Bias detection: response confidence evidence for conflict sensitivity in the ratio bias task

Journal:	*Journal of Cognitive Psychology*
Manuscript ID:	JCP-FA 14-101.R2
Manuscript Type:	Full Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Mevel, Katell; Sorbonne, U8240 Laboratory for the Psychology of Child Development and Education (LaPsyDÉ); Université de Caen Basse-Normandie, Université Paris Descartes, Sorbonne Paris Cité, Poirel, Nicolas; Sorbonne, U8240 Laboratory for the Psychology of Child Development and Education (LaPsyDÉ); Université de Caen Basse-Normandie, Université Paris Descartes, Sorbonne Paris Cité, ; Institut Universitaire de France (IUF), Rossi, Sandrine; Sorbonne, U8240 Laboratory for the Psychology of Child Development and Education (LaPsyDÉ); Université de Caen Basse-Normandie, Université Paris Descartes, Sorbonne Paris Cité, Cassotti, Mathieu; Sorbonne, U8240 Laboratory for the Psychology of Child Development and Education (LaPsyDÉ); Université de Caen Basse-Normandie, Université Paris Descartes, Sorbonne Paris Cité, Simon, Grégory; Sorbonne, U8240 Laboratory for the Psychology of Child Development and Education (LaPsyDÉ); Université de Caen Basse-Normandie, Université Paris Descartes, Sorbonne Paris Cité, Houdé, Olivier; Sorbonne, U8240 Laboratory for the Psychology of Child Development and Education (LaPsyDÉ); Université de Caen Basse-Normandie, Université Paris Descartes, Sorbonne Paris Cité, ; Institut Universitaire de France (IUF), De Neys, Wim; Sorbonne, U8240 Laboratory for the Psychology of Child Development and Education (LaPsyDÉ); Université de Caen Basse-Normandie, Université Paris Descartes, Sorbonne Paris Cité, ; CNRS,
Keywords:	reasoning, conflict detection, response confidence, ratio bias, heuristics

BIAS DETECTION: RESPONSE CONFIDENCE EVIDENCE FOR CONFLICT SENSITIVITY IN THE RATIO BIAS TASK

Running head: Response confidence in the ratio bias task

Katell Mevel[1,2,5]*, Nicolas Poirel[1,2,3,5], Sandrine Rossi[1,2,5], Mathieu Cassotti[1,2,5], Grégory Simon[1,2,5], Olivier Houdé[1,2,3,5], Wim De Neys[1,2,4,5]

1 – CNRS U8240, Laboratory for the Psychology of Child Development and Education (LaPsyDÉ), Sorbonne, 46 rue Saint-Jacques F-75005 Paris, France

2 – Université de Caen Basse-Normandie, Université Paris Descartes, Sorbonne Paris Cité, France

3 – Institut Universitaire de France (IUF), France

4- CNRS, U8240, LaPsyDÉ, France

5 – GIP Cyceron, Bld Henri Becquerel F-14074 Caen, France

*Corresponding author

Dr. Katell MEVEL, GIP de Cyceron, Blvd Henri Becquerel BP 5229, F-14074 Caen Cedex;

Phone: +33(0)231 470 144; Email: katell.mevel@parisdescartes.fr

Prof. Nicolas Poirel: Phone: +33(0)231 470 144; Email: nicolas.poirel@parisdescartes.fr

Dr. Sandrine Rossi: Phone: +33(0)231 470 137; Email: rossi@cyceron.fr

Dr. Mathieu Cassotti : Phone: +33(0)140 462 995; Email: mathieu.cassotti@parisdescartes.fr

Dr. Grégory Simon: Phone: +33(0)231 470 137; Email: simon@cyceron.fr

Prof. Olivier Houdé: Phone: +33(0)140 462 995; Email: olivier.houde@parisdescartes.fr

Dr. Wim de Neys: +33(0)140 462 995; Email: wim.de-neys@parisdescartes.fr

BIAS DETECTION: RESPONSE CONFIDENCE EVIDENCE FOR CONFLICT SENSITIVITY IN THE RATIO BIAS TASK

Running head: Response confidence in the ratio bias task

FUNDING

This work was supported by the French National Research Agency under Grant number ANR-12-JSH2-0007-01. The authors declare that there is no conflict of interests regarding the publication of this paper.

ACKNOWLEDGMENT

The authors would like to thank all the individuals who participated in this study.

ABSTRACT

Human reasoning is often biased by heuristic thinking. A key question is whether people detect that their heuristic answer conflicts with logical considerations. Empirical studies suggest that the detection is typically successful but the generality of these findings has been questioned. The present study focuses on this issue. A response confidence measure was used to validate conflict sensitivity findings in the classical ratio bias task and identify individual differences in conflict detection efficiency. Participants were asked to indicate how confident they were after solving problems for which a cued heuristic response could be inconsistent or consistent with the correct response. Results confirmed that most reasoners showed a confidence decrease when they were biased, suggesting that they acknowledge that their intuitive answers are not fully warranted. However, there were also subgroups of reasoners who failed to show a confidence effect. Implications for the debate on conflict detection during thinking are discussed.

Keywords: reasoning, conflict detection, response confidence, ratio bias, heuristics

INTRODUCTION

Imagine that you are given a choice of two trays containing red and white marbles. In each tray, there are always more white than red marbles. The small tray contains 1 red marble out of a total of 10. The large tray contains 9 red marbles out of a total of 100. You get to draw one marble out of one of the trays. If this drawn marble is red, you win a prize. Of course, you have to pick without looking and the trays are shaken up before you draw. Which one of the trays should you draw from to maximize your chances of winning?

When presented with this problem, a lot of people have a strong intuitive preference for the large tray. From a logical point of view, this is not a smart choice. Indeed, although the large tray contains more red marbles than the small tray, there are also a lot more white marbles in the large tray. If you take the ratio of red and white marbles in both trays into account, it is clear that the small tray is giving you a 10% chance of picking a red marble (i.e., 1/10) while the large tray only offers a 9% chance (i.e., 9/100). The striking thing is that although one doesn't need to be a math genius or logician to figure this out, many educated people nevertheless fail to solve this basic "ratio" problem (Epstein, 1994). The fact that the absolute number of red marbles is higher in the large tray has such a strong intuitive pull that people seem to neglect the ratio principle and erroneously conclude that they should draw from the large tray. This so-called "ratio bias" thus leads people to make "irrational" choices for problems such as those presented in the marbles task.

Decades of research have shown that similar intuitive thinking applies to a wide range of logical and probabilistic reasoning tasks (Evans, 2008; Evans & Over, 1996; Kahneman & Frederick, 2005; Kahneman & Tversky, 1973). In general, human reasoners seem to have a strong tendency to base their judgment on fast intuitive impressions like the one induced by the marble task, rather than on more demanding, deliberative reasoning. In some cases, intuitive or so-called "heuristic" thinking might cue the exact same response that would follow from more deliberate reasoning. In these cases, heuristic thinking will be useful. It is faster and less demanding than deliberate thinking, allowing us to allocate our scarce cognitive resources to other tasks. However, heuristics can also cue responses that conflict with more logical or probabilistic principles and bias our reasoning (Evans, 2003, 2010; Kahneman, 2011; Stanovich & West, 2000). This is what we refer to when mentioning "heuristic bias" in what follows. A central question is whether people know that they are biased and detect that their heuristic conclusions are not logically warranted. In other words, where do such biased responses originate from and **are we aware that we are biased** when thinking?

Two different points of view can be distinguished here, calling upon either a failure to detect the conflict between the heuristic response and logical, normative[i] considerations, or an inhibition failure *per se*. The first one postulates that the widespread heuristic bias can be attributed to a failure in monitoring our intuition and detecting the conflict (Kahneman & Frederick, 2002). Alternatively, some authors have argued that people detect the conflict but do not always manage to inhibit and discard the tempting beliefs (Denes-Raj & Epstein, 1994; Houdé, 2007; Sloman, 1996). In that case, the conflict is successfully detected but people still "behave against their better judgment" (Denes-Raj & Epstein,1994). Recent empirical studies have tried to decide between these views and started to measure the efficiency of the conflict detection process during reasoning (e.g., see De Neys, 2012, 2014 for a review). These studies typically present classic problems of the reasoning and decision-making field (e.g., base-rate neglect tasks, ratio bias tasks, conjunction fallacy, belief bias syllogisms, etc.) in which an intuitively cued heuristic response conflicts with logical considerations. These "conflict" problems are presented along with "no-conflict" problems, for which the intuitively cued heuristic response and logical responses are congruent. For example, a no-conflict version of the ratio bias problem that was introduced above could state that the small tray contains 1 red marble out of a total of 10, and the large tray contains 11 red marbles out of a total of 100 marbles. In that case, both logical and heuristic responses converge. That is, the large tray has both the highest absolute number of red marbles and the highest probability of winning (e.g. 11%). The principle behind this type of design is that if people do not know the relevant logical norms, or if they do not use them for monitoring conflicts, the two versions of the problem should be processed the same way. Accordingly, a range of measures has been introduced to examine whether or not people process the conflict and no-conflict problems differently (e.g., see De Neys, 2012 for a review). Results typically show that reasoners, even biased ones, are sensitive to conflict. This is perceptible through significant variations of response times (Bonner & Newell, 2010; De Neys & Glumicic, 2008; Pennycook, Fugelsang, & Koehler, 2012; Stupple & Ball, 2008; Villejoubert, 2009), confidence ratings (De Neys et al., 2011, 2013; De Neys & Feremans, 2013; Stupple, Ball & Ellis, 2013; Thompson & Johnson, 2014), skin conductance (De Neys, Moyens, & Vansteenwegen, 2010), functional Magnetic Resonance Imaging (De Neys, Vartanian, & Goel, 2008) and Electro-EncephaloGraphy signals (De Neys, Novitskiy, Ramautar, & Wagemans, 2010), eye- and gaze- tracking measures (Ball, **Phillips**, Wade & **Quayle**, 2006; De Neys & Glumicic, 2008; Morsanyi & Handley, 2012), and memory scores (De Neys & Glumicic, 2008; De Neys & Franssens, 2009; Franssens & De Neys, 2009). For instance, increased response times (Bonner & Newell, 2010; De Neys & Glumicic, 2008; Pennycook, Fugelsang, & Koehler, 2012; Stupple & Ball, 2008; Villejoubert, 2009) and decreased confidence rating (De Neys et al., 2011, 2013; De Neys & Feremans, 2013; Stupple, Ball & Ellis, 2013; Thompson & Johnson, 2014) are typically observed when participants solve conflict versions of a problem, as compared to no-conflict versions.

It should be noted, however, that the conflict detection studies are not uncontroversial (e.g., De Neys, 2012; 2014; Klauer & Singmann, 2013; Pennycook et al., 2012; Singmann, Klauer, & Kellen, 2014). Findings have far-stretched implications for reasoning and decision making theories (e.g., De Neys & Glumicic, 2008; De Neys, 2012). For example, one implication of the successful nature of conflict detection is that reasoners can easily and **effortlessly access logical knowledge**. Without such easy, fast access to logical knowledge it is hard to account for how biased reasoners can be sensitive to logical violations (De Neys, 2012). However, as noted by Singman et al., the existence of such an intuitive and fast type of logical intuitions clashes with key assumptions of traditional theories (i.e. the assumption that logical thinking is necessarily slow and effortful). Singman et al. have then rightly underlined that, before making such theoretical revisions based on published detection work, we should make sure that the empirical findings are not driven by experimental confounds. Even proponents of the idea that conflict detection is successful have underlined that the research is still its infancy (De Neys, 2012). In sum, further validation and generalization of the conflict detection findings is paramount (De Neys, 2012; 2014; Klauer & Singmann, 2013; Pennycook et al., 2012; Singmann, Klauer, & Kellen, 2014). The present study centers on this issue, thereby focusing on the ratio bias task. It has been argued that this task provides ample evidence of successful conflict detection and it has been used as a key example of the robustness of the conflict detection findings in recent theorizing (e.g., De Neys & Bonnefon, 2013; De Neys, 2012). However, this claim relies on a single study, showing longer latencies for conflict than for no-conflict problems (Bonner & Newell, 2010). Although highly relevant, it is important to validate these findings.

To this end, three issues will be specifically addressed in this study. First, we focus on one of the most popular measures to investigate conflict detection efficiency, that is the response confidence measure (De Neys et al., 2011, 2013; De Neys & Feremans, 2013). We will explore whether this critical measure also points to successful ratio bias detection. This will allow us to test whether Bonner & Newell's (2010) findings can be validated with a different conflict detection measure. Second, it has been suggested that conflict detection findings might suffer from a training confound. When testing participants with multiple conflict and no-conflict problems, the conflict detection might be cued by the iterative nature of the task (Kahneman, 2000; De Neys et al., 2011). Reducing the number of presented problems is not efficient to sidestep this problem and it has been argued that the most critical test to address this point is a between-subject test, for which only the first problem that is presented is analyzed (Kahneman, 2000; Stanovich & West, 2008). Of note, in Bonner and Newell's (2010) work, such a single item analysis was not performed. Finally, it has been suggested that conflict detection studies also need to start looking for potential individual differences in conflict detection among biased reasoners (De Neys, 2014; De Neys & Bonnefon, 2013; Mata et al., 2014; see also Stanovich, West, & Toplak, 2011). So far, most published work has focused on group level analyses and stated that the "modal" or "average" biased reasoner (i.e. the typical biased reasoner, showing the predominant reasoning profile) shows conflict detection. One interesting exception is the work of Stupple et al. (2011) who found that even the most biased reasoners showed sensitivity to belief-logic conflict in a syllogistic reasoning task, as evidenced by their response times. However, the most logical responders were also the ones showing signs of a more pronounced sensitivity. It cannot be ruled out that such inter-individual differences may arise in the ratio bias task as well. For instance, are there subgroups of biased reasoners who potentially do not detect the conflict? Is there any relationship between conflict detection and performance? In an attempt **to help identify possible individual differences** in conflict detection efficiency during the ratio bias task, we made sure to test a large sample and present complementary analyses at the participant level.

METHODS

Participants

A total of 317 undergraduates who were taking psychology courses at the University of Paris Descartes and Caen Basse-Normandie (France), participated in the study. Fourteen participants were discarded because of missing data, leading to a final sample of 303 participants, characterized by a mean age of 21.68 years (SEM = .25), a mean education of 14.51 years (SEM = .04) and a 17.82% proportion of males.

Material

The problems used in this experiment were based on the ratio-bias task as previously published (see for instance Bonner & Newell, 2010; Denes-Raj & Epstein, 1994; Rudski & Volksdorf, 2002; Stanovich & West, 2008). Participants had to solve a total of ten problems. The tray pairs were developed on the basis of percentage ranges used in previous research. The small tray contained 1 or 3 red marbles out of a total 10 (i.e. 10% - 30% Rudski & Volksdorf, 2002). The large tray had a total of 100 marbles and the proportion of red marbles differed from the small tray by a value of 1% or 6%. Four of the problems were conflict problems and 4 were no-conflict problems as described above, meaning that the participants saw each combination of ratios only once. This resulted in 8 problems altogether: 4 conflict problems (e.g., 1/10 vs. 9/100, 1/10 vs. 4/100, 3/10 vs. 29/100, 3/10 vs. 24/100) and 4 no-conflict problems (e.g., 1/10 vs. 11/100, 1/10 vs. 16/100, 3/10 vs. 31/100, 3/10 vs. 36/100). Figure 1 presents an illustration of the tray setup.

Each problem was presented on a separate page in a booklet, with the marbles being ordered and the proportions given in absolute written numbers. The left most tray was always the small one, and the rightmost tray the large one. On the first page of the booklet the participants received the following general instructions:

"In this booklet, you will be presented with problems. For each of them, you will see a black frame and two trays of white and red marbles. In the example below [Figure 1], the small tray includes 10 marbles among which 4 are red, and the large tray includes 100 marbles among which 49 are red. Imagine that you win a prize if you succeed in drawing a red marble from one of the two trays (you cannot peek when drawing and each set has been shaken up). Hence, you need to figure out from which tray it's more likely that you win (drawing a red marble). For each problem, the number of white and red marbles will differ and you will have to mark the tray from which you would prefer to draw from (i.e. the small or the large tray) in order to win the prize."

After participants had solved the first problem, they found the following instructions on the next page:

"Please indicate how confident you are that the answer you just gave was the correct one, by writing down a number between 0 [totally unconfident] and 100 [totally confident]:

I'm __________ % confident that my answer is correct".

The remaining seven problems were presented with the confidence rating directly below.

After the conflict and no-conflict items, we also presented two control items that were designed to identify inattentive or randomly answering participants. One item presented a choice between a large tray and a small tray with 0 red marbles. Hence, picking the small tray would indicate a genuine failure to comply with task instructions or a tendency to favor the small leftmost tray. The second control item presented a blatantly false statement (e.g., *"Denver is the capital of the United States of America"*) and asked participants to indicate how confident they were it was correct. We decided *a priori* that participants who would give the same (non-zero) rating on this problem and the experimental problems would be discarded. However, results showed that participants were paying basic attention to the task and none of them needed to be excluded based on these criteria. While the first control item was successfully completed by all the participants with an average confidence rating of 91.56% (SEM = 1.22%), the average confidence rating for the second item was of 2.66% (SEM = .90%). This indicates that general inattentiveness or random response tendencies are not confounding the present data.

Procedure

Participants were tested at the same time during a regular course at the University. The 8 ratio-bias problems were presented in one of 8 pseudo-random orders, each version starting with one of the 8 problems. We made sure that half of the presented booklets started with a conflict problem, while the other half started with a no-conflict problem.

RESULTS

Accuracy: On average, 83.33% of the conflict problems were solved correctly (SEM = 1.53%), which replicates rates previously reported (Bonner & Newell, 2010). Accuracy on our no-conflict problems was somewhat lower than previously reported (78.55%, SEM = 1.77%). An Analysis of Variance (ANOVA) with one within factor (Type of problems: conflict or no-conflict) indicated that this difference failed to reach significance ($F_{(1,302)}$ = 3.76, p =.053, η^2_p =.012). For completeness, note that the slightly higher error rate on no-conflict problems is of little concern for our present purposes, since the key interest here lies in the difference between confidence findings for correctly solved no-conflict and incorrect conflict items. As reported, our control items established that general attentional confounds are unlikely to affect the present data.

Response confidence: Overall, our key interest lies in the confidence ratings. Our main question was to check whether confidence measures also point to successful ratio bias detection. If so, we should replicate previous confidence findings reported with other classic tasks (De Neys et al., 2011, 2013; De Neys & Feremans, 2013): that is, biased reasoners who fail to solve the conflict problems correctly should show a lower confidence after solving conflict problems than after (successfully) solving the no-conflict problems. To this end, an average individual response confidence score was computed for the incorrect conflict and correctly solved no-conflict problems. We submitted those scores to an ANOVA in which the Type of problem served as a dependent variable (incorrect conflict vs. correct no-conflict). We found a significant effect of the Type of problem ($F_{(1,109)}$ =18.80, p =.00003, η^2_p =.15). As Figure 2 shows, confidence ratings were lower for the incorrect conflict problems (M = 55.80%, SEM = 2.36%) than for the correct no-conflict problems (M = 61.20%, SEM = 2.22%). Altogether, this confirms people's sensitivity to conflict between the given heuristic response and the logical one. For completeness, we repeated the above analysis, for correctly solved conflict problems too. In line with previous confidence studies, no significant difference was observed when comparing the correct conflict (M = 74.90%, SEM = 1.38%) with the correct no-conflict (M = 74.39%, SEM = 1.39%) problems, $F_{(1,271)}$ = .53, p = .47, η^2_p = .002. Hence, not surprisingly, when people give a correct answer they also seem to know that their answer is correct.

As one reviewer noted, one might also want to test whether confidence for incorrectly solved no-conflict problems is lower than for correctly solved no-conflict problems. In theory, this allows us to verify whether errors *per se* (i.e. errors observed irrespective of the presence of conflict) result in a decreased confidence or whether a decrease in confidence is pertaining to conflict items solely. To this end, we ran an additional within-subject ANOVA in which the Type of problem served as a dependent variable (incorrect no-conflict vs. correct no-conflict).

Results showed that confidence **ratings were similar for correct no-conflict** (M = 63.93%, SEM = 2.08%) **and for** incorrect no-conflict items (M = 60.54%, SEM = 2.10%), $F_{(1,117)}$ =3.09, p =.082, η^2_p =.026. **Though this might indicate that errors *per se* do not seem to result in lower confidence rating, further investigations are needed as the effect borders significance. Provided that future explorations confirm the present finding, this will be** further evidence for the postulate of lowered confidence findings resulting from conflict detection. However, as we noted, the error rates on the no-conflict problems in the current study were somewhat higher than typically reported. It appears difficult to unequivocally interpret such errors, as people who err on no-conflict trials give an erroneous response although there is no cued conflict between logical considerations and heuristic intuitions. In the absence of any clear indication as to what is driving these errors, one might want to interpret the confidence findings for incorrectly solved no-conflict problems with some caution.

First problems response confidence: To address the impact of possible learning effects, we also ran additional between-subjects ANOVAs in which we included only the confidence rating of the first problem that each participant solved (recall that this was a conflict problem for half of the participants and a no-conflict problem for the remaining half). As Figure 3 shows, results replicated the main finding of the overall analysis. In particular, participants who gave a heuristic response on the conflict problem were significantly less confident about their response (M = 55.90%, SEM = 7.31%) than participants who correctly solved a no-conflict problem (M = 72.46%, SEM = 2.66%), $F_{(1,\ 114)}$ = 6.29, p = .013, η^2_p = .05. This significant difference was not found when comparing participants who correctly solved the first conflict problem (M = 74.80%, SEM = 1.99%) and participants who correctly solved a no-conflict problem (M = 72.46%, SEM = 2.66%), $F_{(1,223)}$ = .52, p = .47, η^2_p = .002. This establishes that the observed confidence decrease of the biased reasoners on the conflict problems does not result from a learning confound.

Subgroups of biased reasoners: Finally, we aimed at exploring potential individual differences in conflict detection efficiency. Do all the biased reasoners show sensitivity to conflict or are there subgroups that do not show this effect? To this end we looked at each individual reasoner's average confidence on the incorrect conflict and (correctly solved) no-conflict problems. Next, we tallied which percentage of reasoners showed a confidence decrease when rating correct no-conflict as compared to incorrect conflict problems. Results can be seen in Figure 4. A total of 110 participants (e.g. 36.3% of the whole sample) showed a ratio bias on at least one conflict problem. The majority of these biased participants (i.e., 56%) indeed marked a decrease in the response confidence for the incorrect conflict problems as compared to their rating for correct no-conflict problems. Mean confidence decrease in this group was of 23.50% (SEM = 2.39%). However, there were also 25% of biased reasoners who gave a higher confidence response rating (Mean increase = 17.75%, SEM = 3.68%) and 19% of biased reasoners who gave the same rating for both types of problems (Mean rating = 74.76%, SEM = 4.67%). Hence, although most biased reasoners showed the effect, there are clearly also smaller subgroups of reasoners who did not show conflict sensitivity as measured by their confidence ratings.

Taking it further, we wondered whether these subsets of biased reasoners showing no sensitivity to conflict were providing more incorrect responses than the other biased participants who seem to detect the conflict. In other words, does the amplitude of conflict detection predict the performance? The average accuracy for conflict items was first computed for each biased subgroup. Next, for the "increase" and "decrease in confidence" subgroups, the individual absolute differences in confidence rating between incorrect conflict and correct no-conflict items were calculated (i.e. the individual sizes of the conflict detection effect were obtained). Finally, these absolute confidence differences were correlated to the corresponding individual average accuracy for conflict items using Pearson's correlations. For the sake of completeness, the exact same approach was also applied for the whole biased group. Fully detailed results are available in Table 1.

As Table 1 indicates, it is interesting to note that for the subgroup of "decrease" reasoners we obtain a significant positive correlation between the size of the confidence decrease and accuracy on conflict problems ($r = .36$, $p = .004$). This seems to suggest that when participants are sensitive to conflict (i.e. when they show a decrease in their confidence rating for incorrect conflict items), the size of the detection effect might predict the performance. Although most biased reasoners will be sensitive to conflict as described above, some might be slightly more sensitive, have better detection, and thereby also show (relatively) less bias (higher accuracy). However, if the amplitude of conflict detection was really to predict performance, the groups who do not show the confidence effect (i.e. the "increase" and "equal" groups) should obviously be least accurate. It seems hard to reconcile this interpretation with the observed higher accuracies in these two subgroups or the absence of association between confidence and accuracy for the whole group (Table 1). Although the results are suggestive, the small amount of observations on which they are based combined with the *post hoc* nature of this exploratory analysis imply that we need to be cautious here. It feels reasonable to wait for more research before making any strong claims about the relation between individual conflict detection size and accuracy.

DISCUSSION

The present results validate previous empirical work that pointed to successful conflict detection in the ratio bias task (Bonner & Newell, 2010). Our measures of response confidence showed that when giving a heuristic response, most participants were experiencing a decrease of confidence in the correctness of their answer. This directly establishes that most biased reasoners show sensitivity to their errors and detect that their heuristic answer is questionable. In addition, we observed that this pattern was present for the first problem and could not be attributed to a cueing confound. Lastly we also observed that although the majority of the participants were showing sensitivity to conflict detection, there were substantial individual differences. These findings add to the previously published evidence for successful conflict detection during biased thinking (De Neys et al., 2011, 2013, 2008; De Neys & Feremans, 2013; De Neys, Moyens, et al., 2010; Franssens & De Neys, 2009; Stupple & Ball, 2008; Thompson & Johnson, 2014; Villejoubert, 2009). As we mentioned earlier in the Introduction, these conflict detection studies have not been uncontroversial and it has been stressed that further generalization and validation is paramount (De Neys, 2012; 2014; Klauer & Singmann, 2013; Pennycook et al., 2012; Singmann, Klauer, & Kellen, In Press). The present results present a critical contribution in this respect.

Results for our control items and the analyses of response confidence for the first problem imply that neither a lack of attention nor a training confound can be invoked to explain our results. In addition, our final analysis confirms that at the individual level the predominant, modal biased reasoner profile indeed shows evidence for successful detection between the produced heuristic and the logical response: over half of biased reasoners (56%) showed the predicted lower response confidence trend, which amplitude was found to be significantly associated with the overall performance on conflict problems. Conversely, there was a subset of biased reasoners for whom we did not observe the conflict detection nor the performance-related effect. Clearly, there is some noise related to measurements of confidence ratings at the individual level but the present findings at least indicate that the issue of potential inter-individual differences in conflict detection efficiency needs to be taken seriously (De Neys & Bonnefon, 2013; De Neys, 2014; Stanovich, 2010; Stanovich, Toplak, & West, 2008). Although most biased reasoners might detect conflict, not all of them will.

Recently, De Neys and Bonnefon (2013) discussed potential implications of the conflict detection findings for our view of individual differences in bias susceptibility. They proposed that classic positions on the nature of bias can be ordered on a timeline from early to late in the reasoning process (see also Stanovich et al., 2008, for a related idea). According to De Neys and Bonnefon, in addition to considering "why" individual differences arise, we also need to consider "when" they occur. The basic idea is that according to some bias views, biased and unbiased reasoners will diverge earlier than others. For example, the so-called "Storage Failure view" - which postulates that biased reasoner have not stored the necessary logical knowledge to solve conflict problems - implies that biased reasoners are bound to take a different cognitive route from the onset of the reasoning process. The "Monitoring (or detection) Failure view" posits a somewhat later divergence point at the time of detection. Finally, the so-called "Inhibition Failure view" – which entails that all reasoners detect the conflict between cued heuristic and logical considerations but subsequently diverge in the final conflict resolution stage - implies a late divergence point. De Neys and Bonnefon argued that the empirical evidence for successful conflict detection indeed supports the latter idea of biased and unbiased reasoners diverging late in the reasoning process. The present individual difference findings imply that although this might hold true for the modal biased reasoner, there are subgroups of reasoners who will not detect conflict and show an early divergence. Note that in and by itself, the present finding does not allow us to determine whether this early divergence results from a monitoring failure *per se* or rather from a storage failure. That is, it might be that this subgroup of reasoners fails to detect conflict because they are not monitoring the conflict or because they simply do not know the relevant logical ratio principle.

To better characterize the nature of such a fine-grained individual divergence, future investigations will have to integrate additional measures of each of the elementary reasoning components (i.e., storage, monitoring, inhibition). In addition, it will be paramount to look for potentially mediating individual difference factors that might allow us to predict the efficiency of conflict detection at the individual level. For example, it might be that the subgroup of reasoners who do not show the confidence "decrease" (or show only a weak effect) in the ratio bias task will be those lowest in numeracy (e.g. Cokely et al., 2012; Liberali et al., 2012), cognitive capacity (e.g., Stanovich & West, 2000), showing specific thinking dispositions (e.g., Stanovich & West, 2008; **Pennycook et al., 2014**), or all of these. Future studies might adopt an exhaustive test battery to explore such associations. Similarly, analyses of accuracy and confidence rating over (task) time might help disentangling individual differences due to a storage from a monitoring failure. Ball (2013) demonstrated that participants who repeatedly complete belief bias problems during different sessions improve in their logical accuracy, even without feedback as to the accuracy of their answers. One can imagine extending this from an inter- to an intra-session approach and look at the variations in accuracy as a function of changes in conflict detection. If biased reasoners do not know the relevant logical principles, they should not show any sign of conflict detection (e.g. no decrease in confidence rating) and their performance should not improve over time. In sum, further explorations of differences in conflict detection at the individual level should greatly benefit from additional measures and analyses as described above.

Altogether, we believe that the present study brings substantial elements for generalization and validation of conflict detection studies. We replicated previous findings with a new measure (i.e. with confidence rating measures), while providing the first exploration of individual differences in the ratio bias task. Though our experimental design allowed for a control over a set of confounds (e.g. lack of attention, misunderstanding of the instructions, training effect), our study also faces some limitations. First, the ratio bias task was paper-based, preventing from recording response latencies as a complement to confidence rating (e.g., see Stupple et al. 2011 for an examination of individual differences in belief bias). Ideally, direct comparisons between the results obtained for both types of measures would further validate the use of confidence decrease as a sign of conflict detection. Second, our claim about the existence of differences at the individual level is somewhat limited by the lack of specific predictive mediating variables to look at (e.g. numeracy skills, cognitive profile, thinking disposition) and complementary intra-session analyses. Future work will need to take into account these elements to gain further insight into the nature and extent of individual differences in conflict detection efficiency.

REFERENCES

Ball, L. J. (2013). Microgenetic evidence for the beneficial effects of feedback and practice on belief bias. *Journal of Cognitive Psychology, 25*, 183-191. doi:10.1080/20445911.2013.765856

Ball, L. J., **Phillips**, P., Wade, C. N., & Quayle, J. D. (2006). Effects of belief and logic on syllogistic reasoning: Eye-movement evidence for selective processing models. *Experimental Psychology, 53*, 77–86.doi: 10.1027/1618-3169.53.1.77

Bonner, C., & Newell, B. R. (2010). In conflict with ourselves? An investigation of heuristic and analytic processes in decision making. *Memory & Cognition, 38*, 186–196. doi:10.3758/MC.38.2.186

Cokely, E.T., Galesic, M., Schultz, e., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: the Berlin Numeracy Test. *Judgment and Decision Making, 7*, 25-47.

De Neys, W. (2012). Bias and Conflict: A Case for Logical Intuitions. Perspectives on *Psychological Science, 7*, 28–38. doi:10.1177/1745691611429354

De Neys, W. (2014). Conflict detection, dual processes, and logical intuitions: Some clarifications. *Thinking & Reasoning, 20*, 169–187. doi:10.1080/13546783.2013.854725

De Neys, W., & Bonnefon, J.-F. (2013). The "whys" and "whens" of individual differences in thinking biases. *Trends in Cognitive Sciences, 17*, 172–178. doi:10.1016/j.tics.2013.02.001

De Neys, W., & Franssens, S. (2009). Belief inhibition during thinking: Not always winning but at least taking part. *Cognition, 113*, 45–61. doi: 10.1016/j.cognition.2009.07.009

De Neys, W., Cromheeke, S., & Osman, M. (2011). Biased but in doubt: conflict and decision confidence. *PloS One, 6*, e15954. doi:10.1371/journal.pone.0015954

De Neys, W., & Feremans, V. (2013). Development of heuristic bias detection in elementary school. *Developmental Psychology, 49*, 258–269. doi:10.1037/a0028320

De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition, 106*, 1248–1299. doi:10.1016/j.cognition.2007.06.002

De Neys, W., Moyens, E., & Vansteenwegen, D. (2010). Feeling we're biased: autonomic arousal and reasoning conflict. C*ognitive, Affective & Behavioral Neuroscience, 10*, 208–216. doi:10.3758/CABN.10.2.208

De Neys, W., Rossi, S., & Houdé, O. (2013). Bats, balls, and substitution sensitivity: cognitive misers are no happy fools. *Psychonomic Bulletin & Review, 20*, 269–273. doi:10.3758/s13423-013-0384-5

De Neys, W., Vartanian, O., & Goel, V. (2008). Smarter than we think: when our brains detect that we are biased. *Psychological Science, 19*, 483–489. doi:10.1111/j.1467-9280.2008.02113.x

Denes-Raj, V., & Epstein, S. (1994). Conflict between intuitive and rational processing: when people behave against their better judgment. *Journal of Personality and Social Psychology, 66*, 819–829. doi: 10.1037/0022-3514.66.5.819

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *The American Psychologist, 49*, 709–724. doi: 10.1037/0003-066X.49.8.709

Evans, J. S. B. T. (2003). In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences, 7*, 454–459. doi:10.1016/j.tics.2003.08.012

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59*, 255–278. doi:10.1146/annurev.psych.59.103006.093629

Evans, J. S. B. T. (2010). Intuition and Reasoning: A Dual-Process Perspective. *Psychological Inquiry, 21*, 313–326. doi:10.1080/1047840X.2010.521057

Evans, J. S. B. T., & Over, D. E. (1996). *Rationality and Reasoning* (p. 179).

Franssens, S., & De Neys, W. (2009). The effortless nature of conflict detection during thinking. *Thinking & Reasoning, 15*, 105–128. doi:10.1080/13546780802711185

Houdé, O. (2007). First insights on "neuropedagogy of reasoning." *Thinking & Reasoning, 13*, 81–89. doi:10.1080/13546780500450599

Kahneman, D. (2000) A psychological point of view: Violations of rational rules as a diagnostic of mental processes. *Behavioral and Brain Sciences 23*, 681–683. doi:10.1017/S0140525X00403432

Kahneman, D. (2011). *Thinking, Fast and Slow.* Macmillan.

Kahneman, D., & Frederick, S. (2002). Representativeness Revisited : Attribute Substitution in Intuitive Judgment. In *Heuristics and biases: the psychology of intuitive judgment* (pp. 49–81). Cambridge University Press.

Kahneman, D., & Frederick, S. (2005). A Model of Heuristic Judgment. In The Cambridge Handbook of Thinking and Reasoning (pp. 267–273). Cambridge: Cambridge University Press.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237–251. doi: 10.1037/h0034747

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*, 1265–1273. doi:10.1037/a0030530

Liberali, J.M., Reyna, V.F., Furlan, S., Stein, L.M., & Pardo, S.T. (2012). Individual Differences in Numeracy and Cognitive Reflection, with Implications for Biases and Fallacies in Probability Judgment. *Journal of Behavioral Decision Making, 25,* 361-381. doi: 10.1002/bdm.752

Mata, A., Schubert, A.L., Ferreira, M.B. (2014). The role of language comprehension in reasoning: How "good-enough" representations induce biases. *Cognition, 133*, 457-463. doi: 10.1016/j.cognition.2014.07.011

Morsanyi, K., & Handley, S. (2012). Does thinking make you biased? The case of the engineers and lawyer problem. *Proceedings of the Annual Meeting of the Cognitive Science society, 34,* 2049–2054.

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2012). Are we good at detecting conflict during reasoning? *Cognition, 124*, 101–106. doi:10.1016/j.cognition.2012.04.004

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*, 1-10. doi: 10.3758/s13421-013-0340-7.

Rudski, J., & Volksdorf, J. (2002). Pictorial versus textual information and the ratio-bias effect. *Perceptual and Motor Skills, 95*, 547–554. doi: 10.2466/pms.2002.95.2.547

Singmann, H., Klauer, K. C., & Kellen, D. (In Press). Intuitive Logic Revisited: New Data and a Bayesian Mixed Model Meta-Analysis. *PloS One.* doi: 10.1371/journal.pone.0094223

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*, 3–22. doi: 10.1037/0033-2909.119.1.3

Stanovich, K. (2010). *Rationality and the Reflective Mind* (p. 344). Oxford University Press, USA.

Stanovich, K. E., Toplak, M. E., & West, R. F. (2008). The development of rational thought: a taxonomy of heuristics and biases. *Advances in Child Development and Behavior,36,* 251–285. doi: 10.1016/S0065-2407(08)00006-2

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: implications for the rationality debate? *The Behavioral and Brain Sciences, 23*, 645–665. doi: 0140-525X/00

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology, 94*, 672–695. doi:10.1037/0022-3514.94.4.672

Stanovich, K. E., West, R. F., & Toplak, M. E. (2011). The complexity of developmental predictions from dual process models. *Developmental Review, 31*, 103–118. doi:10.1016/j.dr.2011.07.003

Stupple, E. J. N., & Ball, L. J. (2008). Belief–logic conflict resolution in syllogistic reasoning: Inspection-time evidence for a parallel-process model. *Thinking & Reasoning, 14*, 168–181. doi:10.1080/13546780701739782

Stupple, E. J., Ball, L. J., Evans, J. S. B., & Kamal-Smith, E. (2011). When logic and belief collide: Individual differences in reasoning times support a selective processing model. *Journal of Cognitive Psychology*, *23*, 931-941. doi: 10.1080/20445911.2011.589381

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,* 54-77. doi: 10.1080/13546783.2012.735622

Thompson, V. A., & Johnson, S. C. (2014). Conflict, metacognition, and analytic thinking. *Thinking & Reasoning, 20*, 215–244. doi:10.1080/13546783.2013.869763

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*, 2980–2985.

Table 1: Average confidence rating, size of the conflict detection effect, accuracy (± SEM) and Pearson's correlations as a function of the subgroup of biased reasoners.

	Subgroup Decrease (n=62)	Subgroup Increase (n=27)	Subgroup Equal (n=21)	Whole Biased group (N=110)
Average confidence: *No-conflict correct (%)*	60.91 (± 2.75)	50.54 (± 4.38)	75.71 (± 4.82)	61.19 (± 2.22)
Average confidence: *Conflict incorrect (%)*	47.80 (± 2.80)	58.67 (± 4.60)	75.71 (± 4.82)	55.80 (± 2.36)
Average size of the conflict detection effect *(absolute value)*	13.11 (± 1.48)	8.14 (± 1.15)	-	9.38 (± 1.00)
Average accuracy: *No-conflict items (%)*	82.66 (± 3.12)	77.78 (± 5.05)	73.81 (± 5.84)	79.77 (± 2.42)
Average accuracy: *Conflict items (%)*	50 (± 3.4)	59.26 (± 4.03)	60.71 (± 4.75)	54.32 (± 2.36)
Pearson's Correlation *r (p)*	**.36 (.004)***	**.17 (.38)**	**-**	**.18 (.065)**

* Significant at p <.05

Figure 1: Illustration of the example problem presented on the first page of the ratio bias booklets.

Figure 2: Average response confidence for each type of problem.

Average individual response confidence for correct and incorrect conflict, as well as incorrect no-conflict problems. Error bars are standard errors.

Figure 3: Average response confidence for the first-presented problem.

Average individual response confidence for the correct and incorrect conflict, as well as incorrect no-conflict problems. Error bars are standard errors.

Figure 4: Proportion of biased reasoners as a function of changes in the response confidence for incorrect conflict vs. correct no-conflict problems.

[i] For clarity, 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 (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 labelling.

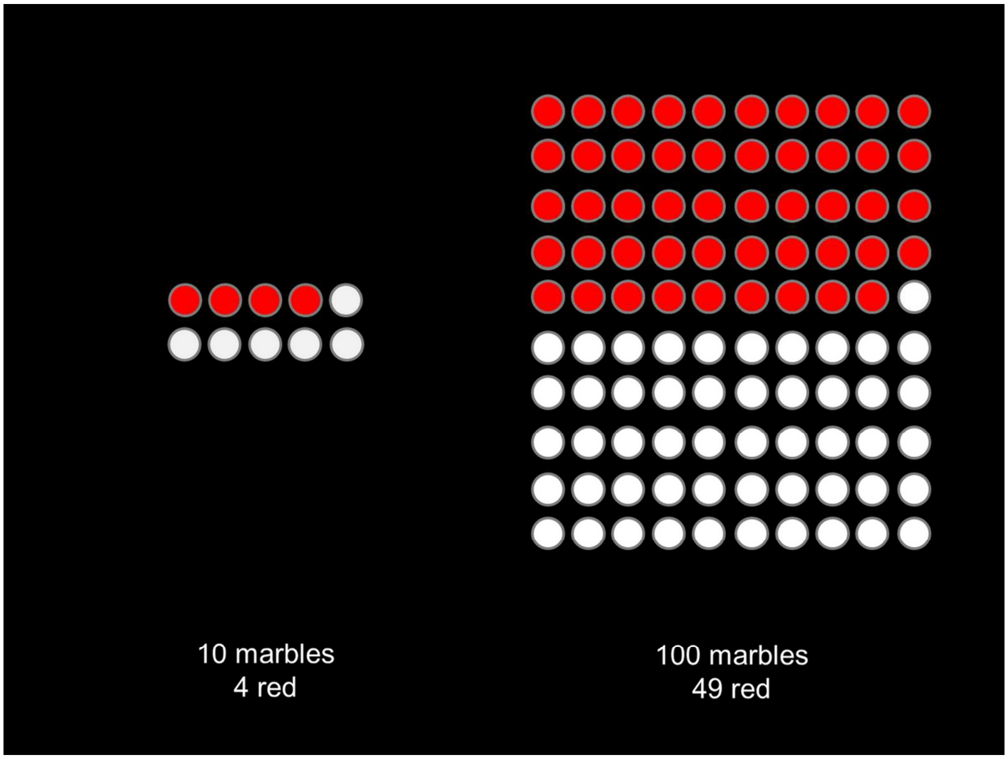
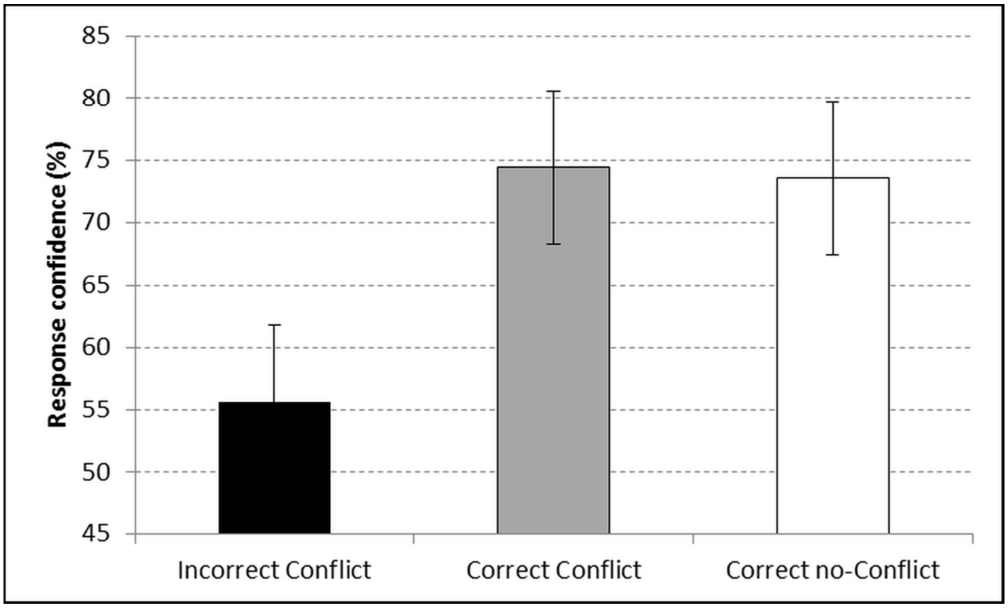
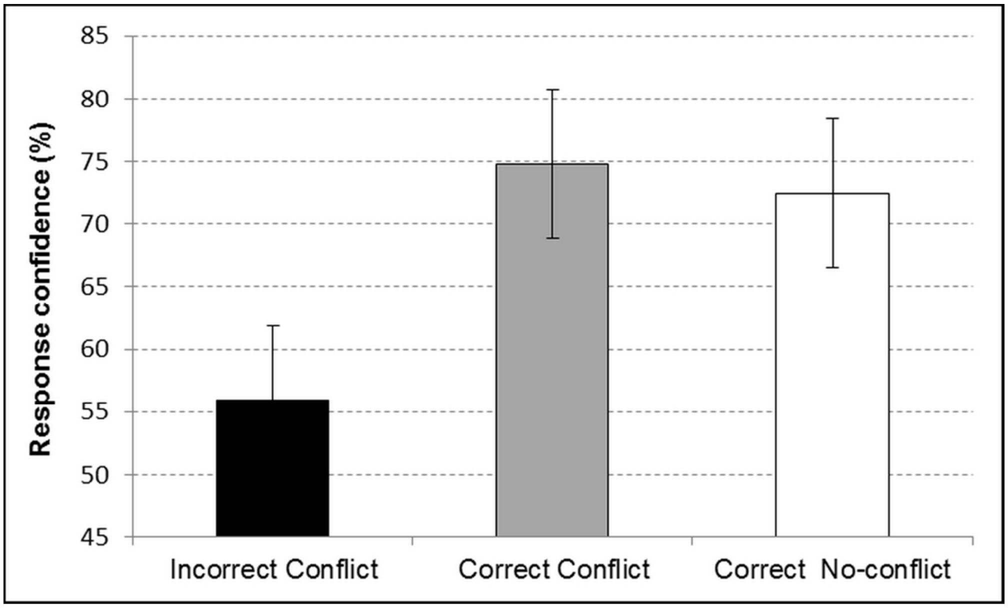


Illustration of the example problem presented on the first page of the ratio bias booklets.
119x90mm (300 x 300 DPI)



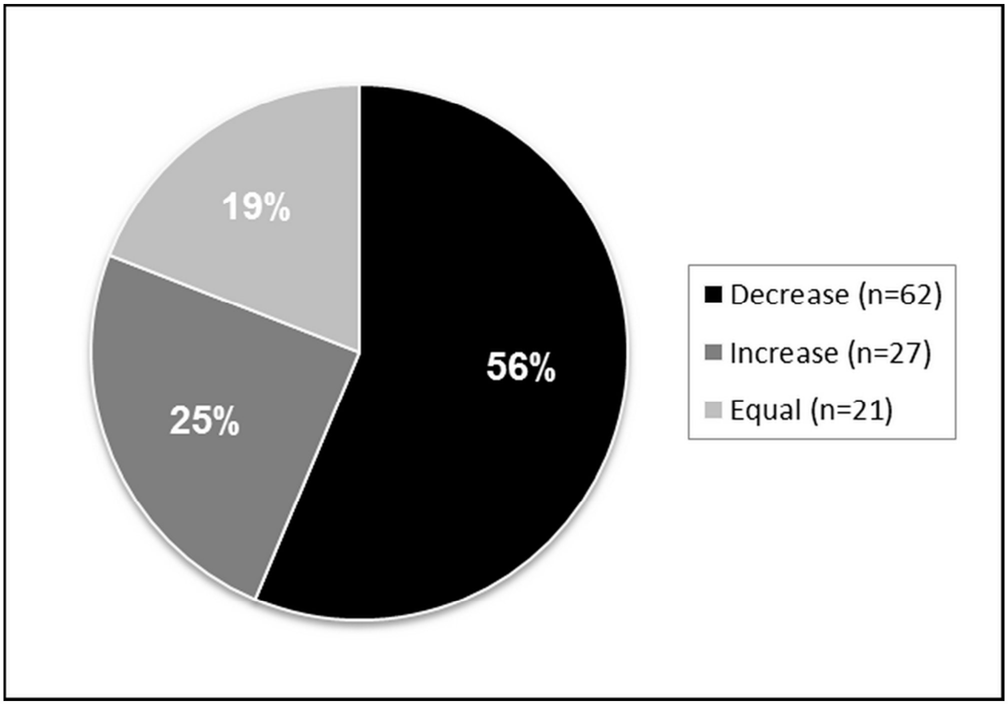
Average response confidence for each type of problem.
Average individual response confidence for correct and incorrect conflict, as well as incorrect no-conflict problems. Error bars are standard errors.

99x60mm (300 x 300 DPI)



Average response confidence for the first-presented problem.
Average individual response confidence for the correct and incorrect conflict, as well as incorrect no-conflict problems. Error bars are standard errors.

99x60mm (300 x 300 DPI)



Proportion of biased reasoners as a function of changes in the response confidence for incorrect conflict vs. correct no-conflict problems.
99x69mm (300 x 300 DPI)