p < .01$, $\eta_p^2 = .30$), and generally more stereotypically for low base rates ($F(1, 50) = 20.87$), $p < .01$, $\eta_p^2 = .29$), with a significant congruence by

base rate extremity interaction ($F(1, 50) = 44.23$), $p < .01$, $\eta_p^2 = .47$). Post hoc analysis showed proportions of stereotype responses were significantly decreased for both incongruent compared to both congruent conditions with an additional significant difference between the two incongruent conditions. The decrease in stereotypical responding (increase of base rate responses) was larger for high base rate compared to low base rate trials in the incongruent conditions.

To test the effects of congruence and base rate extremity, a 2x2 repeated measures ANOVAs were calculated for response times and judgments of confidence. For response times results showed significant main effects of congruence ($F(1, 50) = 18.06$, $p < .01$, $\eta_p^2 = .27$) and base rate extremity ($F(1, 50) = 6.97$), $p < .05$, $\eta_p^2 = .12$), as well as a significant interaction effect ($F(1, 50) = 16.22$, $p < .01$, $\eta_p^2 = .24$). Participants generally reacted more slowly for incongruent and high base rate trials although the slower responses for high base rate trials are a consequence of the interaction effect. Tukey HSD post-hoc tests showed that participants were slower in the high incongruence situation compared to both congruent situations (mean differences: $M = 232.53$ ms; $M = 194.65$ ms). These differences were not significant for the low incongruence situation, even though the same trend remains (mean differences: $M = 71.18$ ms; $M = 33.31$ ms).

The same pattern of results was observed for judgments of confidence, both main effects of congruence ($F(1, 50) = 24.37$, $p < .01$, $\eta_p^2 = .33$), and base rate extremity ($F(1, 50) = 4.35$, $p < .05$, $\eta_p^2 = .08$) as well as the interaction effect ($F(1, 50) = 12.65$, $p < .01$, $\eta_p^2 = .20$) were significant. Participants were generally more confident for congruent and high base rate trials. Post-hoc comparisons showed participants were significantly less confident for high incongruence trials compared to both congruent situations (mean differences: $M = 6.45$ %; $M = 2.92$ %). At the same time, participants

were significantly less confident in the low incongruence compared to the high congruence situation (mean difference, $M = 5.31$ %), but not compared to the low congruence situation (mean difference, $M = 1.78$ %).

Analyses of stereotype based responses

The same 2x2 analyses were conducted responses in accordance with the stereotype. Results were very similar to the previous analysis on the complete data set and can be seen in Figure 4. For response times both the main effects of congruence ($F(1, 48) = 13.74$, $p < .01$, $\eta_p^2 = .22$) and base rate extremity ($F(1, 48) = 7.70$, $p < .01$, $\eta_p^2 = .14$), as well as the interaction ($F(1, 48) = 16.40$, $p < .01$, $\eta_p^2 = .25$) were significant. Participants were generally slower for incongruent trials and for high base rates. Post-hoc analysis showed participants were significantly slower in the high incongruence situation ($M = 1297$ ms) compared to both congruent situations ($M = 1045.80$ ms; $M = 1078.80$ ms) as well as for the low incongruence situation ($M = 1127.80$ ms). Participants were not significantly slower for the low incongruence situation compared to the congruent situations.

For judgments of confidence the main effect of congruence ($F(1, 48) = 19.20$, $p < .01$, $\eta_p^2 = .29$) and the interaction effect ($F(1, 48) = 11.94$, $p < .01$, $\eta_p^2 = .20$) were significant. Participants were less confident for incongruent trials while post-hoc analysis showed that this decrease in confidence was significant for the high incongruence situation ($M = 79.08\%$) compared to both congruent situations ($M = 85.92\%$; $M = 82.98\%$). At the same time participants showed significantly less confidence in the low incongruence situation ($M = 81.08\%$) compared to the high congruence but not the low congruence situation.

[Figure 4 about here]

Rationalization vs. Decoupling analysis

In order to analyse responses made in accordance with the base rates, only responses for the high incongruence situation (28.82% of all responses) were included due to a low percentage of these types of responses in the low incongruence situation (12.85%). We compared these responses to both the stereotypical responses in the high incongruence situation and the high congruence situation. Two one-way ANOVA-s were conducted for the response times and judgments of confidence. For response times the main effect was significant ($F(2, 80) = 8.76$, $p < .01$, $\eta_p^2 = .18$). Post-hoc analysis showed participants were significantly slower for both types of responses in the incongruent compared to the congruent situation. There was no significant difference between the base rate responses ($M = 1474.40$ ms) and stereotypical responses ($M = 1339.40$ ms) in the incongruent situation.

For judgments of confidence the main effect was significant ($F(2, 82) = 18.42$, $p < .01$, $\eta_p^2 = .31$). Post-hoc analysis showed participants were significantly less confident for both types of responses in the incongruent compared to the congruent situation. The difference in confidence between the base rate responses ($M = 74.30\%$) and the stereotypical responses ($M = 77.57\%$) in the incongruent situation was not significant. Response times and confidence judgments can be seen in Figure 5.

[Figure 5 about here]

Stereotypical responses in incongruent situations include both Type 1, and Type 2 rationalization responses (stereotypical responses after conflict was detected). On the other hand, base rate responses in the same situation indicate largely decoupling processes (overriding of the stereotypical responses). Since the previous analysis was

not a pure comparison of rationalization and decoupling, a more refined approach was needed. The goal was to eliminate as many Type 1 responses as possible from the incongruent situation. This would presumably result in a large proportion of Type 2 responses and provide an opportunity to better compare rationalization and decoupling. We assumed that most responses which were faster than the average responses in the high congruence situation were pure Type 1 processing. By this logic, for each participant, responses in the incongruent situation which were faster than his or her individual average in the congruent situation were eliminated. By doing this 42.73% of stereotypical, and 35.37% of the base rate responses were eliminated. Two t-tests were conducted, for response times and judgments of confidence. Results showed there was no significant difference ($t(32) = 1.12$, $p > .05$) in response times between the stereotypical ($M = 1673.08$ ms) and base rate ($M = 1826.96$ ms) responses. The same was found for the judgments of confidence ($t(32) = 0.69$, $p > .05$), with the difference becoming even smaller than in the previous analysis (stereotypical responses $M =$ 74.84%; base rate responses $M = 73.29\%$).

Rationalization as function of conflict level

Analyses of stereotype based responses revealed the usual result which shows participants are slower when responding based on the stereotype in the high compared to the low incongruent condition. The result is usually interpreted as the influence of base rate extremity on the probability of conflict detection (Pennycook et al., 2015). However, it may be more difficult to rationalize the stereotypical response when conflict is high than when it is low, so it could take more time and/or reduce confidence to a greater degree. Using the same logic as in the *Rationalization vs. Decoupling analysis,* we compared stereotypical responses in the two incongruent situations after removing what we could presume to be mostly Type 1 responses. Two t-tests were

conducted for response times and confidence ratings. The results showed there was no significant difference ($t(44) = 1.76$, $p > .05$) in response times between the high (M = 1580.52 ms) and low incongruence (M = 1472.43 ms) situations. The same was true for judgments of confidence ($t(44) = 1.62$, $p > .05$), there was no difference in confidence between the high (M = 77.05%) and low (M = 79.30%) incongruence situations.

Decoupling as a function of conflict level

Additionally, we conducted a limited analysis comparing response times and confidence of base rate responses between the high and low incongruence situations. Results showed no significant difference ($t(20) = 0.02$, $p > .05$) for response times between the high (M = 1204.72 ms) and low incongruence (M = 1206.86 ms). However, there was a significant difference for judgments of confidence ($t(20) = 2.70$, $p < .05$). Participants were more confident for base rate responses in the high (M = 76.84%) compared to the low incongruence situation (M = 68.18%).

Correlation analysis

The same two-level correlational analysis was conducted as for Experiment 1. Correlations between response times and confidence ratings in all of the experimental conditions show that participants who were faster were also more confident. For the high ($r(49) = -.39$, $p < .01$) and low congruence ($r(49) = -.40$, $p < .01$) conditions the correlations were slightly lower than for the high ($r(49) = -.53$, $p < .01$) and low incongruence ($r(49) = -.53$, $p < .01$) conditions. The item-level correlation was also significant ($r(38) = -.52$, $p<.01$). Confidence ratings were higher for items that were solved faster.

Low conflict sensitivity analysis

Thus far the analyses of Experiment 2 data have not provided any insight into the unique results of Experiment 1. In Experiment 1 participants showed a large degree of sensitivity to low conflict (low base rates) which resulted only in a main effect of congruence but not an expected congruence by base rate extremity interaction when responding stereotypically. In Experiment 2, on the other hand, we found the usual result, high base rates reliably induced conflict which was reflected in response times and confidence judgments, but low base rates did not. If the difference in results was not caused by chance and low sample size in Experiment 1 we could assume the participants in Experiment 1 were more sensitive to low conflict than the participants in Experiment 2. In order to test whether there was a substantial subset of participants in Experiment 2 which were also sensitive to low base rates, we divided the sample into two groups. We presumed individuals who gave a larger proportion of base rate responses in incongruent trials were generally more sensitive to conflict. Individuals were assigned to a low and high sensitive group based on the proportion of base rate responses in relation to the sample median. To test whether the high sensitivity group displayed significant conflict detection, two 2(group) x 2(congruence) ANOVA-s were conducted for response times and judgments of confidence in the low base rate condition. For response times both main effects and the interaction effect were significant. Participants in the high sensitivity group were generally slower ($F(1, 48) = 11.11$, $p < .01$, $\eta_p^2 = .19$). Response times were prolonged for conflict trials ($F(1, 48) = 4.06$, $p < .05$, $\eta_p^2 = .08$) which is qualified by the interaction effect ($F(1, 48) = 11.68$, $p < .01$, $\eta_p^2 = .20$). Participants in the high sensitivity group were significantly slower for incongruent ($M = 1310.20$ ms) compared to the congruent trials ($M = 1165$ ms) while there was no significant difference for the low sensitivity group. For judgments of

confidence results showed a significant main effect of congruence ($F(1, 48) = 5.95$, $p < .05$, $\eta_p^2 = .11$) and an interaction effect ($F(1, 48) = 4.05$, $p < .05$, $\eta_p^2 = .08$). Participants were generally less confident in incongruent trials, but post-hoc analysis showed this difference was only significant for the high sensitivity group. Response times and confidence ratings can be seen in Figure 6.

[Figure 6 about here]

Discussion

The goal of Experiment 2 was to replicate the results of the first experiment on a larger, independent sample which also enabled a more in-depth analysis in order to better understand the underlining processes. When comparing the results of analyses on the total data sets there were no differences between the experiments. Incongruence generally prolonged response times, decreased confidence, and increased the use of base rates. However, there were differences when analysing stereotype based responses in the two experiments. Experiment 1 did not replicate the usual interaction effect which would show that low base rates had a lower impact on response times and confidence judgments (Newman et al., 2017; Pennycook et al., 2015). The results showed a main effect of congruence on response times, and only a marginal interaction effect for judgments of confidence. Experiment 2, on the other hand, clearly showed there was an interaction effect. The impact of conflict on response times and judgments of confidence was significantly larger in the high incongruence situation.

However, further analysis showed there were two distinct groups of participants. Those who responded according to base rate information to a larger degree also showed higher sensitivity to conflict when responding stereotypically in incongruent trials. For this subset of participants, there was a significant decrease in response speed and

confidence even when responding stereotypically in the low incongruence condition. The low sensitivity group showed no such differences, mirroring the results of Pennycook and colleagues (2012; 2015). This may indicate the results of Experiment 1 were a consequence of a combination of small sample size and a large proportion of highly conflict sensitive participants. While our findings indicate there is a subgroup of participants who are highly sensitive to conflict, other research has shown individual differences in conflict sensitivity by highlighting participants who fail to detect conflict (Frey et al., 2017; Mevel et al., 2015). Future research of automatic conflict detection and monitoring processes needs to take into account these individual differences.

The larger sample size in the second experiment allowed for a more detailed analysis and comparison between the different experimental conditions as a function of response type. Results showed there was no difference between stereotypical and base rate responses in the high incongruence condition, while participants were significantly slower and less confident in both compared to the congruent condition (Figure 5). If interpreted within the framework proposed by Pennycook et al. (2015) in Figure 1, this would suggest rationalization and cognitive decoupling have a similar impact on response times and confidence. It is important to note that this interpretation presumes that most of base rate responses in the conflict situation represent cognitive decoupling, rather than Type 1 base rate responses or rationalization. However, it is not possible to completely isolate rationalization and decoupling processes. This is due to the fact that stereotypical responses in the incongruent situation consist of both Type 1 stereotype-based answers and Type 2 rationalization responses. On the other hand, based on the proposed model, it is possible that base rate responses consist of Type 1 base rate responses, rationalized base rate responses, and to a larger degree, decoupling based responses. In order to control for this we attempted to eliminate Type 1 responses in the

incongruent situation by removing responses which were faster than the average in the congruent situation. By doing this we presumed most of the remaining data would represent responses in which Type 2 processing had been initiated which would enable a better comparison between rationalization and decoupling processes. Analyses of response times and judgments of confidence for the reduced data set again showed no significant differences between base rate and stereotypical responses in the incongruent situation. Although this approach is by no means perfect or exact, it may provide insight into the relative impact of conflict detection and resolution on reasoning and confidence. It would seem that once conflict has been detected, the resolution of that conflict has a similar timeline and impact on confidence for both rationalization and decoupling responses.

When comparing rationalization responses for the two incongruent situations, there were no significant differences in response times and confidence. There was a trend which showed participants were slightly faster and more confident in the low incongruence situation when they rationalized the stereotypical response, but it would seem base rate extremity has little, if any, impact on this process.

An additional, interesting result was obtained by comparing base rate responses in the high and low incongruence conditions. Even though there was no significant difference in response times between the two situations, there was a large difference in confidence levels. Participants were more confident when responding according to the base rate in the high incongruence (high base rate) situation. The result suggests the time required to resolve conflict in favour of the base rate is not under the impact of base rate extremity. On the other hand, confidence is greatly impacted by base rate extremity when the final response is a base rate response. Participants were more confident for high base rate responses compared to low base rate responses. Previous

research has shown fluency and the experience of conflict serve as major cues for the generation of metacognitive judgments (Ackerman & Zalmanov, 2012; Thompson et al., 2013). Our results showed base rate information can impact confidence independently from other major cues, depending on the type of response.

Correlational analysis on the total data set confirmed faster participants were also more confident. Additionally, an item-level analysis showed higher levels of confidence for items that required less time. This is in line with previous research which identifies fluency as a major metacognitive cue.

General discussion

The goal of this study was to investigate the connections between reasoning and metacognition by combining the meta-reasoning framework and a recent approach to dual process theory. By doing this we hoped to provide further insight into metacognitive processes depending on the level of conflict and outcomes of specific reasoning sub-processes. On a general level our data replicated the main findings of previous research. Response times, as an indicator of fluency, were significantly correlated with confidence judgments on both an inter-individual and inter-item level. The second general result showed decreased confidence and prolonged response times for incongruent compared to congruent trials. These results, as well as the fact that high base rates lead to a higher proportion of base rate responses compared to low base rates, mirror findings by other researchers (Newman et al., 2017; Pennycook et al., 2015).

A more in-depth analysis of response times and confidence judgments provides a more complete view of the relationships between reasoning and metacognition. Conflict detection has a major impact on confidence ratings, if the conflict is detected then confidence decreases. When there is no conflict, base rate information becomes the differentiating factor between the two congruent situations. Participants showed higher

levels of confidence for high base rate congruent trials even though both high and low congruent trials were solved at the same speed. For the incongruent conditions the impact (salience) of base rate information seems to depend on the type of response following different reasoning sub-processes in stage 3 of the model depicted in Figure 1. Results showed confidence was greatly impacted by base rate information during the decoupling process. Participants were equally fast when responding according to base rate information in both incongruent conditions, but confidence was significantly higher for high base rate ratios. On the other hand, base rate information had very little impact during the rationalization process. There were small but significant differences between the stereotypical responses in the high compared to the low incongruence condition, but after filtering what we presumed to be mainly Type 1 responses (responses where the conflict was not detected) these differences were no longer significant. The fact that this process resulted in a stereotypical response seems to have a greater impact on confidence than the differences in base rate ratios. We can speculate that adjustments in metacognitive processes follow reasoning and the outcomes of reasoning processes. Confidence judgments then represent the final response based on those adjustments as well as other cues such as fluency of response.

Base rate extremity did not have a significant influence on response times of stereotypical responses in incongruent conditions in Experiment 1. The expected result was that high base rates would have a greater impact than low base rates. On the other hand, this pattern of results was obtained in Experiment 2 which replicates the findings of Pennycook et al. (2015). However, a detailed analysis of Experiment 2 data revealed two distinct subsets of participants. One group, which we labelled *high sensitivity*, produced results similar to participants in Experiment 1. For this group, there was a significant decrease in response speed and confidence for incongruent compared to

congruent low base rate trials. The second group, *low sensitivity,* produced results in line with those of Pennycook et al. (2012). For that group low base ratios did not lead to conflict detection, specifically there was no decrease in response speed or in confidence compared to the congruent condition. De Neys (2014) and Pennycook et al. (2012) raise the question of a base rate ratio threshold for conflict detection in this type of task. Both articles concluded the threshold might be between 700/300 (moderate) and 900/100 (high). Our results indicate there are people who show sensitivity to conflict even at low base rate ratios (e.g. 575/425). These results show individual differences in conflict sensitivity can be detected even at low levels of conflict. Other studies have shown there are participants who fail to detect conflict entirely (Frey et al., 2017; Mevel et al., 2015). Together, these findings imply a rather large range of individual differences in conflict sensitivity. There remains a question about the nature of these individual differences. For more sensitive participants, base rate information may increase the weight of initially generated base rate responses, thus increasing the probability of conflict detection. Another possibility is that these participants are more sensitive to the disparity between the two sources of information. Both conclusions are plausible with the available data.

Considering research in the field of reasoning from the past decade there is potential for a different interpretation of the current results. One could argue that, due to the nature of the procedure and speed of responses, all of the responses are actually pure Type 1 processing. In this view, the differences in response times would be a result of the difference in the strengths of the two Type 1 processes running in parallel (belief based, and probability based Type 1 processes). By using a two-response paradigm, researchers have attempted to isolate Type 1, and Type 2 responses (e.g. Bago & De Neys, 2017a). Comparing the current procedure with the two-response paradigm, there

is an argument to be made that our data represents initial responses, and that little or no Type 2 processing occurred. However, our data indicates the detection and resolution of conflict can be completed within the observed response times. When participants responded stereotypically in the high base rate incongruent condition they were significantly slower and less confident than when giving the same type of response for congruent trials. Additionally, high sensitive participants showed the same pattern of results for low base rate conditions, while low incongruence had no effect on response times and confidence when responding stereotypically for the low sensitive group. This would indicate that different processes occurred for the two groups, even when providing the same type of response. We propose Type 2 processes (detection + resolution of conflict) are not exclusively a product of rethinking, and are not necessarily conscious, especially for a relatively simple base rate task. Taking this into account, there may be a need to differentiate between these rapid Type 2 processes, and what is traditionally considered Type 2 processing (deliberate, analytical processing). Such a classification would recognize multiple intuitions (Type 1 processes), rapid Type 2 processes (conflict detection + resolution), and traditional analytical processes, which could be designated as Type 3 processing.

Conclusions

First, participants took a longer time and showed decreased confidence for conflict trials in both experiments. On a more specific level, confidence ratings depend on base rate extremity, conflict presence/detection, and response type. Base rate ratios may be a salient cue for confidence judgments during decoupling compared to the rationalization process. Additionally, results indicate the existence of individual differences in conflict sensitivity which lead to significant conflict detection in the low base rate ratio condition for a subset of participants. The results fit in well with the recently proposed

model by Pennycook et al. (2015) while providing more detailed insight into stage 2 and 3 processes (conflict detection and resolution). Finally, we propose that, in addition to the existence of multiple Type 1 processes, there are rapid, possibly non-conscious Type 2 processes (conflict resolution) which are different from what is traditionally thought of as analytical processing.

References

Ackerman, R., & Thompson, V.A. (2015). Meta-reasoning: What can we learn from meta-memory? In A. Feeney & V.A. Thompson (Eds.), *Reasoning as Memory* (pp. 164-182), New York: Psychology Press.

Ackerman, R., & Thompson, V. A. (2017). Meta-reasoning: Monitoring and control of thinking and reasoning. *Trends in Cognitive Sciences, 21*(8), 607-617.

Ackerman, R., & Zalmanov, H. (2012). The persistence of the fluency-confidence association in problem solving. *Psychonomic Bulletin Review, 19,* 1187-1192.

Bago, B., & De Neys, W. (2017a). Fast logic?: Examining the time course assumption of dual process theory. *Cognition, 158*, 90-109.

Bago, B., & De Neys, W. (2017b). Rise and fall of conflicting intuitions during reasoning. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 87-92). Austin, TX: Cognitive Science Society.

Bajšanski, I., Močibob, M., & Valerjev, P. (2014). Metacognitive judgments and syllogistic reasoning. *Psychological Topics, 23*(1), 143-165.

Bialek, M. (2017). Not that neglected! Base rates influence related and unrelated judgments. *Acta Psychologica, 177,* 10-16.

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

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

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.

Evans, J.St.B.T. (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking & Reasoning, 13*(4), 321-339.

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

Evans, J.St.B.T. (2012a). Dual-Process theories of deductive reasoning: Facts and fallacies. In K.J. Holyoak & R.G. Morrison (Eds.), *The Oxford Handbook of Thinking and Reasoning*, (pp. 115-133), New York: Oxford University Press.

Evans, J.St.B.T. (2012b). Questions and challenges for the new psychology of reasoning. *Thinking & Reasoning, 18*(1), 5-31.

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

Frey, D., Johnson, E.D., & De Neys, W. (2017). Individual differences in conflict detection during reasoning. *Quarterly Journal of Experimental Psychology*. Advance online publication. doi: 10.1080/17470218.2017.1313283.

Glockner, A., & Witteman, C. (2010). Beyond dual-process models: A categorization of processes underlying intuitive judgment and decision making. *Thinking & Reasoning, 16*(1), 1-25.

Howarth, S., Handley, S.J. & Walsh, C. (2016). The logic-bias effect: The role of effortful processing in the resolution of belief-logic conflict. *Memory & Cognition, 44*, 330-349.

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

Markovits, H., Thompson, V.A., & Brisson, J. (2015). Metacognition and abstract reasoning. *Memory & Cognition, 43*, 681-693.

Mevel, K., Poirel, N., Rossi, S., Cassotti, M., Simon, G., Houdé, O., & De Neys, W. (2015). Bias detection: Response confidence evidence for conflict sensitivity in the ratio bias task. *Journal of Cognitive Psychology, 27*(2), 227–237.

Nelson, T.O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Brower. (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory* (Vol. 26, pp. 125-173). San Diego: Academic Press.

Newman, I.R., Gibb, M., & Thompson, V.A. (2017). Rule-based reasoning is fast and belief-based reasoning can be slow: Challenging current explanations of belief-bias and base-rate neglect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(7), 1154-1170.

Obrecht, N.A., & Chesney, D.L. (2016). Prompting deliberation increases base-rate use. *Judgment and Decision Making, 11*(1), 1-6.

Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin & Review, 11*(6), 988-1010.

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

Stanovich, K.E., & West, R.G. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology, 94*(4), 672-695.

Thompson, V.A., Evans, J.St.B.T., & Campbell, J.I.D. (2013). Matching bias on the selection task: It's fast and feels good. *Thinking & Reasoning, 19*(3), 431-452.

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

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

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

Thompson, V.A., Prowse Turner, J.A., Pennycook, G., Ball, L.J., Brack, H., Ophir, Y., & Ackerman, R. (2013). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition, 128*, 237-251.

Trippas, D., Thompson, V.A., & Handley, S.J. (2017). When fast logic meets slow belief: Evidence for a parallel-processing model of belief bias. *Memory & Cognition, 45*(4), 539-552.

Figure Captions

Figure 1. Three stage, dual process model of reasoning (Pennycook et al., 2015, p. 39).

Figure 2. Example of a single trial procedure.

Figure 3. Response times (left) and confidence judgments (right) as a function of congruence and base rate extremity for stereotype based responses in Experiment 1 (spreads represent 95% confidence).

Figure 4. Response times (left) and confidence judgments (right) as a function of congruence and base rate extremity for stereotype based responses in Experiment 2 (spreads represent 95% confidence).

Figure 5. Response times and confidence judgments for correct congruent, stereotype, and base rate responses in Experiment 2 (spreads represent 95% confidence).

Figure 6. Response times (left) and confidence judgments (right) as a function of congruence and group conflict sensitivity in low base rate conditions (spreads represent 95% confidence).

Table 1. *Meta-reasoning judgments*

Judgement of Solvability (JOS)	A judgement of the probability that a particular task is solvable given the type of task, experience, knowledge and capacity of the person making this judgement.
Feeling of Rightness (FOR)	Feeling elicited by Type 1 generated responses, a mediator between Type 1 and Type 2 processes (lower FOR → higher probability of initiating Type 2 processes).
Ongoing judgements	Warmth rating, intermediate confidence rating, dynamic prediction of knowing.
Final confidence judgements (FCJ)	Retrospective confidence rating of the final solution to a problem or task.

Table 2. *Examples of the four experimental situations*

	Congruence	
Base rate	Congruent	Incongruent
High	Person A is organized The group consists of 874 accountants and 126 artists.	Person B is physically strong The group consists of 860 teachers, and 140 boxers
Low	Person C is romantic The group consists of 580 poets, and 420 surgeons.	Person A is authoritative The group consists of 568 janitors and 432 policemen.

Table 3. *Mean values and standard deviations (in brackets) of the total data set in Experiment 1*

	Response times (ms)	Confidence (%)	Stereotypical answers (%)
Congruent – high base rate	897.01 (303.80)	88.70 (8.45)	93.00 (15.57)
Congruent – low base rate	979.18 (431.03)	87.17 (8.91)	94.67 (9.00)
Incongruent – high base rate	1264.02 (528.47)	82.50 (11.31)	72.33 (26.22)
Incongruent – low base rate	1172.26 (620.46)	85.03 (10.37)	88.00 (24.97)

Table 4. *Mean values and standard deviations of the total data set in Experiment 2*

	Response times (ms)	Confidence (%)	Stereotypical answers (%)
Congruent – high base rate	1049.84 (320.30)	85.51 (8.27)	95.10 (9.03)
Congruent – low base rate	1087.70 (313.19)	81.98 (10.05)	92.55 (9.13)
Incongruent – high base rate	1282.36 (452.46)	79.06 (11.83)	71.18 (26.96)
Incongruent – low base rate	1121.02 (345.20)	80.20 (11.31)	87.25 (22.72)

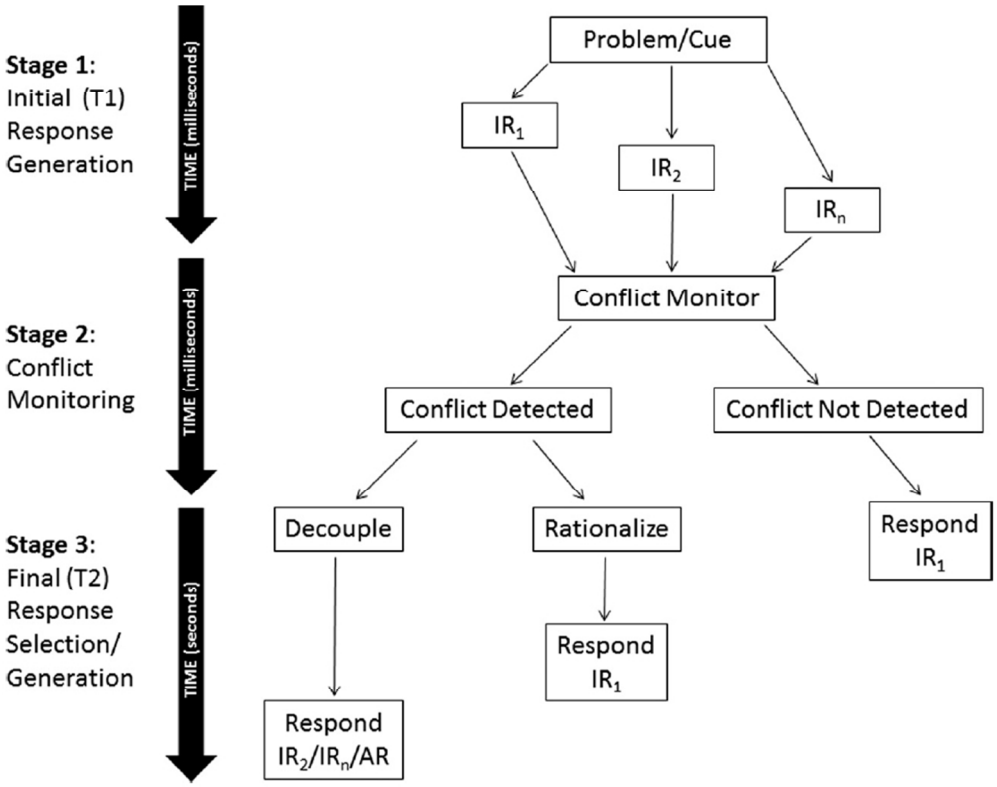

Figure 1. Three stage, dual process model of reasoning (Pennycook et al., 2015, p. 39).

81x65mm (300 x 300 DPI)

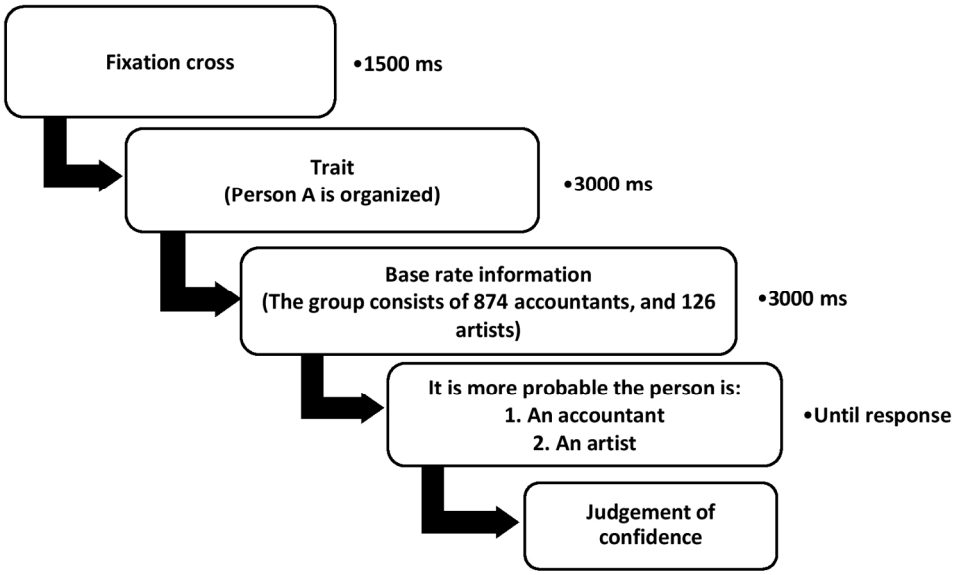

Figure 2. Example of a single trial procedure.

154x95mm (300 x 300 DPI)

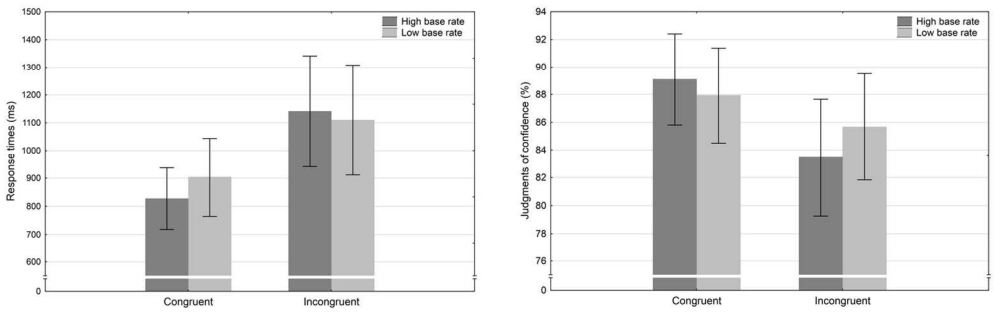

Figure 3. Response times (left) and confidence judgments (right) as a function of congruence and base rate extremity for stereotype based responses in Experiment 1 (spreads represent 95% confidence).

135x46mm (300 x 300 DPI)

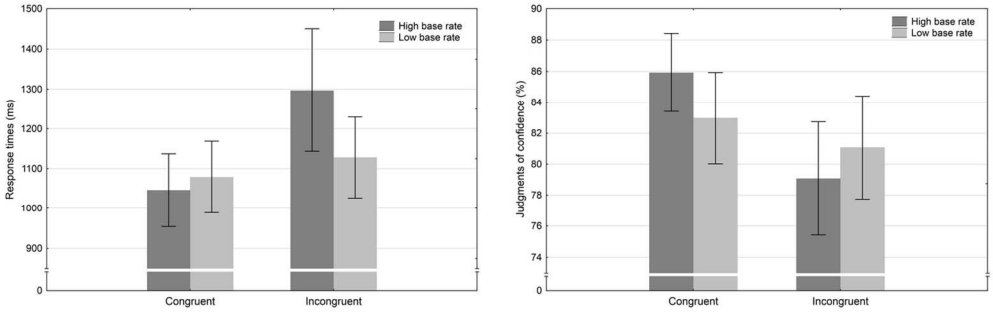

Figure 4. Response times (left) and confidence judgments (right) as a function of congruence and base rate extremity for stereotype based responses in Experiment 2 (spreads represent 95% confidence).

124x39mm (300 x 300 DPI)

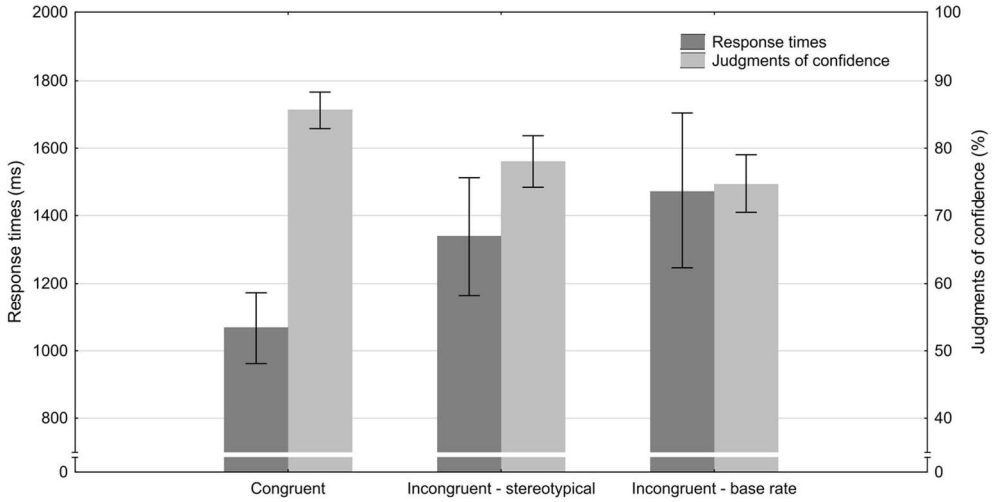

Figure 5. Response times and confidence judgments for correct congruent, stereotype, and base rate responses in Experiment 2 (spreads represent 95% confidence).

123x63mm (300 x 300 DPI)

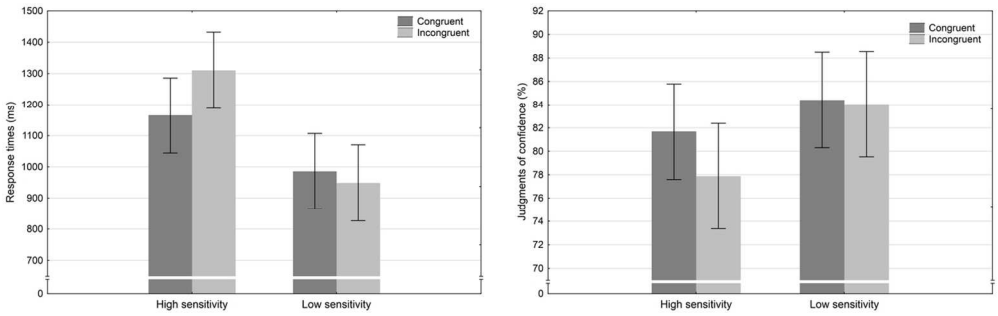

Figure 6. Response times (left) and confidence judgments (right) as a function of congruence and group conflict sensitivity in low base rate conditions (spreads represent 95% confidence).

120x37mm (300 x 300 DPI)