System 2 and cognitive transparency: Deliberation helps to justify sound intuitions during reasoning

Nicolas Beauvais[1], Aikaterini Voudouri[1], Esther Boissin[2], Wim De Neys[1]

[1] Université Paris Cité, LaPsyDÉ, CNRS, Paris, France

[2] Cornell University, Department of Psychology, Ithaca, USA

Corresponding author:

Nicolas Beauvais

nicolas1beauvais@gmail.com

LaPsyDE, UMR CNRS 8240, 46 rue Saint-Jacques, 75005 Paris, France.

Keywords: reasoning, dual-process, justification, heuristics and biases, decision-making

Acknowledgments:

This research was funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency (HaDEA). Neither the European Union nor the granting authority can be held responsible for them. Grant Agreement no. 101120763 – TANGO.

System 2 and cognitive transparency: Deliberation helps to justify sound intuitions during reasoning

Abstract

Recent dual-process work suggests that sound responses to reasoning tasks are often generated intuitively. This has led to the hypothesis that deliberation contributes to sound response justification rather than mere response generation. However, direct evidence supporting this justificatory role of deliberation remains sparse. To test this hypothesis, three studies examined whether people could properly justify their responses to classic bias problems (base rate neglect, bat and ball, and risky choice problems) when responses were given intuitively or after deliberation. Participants provided fast, intuitive responses under time pressure in a first block, followed by unconstrained, deliberative responses to the same problems in a second block, with justifications required in both cases. Results show participants generally struggled to properly justify intuitive correct responses but tended to properly justify deliberative correct responses. These findings suggest that deliberation helps articulate reasons for intuitions, highlighting its role in the justification of decisions.

Keywords: reasoning, dual-process, justification, heuristics and biases, decision-making

1. Introduction

Dual process theories of human thinking have attracted much attention over these past decades. They propose that human reasoning and decision-making involve an interplay between two types of processes: intuitive and deliberative processes (Evans, 2008; Evans & Stanovich, 2013; Kahneman & Frederick, 2005; Sloman, 1996). While intuitive processes are fast and automatic, deliberative processes are slower and require more cognitive resources (Evans, 2019; Evans & Stanovich, 2013). Although our intuitions are useful in simplifying decision-making, they can sometimes lead to biases (Evans & Stanovich, 2013; Kahneman, 2011). A classic example illustrating these biases is the famous bat-and-ball problem, initially presented by Frederick (2005):

A bat and a ball together cost 1.10$. The bat costs 1$ more than the ball.
How much does the ball cost?

When faced with this problem most individuals respond with 10 cents, however the correct answer is 5 cents. Indeed, the total cost of the items being $1.10, if the ball costs 5 cents then the bat will cost $1.05, resulting in a total of $1.10. Although the solution to this problem can be found with a simple mathematical equation (x + 1 + x = 1.10), most people respond incorrectly when faced with it (Bourgeois-Gironde & Van Der Henst, 2009; Frederick, 2005).

Dual process theories have provided a popular explanation for this type of error. They support that when faced with decisions, individuals generally rely on their intuitions. Those intuitions are based on shortcuts and subjective beliefs (also known as heuristics) and sometimes lead to biased responses that conflict with logical and mathematical principles. Dual-process theories suggest that to overcome a biased intuitive response, reasoners need to engage in deliberation and correct their rapid intuitions (Evans & Stanovich, 2013; Kahneman, 2011; Tversky & Kahneman, 1974). However, since deliberation requires time and cognitive effort, people often avoid it once they have already arrived at an intuitive and effortless response (Kahneman, 2011) and remain biased.

Consequently, the influential corrective assumption of dual process theories assumes that deliberate reasoning typically generates logical responses by correcting erroneous heuristic intuitive responses (Evans, 2008; Kahneman & Frederick, 2005). However, recent findings show that individuals can sometimes give correct responses intuitively, without deliberation

(Bago & De Neys, 2019; De Neys, 2023; Raoelison et al., 2021). To distinguish between intuitive and deliberative processes, these studies used a two-response paradigm in which participants provide two responses to the same problem. First, participants need to give their first, intuitive hunch as fast as possible (under tight time constraints). Afterward, they can take all the time they want to deliberate on the problem and give a final answer (Thompson et al., 2011). Several studies using this method have demonstrated that correct deliberative responses are generally preceded by already correct intuitive responses (Bago & De Neys, 2019; Newman et al., 2017). These results suggest that sound reasoners' intuitions do not always require correction since they often are already correct, hence challenging the idea that correction is the sole function of deliberative reasoning. Further evidence from other paradigms also questioned the link between deliberation and correct responding (e.g., Byrd et al., 2023; Markovits et al., 2021; Szaszi et al., 2017). In sum, intuitive processes do not necessarily provide biased responses, and deliberate processes are not always necessary for sound reasoning.

However, this does not imply that people do not need to deliberate per se. Rather, deliberation may serve other roles in the reasoning process than mere correction. Previous work suggested that deliberation may also have a justificatory function, aiming to rationalize our intuitions (Bago & De Neys, 2019; De Neys, 2025; Evans, 2019; Pennycook et al., 2015). In one of their studies, Bago and De Neys (2019) asked participants to provide a justification for their answer after the intuitive and deliberate response stages. Results suggested that while most correct responses were already generated in the intuitive stage of the two-response paradigm, participants were more capable of *justifying* these responses after they had had time to deliberate. Indeed, a characteristic frequently attributed to deliberative reasoning is its "cognitive transparency": the fact that the output of deliberation comes "with an awareness of how it was derived" (Bonnefon, 2013, 2016; Evans & Stanovich, 2013). Intuitive processing typically lacks this explanatory property. As noted by Bago and De Neys (2019), the absence of introspective insight or justification is often considered a hallmark of intuitive processing and a reason why it is commonly referred to as "gut-feelings" (Marewski & Hoffrage, 2015; Mega & Volz, 2014). Therefore, it is possible that an alternative role of deliberation is to provide individuals with explicit reasons for justifying their decisions and actions, both to themselves and to others. These hypotheses resonate with the work of Mercier and Sperber (2011, 2017), who argue that the main purpose of reasoning is to provide reasons to adhere to a discourse or an action.

While these lines of work suggest that deliberation may enable justification, the justificative function of deliberative (vs intuitive) reasoning remains to be tested directly. Bago and De Neys (2019) laid out exploratory results pointing in this direction, but limited to a single problem type. Hence, testing the pattern's robustness and generality is crucial. The present study addresses this issue. We selected three classic tasks from the reasoning and decision-making literature, on which sound intuitive response generation has been demonstrated. In addition to bat-and-ball problems, we used base-rate problems (Bago & De Neys, 2017; Tversky & Kahneman, 1974) and risky-choice bets (Kahneman & Tversky, 1979, 1984; Voudouri et al., 2024). By exploring tasks with a wider array of normative principles (i.e., mathematics, probabilistic base-rate principles, or expected value calculations) and heuristics (i.e., faulty math, stereotypical associations, or loss aversion), we tried to validate whether the findings hold true across various contexts and provide a strong test of the theoretical claim that deliberative reasoning helps to generate explicit reasons and explanations, a process we refer to as justification.

In the current study, we used a two-block paradigm in which participants were presented with an initial intuitive block of problems to which they had to give their first hunch under time pressure. In a second, deliberative block, participants responded to the same problems again but without constraint and were encouraged to take all the time they wanted to deliberate on each problem. They also provided a confidence rating and an explanation for each response. We hypothesized that individuals would be more likely to come up with sound explanations for responses they gave after deliberation than for their intuitive responses.

2. Method

We ran three independent studies on the different tasks. Given the identical design and closely aligned core findings across these studies, we present the results combined for the sake of clarity and ease of presentation.

2.1. Pre-registration and data availability

The study design and hypothesis were pre-registered on the Open Science Framework for all three studies. The pre-registrations, data, material, and analysis scripts are available on OSF at https://osf.io/pm3je/.

2.2. Participants

Study 1 (Bat-and-ball problems). We recruited 100 participants on the Prolific Academic platform (www.prolific.ac). We included only native English speakers from the United Kingdom, United States, Ireland, Australia, Canada, or New Zealand. Participants' mean age was 37.5 years (SD = 15.0). There were 49 women, 49 men, and two reported "other" as a gender. Regarding education level, 36% had a high school diploma and 63% a university degree. Only 1 participant reported an education level of less than high school.

Study 2 (Base-rate neglect problems). We recruited 100 participants on Prolific Academic, also including only native English speakers from the same countries as in Study 1. Participants' mean age was 37.7 years (SD = 13.42). There were 50 women, 49 men, and one "other" gender. Among them, 26% reported a high school diploma as their higher education level, and 74% reported a university degree.

Study 3 (Risky-choice problems). We recruited 100 participants on Prolific Academic. All were native English speakers from the same countries as in Study 1. However, due to a software issue, only 99 participants completed the study. The mean age was 37.9 years (SD = 13.28). There were 49 women, 48 men, and two selected "other" as a gender. Among them, 28% reported their higher education level to be a high school diploma, and 72% reported a university degree.

2.3. Material

2.3.1. Reasoning problems

Study 1 (Bat-and-ball problems). Participants were asked to respond to four bat-and-ball problems which were variants of the original bat-and-ball problem (Frederick, 2005). The bat-and-ball problem has been extensively studied in scientific research and popular science writing, which implies that some participants may already be familiar with it (Haigh, 2016). Previous studies have shown that participants with more experience in cognitive tasks, including those who have encountered the bat-and-ball problem before, tend to perform better than naïve participants (Stieger & Reips, 2016). However, Chandler et al., (2014) showed this correlation disappeared when presenting structurally identical but content-modified versions of the problem. In this study, as in Raoelison and De Neys (2019), we thus used content-modified

versions of the original bat-and-ball problem to minimize the impact of familiarity or prior exposure on task performance. These problems had the same structure as the original problem but different content: each problem specified two types of items with different quantities (instead of prices), and participants were asked to determine the quantity of one item based on the information of the total amount and the quantity relation between the two types of items. The format is illustrated by the following example:

In a building residents have 370 dogs and cats in total.
There are 300 more dogs than cats.
How many cats are there?

- *7*
- *35*
- *70*
- *105*

In each problem, participants had to choose between four possible answers: the correct response (i.e., "5 cents" in the original bat-and-ball scenario, “35” in the previous example), the intuitive "heuristic" response (i.e., "10 cents" in the original scenario, “70” in the example), and two foil options. To determine the correct answer option, we used the form of the correct mathematical equation for solving the standard bat-and-ball problem, which is $100 + 2x = 110$. To obtain the heuristic answer option, we used the form of the mathematical equation that individuals are thought to be using in this kind of problem when giving their initial intuitive response: $100 + x = 110$ (Kahneman, 2011). Following Bago and De Neys (2019), the two foil options were determined as being always the sum of the correct and heuristics options (e.g., “15 cents” in the original bat-and-ball units) and their second greatest common factor (e.g., "1 cent" in original units). The order in which the four response options appeared was randomized for each item. However, we ensured that the order for a given problem was the same in the intuitive and deliberate response block so that it would not influence performance across blocks.

Half of the problems were presented in a classic “conflict” format in which the intuitively cued “heuristic” response conflicts with the correct answer (see above) as in the classical bat-and-ball problem. We also included control no-conflict problems, in which both the logical and the heuristic information cue the same response. This was achieved by removing the critical relational statement “more than” (De Neys et al., 2013). The no-conflict version of the previous example problem would then be:

In a building residents have 370 dogs and cats in total.
There are 300 dogs.
How many cats are there?

- *7*
- *35*
- *70*
- *105*

These no-conflict trials were used to ensure that participants paid minimal attention throughout the study. If participants refrained from random guessing, performance should be near ceiling. Overall, participants thus responded to two conflict and two no-conflict problems, in each block.

Study 2 (Base-rates neglect problems). In Study 2, participants were presented with base-rate problems in which they always received a description of the composition of a sample (e.g., "This study contains high school students and librarians"), a base rate information (e.g., "There are 5 high school students and 995 librarians"), and a description designed to cue a stereotypical association (e.g., "Person M. is loud"). The participant's task was to indicate to which group the person most likely belonged. The task instructions stressed that the person was drawn randomly from the specified sample. The problem presentation format was based on Pennycook et al.'s (2015) rapid-response paradigm. The base rates and descriptive information were presented serially and the amount of text that was presented on screen was minimized. Base-rates varied between 995/5 and 997/3. The full problem format is illustrated below:

This study contains high school students and librarians.
Person 'M' is loud.
There are 5 high school students and 995 librarians.
Is Person 'M' more likely to be:

- *A high school student*
- *A librarian*

We presented two conflict and two no-conflict problems in each block. In conflict problems, the base rate probabilities and the stereotypical information cued conflicting responses (see example above). In the no-conflict control problems, the description triggered a stereotypical trait of a member of the largest group. As in no-conflict bat-and-ball problems, these no-conflict problems should be easy to solve. If participants are paying minimal attention to the task and

refrain from random guessing, they should show very high accuracy. Note that critics of the base-rate task (Barbey & Sloman, 2007; Gigerenzer et al., 1988) have pointed out that if reasoners adopt a Bayesian approach and combine the base rate probabilities with the stereotypical description, this can lead to interpretative complications when the description is extremely diagnostic. The moderate descriptions used in this study (such as 'strong' or 'loud') as well as the extreme base-rates help to avoid this potential problem and guarantee that even a very approximate Bayesian reasoner would need to pick the response cued by the base-rates (see De Neys, 2014).

Study 3 (Risky-choice problems). In Study 3, participants were presented with bets that could result either in a gain or a loss (based on Keysar et al., 2012; and Voudouri et al., 2024). Every bet stated the probability of winning a certain amount of money and the probability of losing a certain amount of money. Participants were asked whether they wanted to take the bet or not, and indicated their choices by clicking on one of two options, labeled as "Yes" (take the bet) and "No" (do not take the bet). Half of the presented bets in each block were conflict problems, which had a positive expected value and a high probability of losing money. Based on the expected value, it is in one's best financial interest to take the bets. However, when faced with such gambles most people avoid taking them given the high chance of losing money (Kahneman & Tversky, 1984). People often make biased decisions when evaluating risks, and tend to overestimate the impact of losses compared to the prospect of comparable potential gains (Kahneman & Tversky, 1979). This bias is often referred to as "loss aversion" (Camerer, 2005; Kahneman & Tversky, 1979; see also Gal & Rucker, 2018 for a discussion). In essence, in these items a conflict is created between avoiding a potential loss (i.e., not taking the bet) and taking a risk in order to acquire a bigger potential gain (i.e., taking the bet). An example of a conflict bet is presented below:

If you take this bet you have:
5%** probability to **WIN €110
95%** probability to **LOSE €5
Do you take the bet?

- *Yes*
- *No*

In the instructions, participants were told that their goal was to make as much profit as possible. Hence, according to an objective outcome calculation, participants should always take the (positive-expected value) bet. We adapted the material from Voudouri et al. (2024) so that

one conflict problem was an "easy" problem, with a very high expected value, and one was a "hard" problem, with a positive but comparatively lower expected value. Probabilities varied between a 5% probability of winning €290 and 95% probability of losing €1 ("easy" conflict problem), to a 10% probability of winning €100 and 90% probability of losing €10 ("hard" conflict problem). We added the easy and hard versions for exploratory purposes. Analyses focus on the aggregate results but the interested reader can find an overview of the individual item results in Figure S3 of the Supplementary Material, section A. Overall, participants were thus presented with two conflict problems (one hard, one easy) per block, and two no-conflict problems. The control no-conflict problems had a positive expected value, with a high probability of winning a significant amount of money and a low probability of losing a small amount of money. For example, 95% probability to win €120 and 5% probability to lose €5. Hence, these problems should not cue loss aversion–or only minimally–and participants should easily choose to take these bets, meaning that performance should be at ceiling.

2.3.2. Two-block paradigm

We used a two-block (or "fast and slow") paradigm for the presentation of the items in all three studies. In this paradigm, participants are presented with two successive blocks (Markovits et al., 2019; Raoelison, Keime, et al., 2021). In an initial intuitive block participants respond to each problem under a strict time limit and are instructed to respond as fast as possible with the first intuitive response that comes to mind ("fast" trials). Then, in a second block (final, deliberative block), participants are presented with the same problems again but without any time constraint, and are instructed to take their time to reflect on the problem before providing their response ("slow" trials). Restricting the processing time and instructing participants to give the first response that comes to mind in the initial intuitive block minimizes deliberation, maximizing the likelihood that responses are provided intuitively. By contrast, the conditions in the final deliberative block allow participants to engage in deliberation while solving the problem. The time limit in the fast trials was adjusted for each type of problem, based on previous pre-testing indicating for each problem type the time needed to merely read the problem, move the mouse, and select a response option (Bago & De Neys, 2017, 2019; Boissin et al., 2022; Franiatte et al., 2024; Voudouri et al., 2024). In Study 1 featuring bat-and-ball problems, participants had to respond within a time limit of 5 seconds. In Study 2 the time limit to respond to base-rate problems was 3 seconds. In Study 3, the time limit to respond to risky-choice problems was 4 seconds. In all three studies, the background color turned yellow one

second before the deadline to warn participants to submit their response. Note that the literature indicates that these deadlines impose a stringent time pressure that forces participants to respond significantly faster than in a traditional unconstrained, one-response test format (Bago & De Neys, 2017, 2019; Boissin et al., 2022; Franiatte et al., 2024; Voudouri et al., 2024).

2.4. Procedure

All three experiments were run online on the Qualtrics platform (www.qualtrics.com). Participants were first informed about the study's length and content and asked to give their consent. They were then presented with instructions about the two types of blocks, followed by two practice trials to get them acquainted with the format of both fast and slow trials. Participants were then presented with the bat-and-ball problems (Study 1), base-rate problems (Study 2), or risky-choice problems (Study 3) in a two-block paradigm.

A previous study using the two-block paradigm manipulated the order of intuitive and deliberative blocks and showed that the order factor did not affect response accuracy (Raoelison, Keime, et al., 2021). Therefore, we did not alternate the order of the blocks across participants in the present set of studies. Participants always started with the intuitive block and afterward moved on to the deliberative block. This also maximally guarantees that performance in the intuitive block is not boosted by deliberation in the deliberative block (Markovits et al., 2019).

In both blocks, after each problem, participants were asked to provide a confidence judgment in the correctness of their response, from 0 (absolutely not confident) to 100 (absolutely confident). Finally, they were asked to provide an explanation for their answer, in an open text entry. The specific instructions were the following: "Could you please try to explain why you selected this answer? Can you justify why you believe it might be correct? Please be as detailed as you can.". Critically, participants had no constraints when providing their explanations in either block. In the initial intuitive block, participants were under a strict time limit only to respond to the presented problem; there were no limitations of any kind on the explanation-justification step.

Whenever participants missed the deadline for the reasoning problem in fast trials, the confidence rating and explanation question were not displayed. Rather, they were presented with a message reminding them to make sure to respond before the deadline on the next item.

We also recorded participants' response time at each step of the procedure. At the end of each study, participants completed demographic questions and were presented with a debriefing message.

2.4.1. Justification Analysis

Two independent raters classified participants' explanations into predefined categories for each problem. The categories were adapted from exploratory data from Bago and De Neys (2019) for bat-and-ball problems, Boissin et al. (2022) for base-rate problems, and Voudouri et al. (2024) for risky-choice problems. The raters' classifications agreed in 76.5% of cases in Study 1 (612/800 explanations), 92.6% in Study 2 (741/800 explanations), and 74% (586/792 explanations) in Study 3 (for conflict and no-conflict problems altogether). When ratings differed, explanations were discussed, and an agreement was reached in all cases.

For the critical conflict problems, an explanation was considered correct when referring to the correct calculation in Study 1 with bat-and-ball problems (e.g., "1.10 in total, and 1 more for the bat than the ball. So .05 for the ball + 1.05 for the bat", "(1.10-1.00)/2 = .05". For simplicity, examples here are rephrased in the original bat-and-ball units). In Study 2 with base-rate problems, an explanation was considered correct when referring to the use of the base-rate (e.g., "Although high school students can be loud, so can librarians. The huge number of librarians means that the chance of it being a librarian is high."; "Lawyers are paid to put forward a strong argument. However, with there being so few lawyers, then it could be a gardener."). In Study 3 with risky-choice problems, an explanation was considered correct when referring to the (positive) expected value of the bet (e.g., "Although you wouldn't win often on this, when you do, overall the amount to bet is worth it. 95/100 times you'd lose the dollar, but then 5/100 times you'd win a lot more than you lost. Overall it is good value."; "The loss chance was high, but the gain was substantial compared to the minimal amount lost".).

For completeness, we also analysed the explanations on the control, no-conflict problems. Note that for the control no-conflict problem–on which the heuristic response is also correct–any reference to the corresponding mathematical calculation in Study 1 (e.g., "$1.10 – $1 = 10 cents"), the base-rate information and/or stereotype in Study 2, and the probabilities and/or values of the bet in Study 3 was considered a correct explanation.

Our main concern lies in correct explanations. Note that for exploratory purposes, we also distinguished different subcategories of incorrect explanations. The interested reader can find a

full overview of the different explanation categories in the Supplementary Material section A. Figures S4 and S5 in the Supplementary Material section A display the proportion of each explanation category by response stage and accuracy, for conflict and no-conflict problems.

2.5. Exclusion criteria

Study 1 (Bat-and-ball problems). Upon completion of the experiment, participants were presented with the standard bat-and-ball problem: they were asked if they had seen it before, and to provide the solution to the problem in free response format. Overall, 36 participants reported having seen, solved, or read about the classical bat-and-ball problem before *and* responded correctly to this problem. According to the preregistration, we performed separate data analyses with and without these participants who were familiar with the classical version of the bat-and-ball problem. Results showed that all trends and significance tests outcomes were similar with and without the exclusion. Considering these results, and as planned in the preregistration, we decided to report the results of the full analysis including all participants.

Fast trials for which participants did not respond before the deadline were removed from the analysis, because it couldn't be guaranteed that the response resulted from mere intuitive processing (i.e., if participants took longer than the deadline, they might have engaged in deliberation). In 15% of fast conflict trials (30 out of 200) and 6.5% of fast no-conflict trials (13 out of 200), participants failed to answer before the deadline. Overall, we thus kept 89.2% of all fast trials (357 out of 400), by rejecting trials in which participants missed the deadline. As per our preregistration, we also excluded the corresponding slow trials from these missed fast trials.

Study 2 (Base-rate problems). On 10.5% of fast conflict trials (21 out of 200) and 6.5% of fast no-conflict trials (13 out of 200) participants failed to answer before the deadline (8.5% of all fast trials). Overall, we thus kept 91.5% of all fast trials (366 out of 400), by rejecting trials in which participants missed the deadline. We excluded the corresponding slow trials from the analyses.

Study 3 (Risky-choice problems). On 19.7% of fast conflict trials (39 out of 198) and 3.5% of fast no-conflict trials (7 out of 198) participants failed to answer before the deadline. Overall, we kept 88.4% of all fast trials (350 out of 396), by rejecting trials in which participants missed the deadline. Corresponding slow trials were also excluded from the analyses.

2.6. Statistical analyses

Throughout the article, we used mixed-effects regression models (Baayen et al., 2008; Meteyard & Davies, 2020) in which participants and items were entered as random effect intercepts. When such models failed to converge, simpler models with participants as random effect intercept were used. We used logistic regression models for the binary data (e.g., correct vs incorrect responses and explanations), and linear regression models for continuous data (e.g., confidence). The interested reader can find the full specification of the model tables in section B of the Supplementary Material.

3. Results

3.1. Accuracy

Consistent with previous findings (Bago & De Neys, 2019; Bourgeois-Gironde & Van Der Henst, 2009; De Neys et al., 2013; Raoelison & De Neys, 2019), in most cases participants failed to correctly solve the conflict version of the tasks, both in fast and slow trials. Figure 1 shows the average proportion of correct responses in fast and slow trials for conflict problems at each task. In bat-and-ball problems (Study 1), the average accuracy in conflict problems increased from 9.7% (SEM = 2.7%) in fast trials to 25.5% (SEM = 4.4) in slow trials ($\chi^2(1) = 46.12$, $p < .001$). In base-rate problems (Study 2), response accuracy to conflict problems went from 33.5% (SEM = 4.2) in fast trials to 68.0% (SEM = 4.1) in slow trials ($\chi^2(1) = 59.69$, $p < .001$). In conflict risky-choice problems (Study 3), accuracy increased from 31.6% (SEM = 3.9) in fast trials to 43.4% (SEM = 4.0) in slow trials ($\chi^2(1) = 8.15$, $p < .01$).

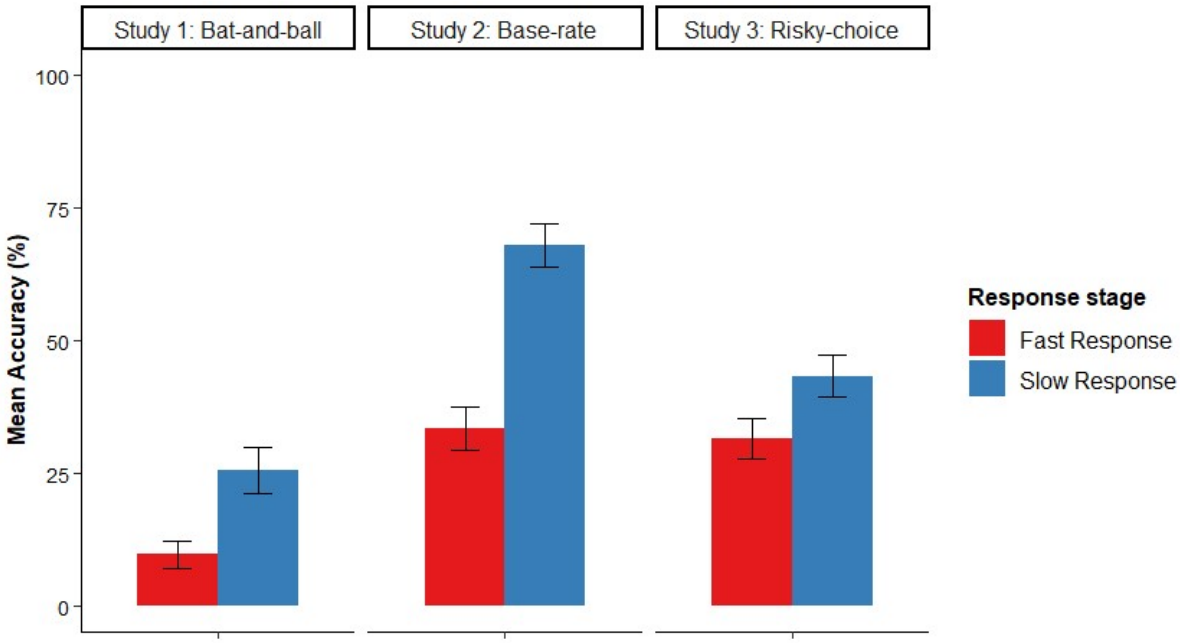

Figure 1. Mean accuracy (%) in fast and slow trials for conflict problems, at each task. Error bars represent the Standard Error of the Mean (SEM).

Thus, although performance in conflict problems was generally low, response accuracy was higher in slow trials when participants had time to deliberate than in fast trials.

In contrast, on the no-conflict control problems, as expected, the accuracy at the slow trials was at ceiling (Study 1 with bat-and-ball problems: M = 97.5%, SEM = 1.3; Study 2 with base-rate problems: M = 97.0%, SEM = 1.2; Study 3 with risky-choice problems: M = 98.0%, SEM = 1.0, see Figure S10 in the Supplementary Material section C). Most importantly, the accuracy at the initial fast trials was also at ceiling (Study 1 with bat-and-ball problems: M = 98.0%, SEM = 0.99; Study 2 with base-rate problems: M = 98.0%, SEM = 0.99; Study 3 with risky-choice problems: M = 94.4%, SEM = 1.7), indicating that the time constraints did not result in blind random responding[1]. Overall, these accuracy results are consistent with previous studies adopting the two-block design (Markovits et al., 2019; Raoelison et al., 2021).

[1] Note that the possibility of participants randomly guessing during fast trials is also ruled out by the analysis of response distributions on conflict bat-and-ball problems. If time constraints forced guessing on these problems' fast trials, incorrect responses would be evenly distributed across the options (as there were one correct and three incorrect options: 1 heuristic and 2 fillers). Instead, errors were predominantly heuristic, making up 95.45% of incorrect responses in fast trials (86.47% of all responses). This strongly suggests that responses on fast trials were not due to random guessing but reflect a reliance on intuitive heuristic processing.

3.2. Direction of change

To gain a deeper understanding of how participants changed their response (or not) after deliberation, we also conducted a direction of change analysis on the conflict problems (Bago & De Neys, 2017; Raoelison & De Neys, 2019). This analysis examines how participants' answers to the same problem change (or do not change) from the initial fast trial to the final slow trial. Participants can provide correct or incorrect responses in every trial of each block, resulting in four possible answer change patterns: "00" (incorrect response in both response stages), "11" (correct response in both stages), "01" (initial incorrect fast and final correct slow response), and "10" (initial correct fast and final incorrect slow response). Figure 2 shows the distribution of each direction of change category in conflict problems for the different tasks.

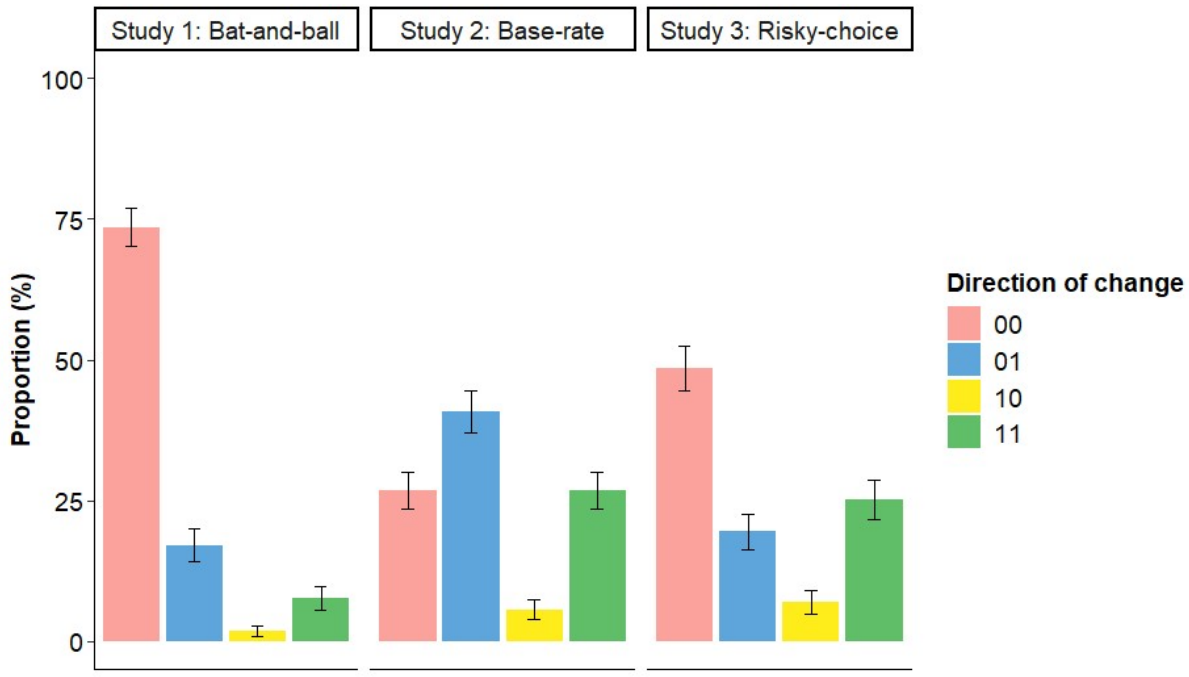

Figure 2. Proportion (%) of each direction of change category for conflict problems, at each task. Error bars represent the Standard Error of the Mean.

In Study 1 with bat-and-ball problems, the most prevalent category was the “00” one: participants provided an incorrect response at both response stages and remained biased in 73% of the trials (SEM = 4.4). The second most common pattern was “01” (17.3%, SEM = 3.6). These are cases in which reasoners generated incorrect responses in the fast trial of the initial intuitive block but corrected them when they solved the same problem in the slow trial of the final deliberative block. The third most common pattern was “11”: in 8.2% (SEM = 2.6) of the

cases, participants gave the correct response to the problem at both response stages. There were only very few responses corresponding to the "10" pattern (1.5%, SEM = 0.9). As in Bago and De Neys (2017), we calculated the so-called non-correction rate, indicating the proportion of responses in the final deliberative block that were already correct in the initial intuitive block (i.e., "11" / "11"+"01" ratio). In other words, it shows the proportion of cases for which participants did not need to deliberate to arrive at a correct response. The mean non-correction rate for the conflict problems reached 32.2% in Study 1. Hence, on average, when participants managed to give a correct answer to conflict problems in the final deliberative block, they already gave a correct answer in the corresponding trials of the initial intuitive block about 32% of the time.

In Study 2 using base-rate problems, the "01" pattern was the most prevalent, with 40% (SEM = 4.2) of cases. "11" was the second most common pattern, with 28% (SEM = 4.0), closely followed by "00", with 26.5% (SEM = 3.8). There were only 5.5% (SEM = 1.7) responses of the "10" pattern. The non-correction rate for conflict problems in this study was 41.2%.

In Study 3 with risky-choice problems, "00" was the most common pattern with 49% (SEM = 4.0) of cases. The "11" pattern was the second most common (24%, SEM = 3.4), followed by the "01" pattern (19.4%, SEM = 3.4). Lastly, the "10" pattern represented 7.7% of responses (SEM = 2.3). The non-correction rate in Study 3 was 55.3%.

Note that in these three studies, sound intuiting (corresponding to the "11" direction of change pattern) was overall less prevalent than in previous studies using the same types of problems (e.g., Bago & De Neys, 2017, 2019; Raoelison & De Neys, 2019; Voudouri et al., 2024). It was, in particular, less prevalent than the "01" pattern (as shown by the non-correction rates inferior or very close to 50%). However, within all 3 studies participants were able to generate the correct response in the initial intuitive block, when deliberation was restrained, in a substantial number of cases.

For completeness, we also conducted the direction of change analysis on no-conflict control problems. As expected, in the no-conflict control problems the overwhelming majority of responses were of the "11" category, i.e. correct responses given at both the initial fast and the final slow trials, in all three studies (see Figure S11 in the Supplementary Material section C).

3.3. Explanations

We now turn to the critical explanation data. For completeness, we first look at the overall, absolute number of correct explanations (i.e., irrespective of response accuracy). Figure 3A shows these results for conflict problems. Consistent with our predictions, we found that correct explanations were more likely after slow trials than after fast trials. Overall, participants only managed to provide sound explanations for their responses to fast trials in 3.06% (SEM = 1.59) of cases in Study 1 with bat-and-ball problems, 31.0% (SEM = 4.25) in Study 2 with base-rate problems, and 9.18% (SEM = 2.10) in Study 3 with risky-choice problems. Sound explanations after slow trials increased to 24.49% (SEM = 4.24) in Study 1 with bat-and-ball problems (χ^2 (1) = 115.75, $p < .001$), 70.5% (SEM = 4.21) in Study 2 with base-rate problems (χ^2 (1) = 104.96, $p < .001$), and 20.92% (SEM = 3.16) in Study 3 with risky-choice problems (χ^2 (1) = 10.91, $p < .001$).

Note that overall, even after deliberation the absolute number of correct explanations remains fairly low. Obviously, this is not surprising given that our accuracy data showed that participants often failed to solve classic conflict problems correctly. Figure 3B shows the proportion of correct explanations on trials that were also solved correctly. Here it becomes clear that when participants answered the problem correctly, they predominantly managed to provide a correct explanation on the slow trials (i.e., when they had had the time to deliberate). Critically, however, these correct explanations were less likely for correctly solved fast trials. In Study 1 with bat-and-ball problems, while participants were able to correctly explain their fast correct response in 26.92% (SEM = 12.16%) of cases, this rose to 90.38% (SEM = 5.56%) for slow trials (χ^2 (1) = 31.29, $p < .001$). Likewise, in Study 2 with base-rate problems, while participants correctly explained their fast correct response in 72.72% (SEM = 6.79%) of cases, this percentage went up to 90.38% (SEM = 3.30%) for slow trials (χ^2 (1) = 15.22, $p < .001$). In Study 3 with risky-choice problems, while participants correctly explained their fast correct response in 34.44% (SEM = 6.89%) of cases, this increased to 52.46% (SEM = 6.23%) for slow trials–although the increase was statistically marginal (χ^2 (1) = 2.65, $p = .10$).

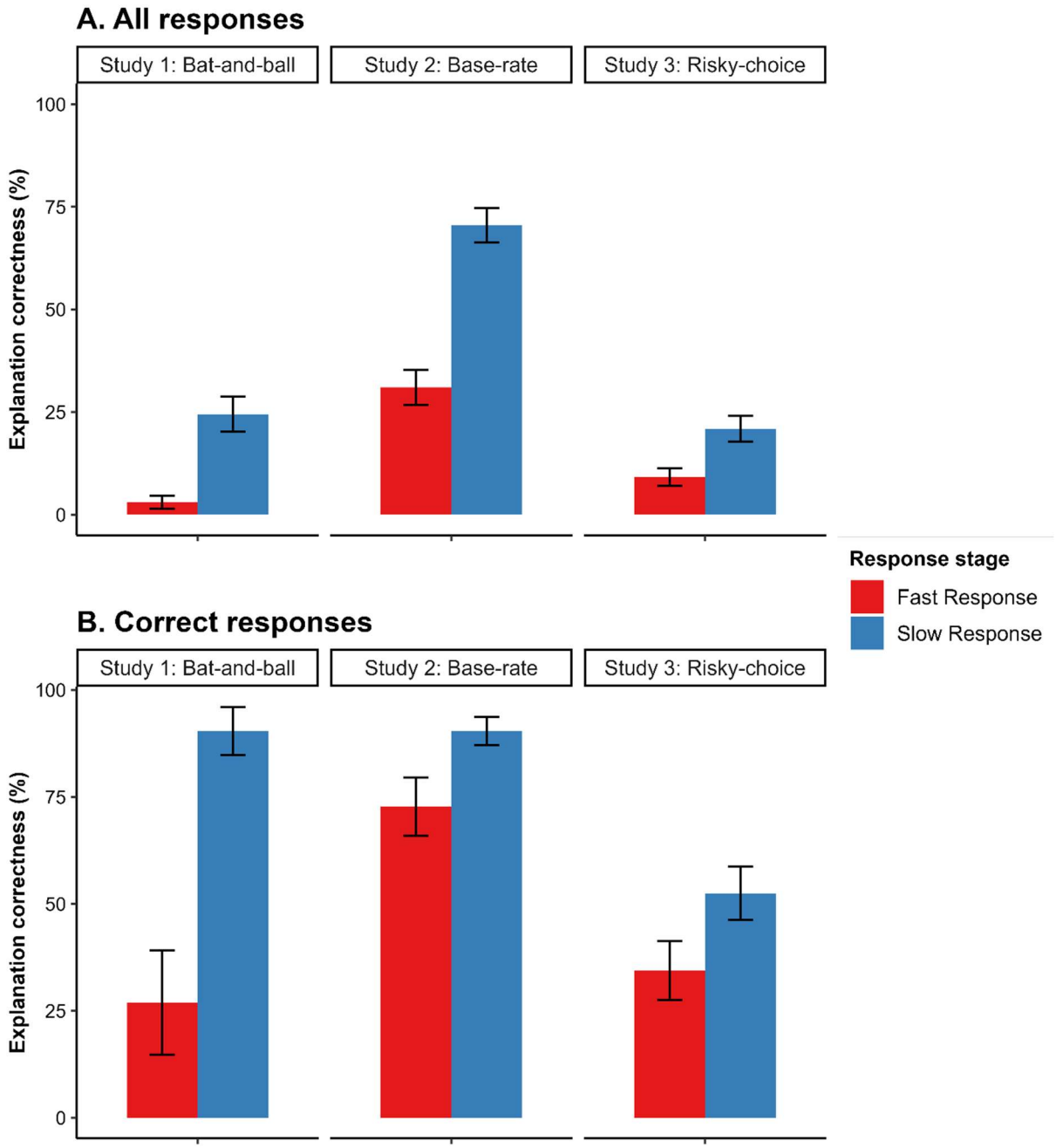

Figure 3. Mean proportion of correct explanations after fast and slow responses for correct and incorrect responses combined (A) and for correct responses only (B) in conflict problems, at each task. Error bars represent the Standard Error of the Mean.

In sum, while people can sometimes give the correct response intuitively, they typically tend to be better at explaining or justifying it after deliberation. At the same time, it is clear that some correct fast responses were correctly justified. This means that the absence of deliberation does not entail a complete incapacity to justify, the point is simply that such justification is less likely and common in the absence of deliberation.

As a sidenote, when eyeballing Figure 3B one may remark that the proportion of correct explanations for slow correct responses to conflict risky-choice problems (Study 3) was rather low. When participants correctly answered conflict risky-choice problems, they managed to properly justify their slow correct responses in slightly more than 50% of cases only. This can partly be explained by the fact that we adopted a conservative coding concerning the classification of explanations in the correct "Expected value" category. According to our coding criteria, an explanation had to explicitly mention the probabilities and values of the bet as well as their relationship to enter the "Expected value" category to be labeled as a correct explanation. An important number of explanations did not meet this level of detail while justifying that the bet was worth taking (e.g., "It's a low amount to lose", "Yeah, will take my chances."). These were coded in a separate category labelled as "Gambler" (37.26% of explanations to fast correct trials, 38.03% of explanations to slow correct trials, see Figure S4 in the Supplementary Material section A), and were considered incorrect explanations. Looser criteria for what constitutes a reference to the positive expected value of bets may have allowed more explanations to be considered correct. For example, when we exploratorily included the "Gambler" category as correct explanation, we obtained 90.16% correct explanations for correct slow responses after deliberation. Critically, even with this less strict scoring criterion, our key point holds that sound justification was more likely after "slow" deliberation than after more intuitive "fast" response generation (90.16% vs 74.44%, χ^2 (1) = 4.47, $p < .05$).

To provide clearer evidence for the added value of deliberation in response justification, and to assess whether this justification function can coexist with a corrective role, we specifically examine participants' explanations for responses that were correct on both fast and slow trials ("11" direction of change pattern), and contrast these with explanations given when the correct response was found only on the slow trial ("01" pattern). Figure 4 presents these data for conflict problems. Results show that when participants responded correctly on both fast and slow trials, correct explanations were more likely after slow correct responses than after fast ones. Specifically, in Study 1 with bat-and-ball problems, the proportion of sound explanations on "11" responses increased from 35.0% (SEM = 15.0) on fast trials to 90.0% (SEM = 10.0) on slow trials (χ^2 (1) = 13.90, $p < .001$). Similarly, in Study 2 with base-rate problems, this proportion increased from 83.78% (SEM = 6.14) on fast trials to 91.89% (SEM = 4.55) on slow trials (χ^2 (1) = 14.54, $p < .001$). In Study 3 with risky-choice problems, this proportion increased from 44.59% (SEM = 8.17) on fast trials to 52.70% (SEM = 8.21) on slow trials–although this increase was statistically not-significant (χ^2 (1) = .65, $p = .42$). Overall, this

pattern mirrors the results obtained when considering all correct responses, irrespective of the direction-of-change pattern.

On "11" trials, since participants initially provided the correct intuitive response, deliberation could not have served a corrective purpose. Conversely, on "01" trials, the fast intuitive response was incorrect, and the subsequent slow response was correct, indicating a corrective role of deliberation. However, the proportion of correct explanations following correct slow responses was similar between the "01" and "11" patterns in each study (Study 1 with bat-and-ball problems: χ^2 (1) = .01, p = .93 ; Study 2 with base-rate problems: χ^2 (1) = .01, p = .94 ; Study 3 with risky-choice problems: χ^2 (1) = .10, p = .75). Thus, although deliberation served to correct erroneous intuitions in the "01" pattern, it also consistently enabled participants to properly justify their slow correct responses (at least to the same extent as in "11" trials).

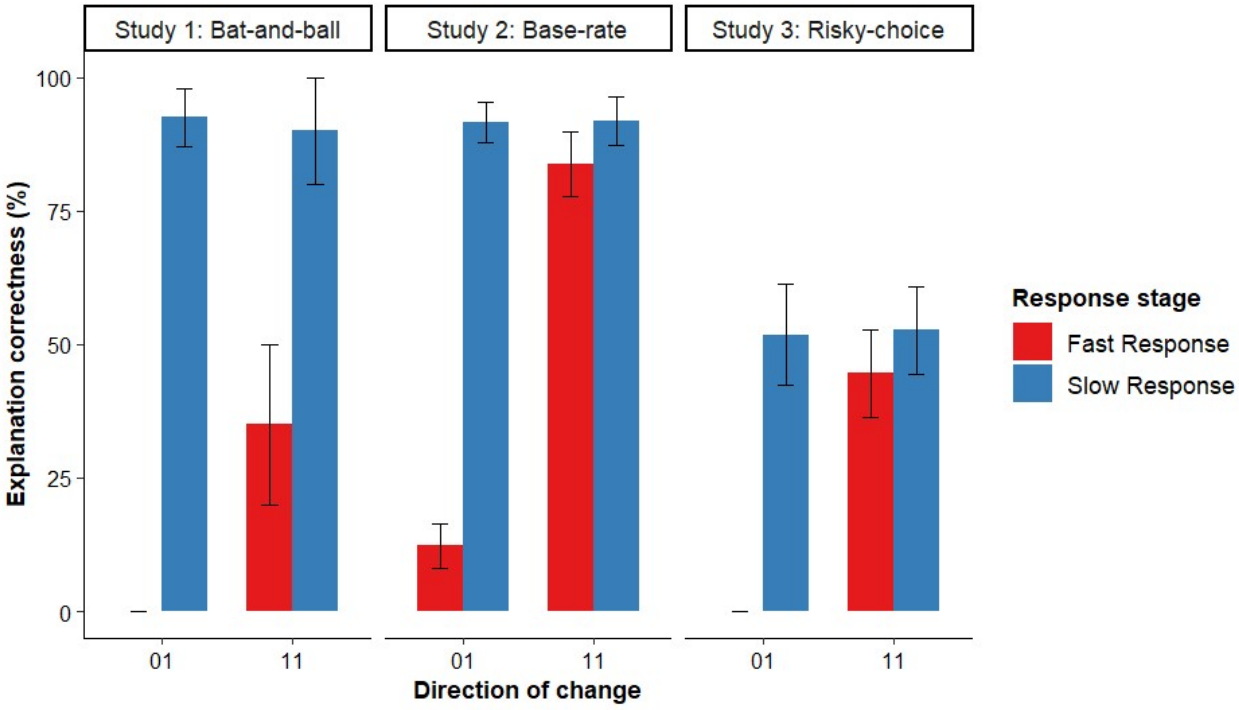

Figure 4. Mean proportion of correct explanations after fast and slow responses on trials of the 01 and 11 direction of change pattern, in conflict problems and at each task. Error bars represent the Standard Error of the Mean.

For completeness, we also analyzed the explanation data for control, no-conflict problems, in which the heuristically cued answer is also correct. Figure S1 in the Supplementary Material

section A shows the explanation data for control, no-conflict problems in each task. The results show that not surprisingly, people had little trouble explicitly justifying their correct deliberate *and* intuitive answers to these problems (e.g., overall in 83.89% and 83.05% of cases, respectively), as they typically referred to the simple heuristic answers (e.g., simple subtraction on bat-and-ball problems, stereotype on base-rate problems, probabilities and/or values in risky-choice problems), which are correct explanations on these problems.

3.4. Additional analyses

For exploratory purposes, we conducted an analysis of the impact of problem difficulty on explanations' correctness in Study 3 to test whether easier problems also tend to be easier to justify (beyond the mere contrast between conflict and no-conflict problems). Figure S3 in the Supplementary Material section A shows the proportion of correct explanations for "easy" and "hard" risky-choice conflict problems. Problem difficulty significantly impacted explanation correctness: participants provided slightly more correct explanations for their correct responses to the "easy" version compared to the "hard" version of risky-choice problems (χ^2 (1)= 8.66, p < .01). These results suggest that while deliberation helps individuals to justify their decisions, it is also generally easier to come up with a proper justification when the problem faced is simpler.

We also examined the explanations for incorrect responses and incorrect explanations in more detail. The interested reader can find a complete overview of the distribution of the explanation category types in Figures S4 and S5 in the Supplementary Material section A. Results indicate that for both fast and slow incorrect conflict responses participants predominantly generated an explanation based on the cued heuristic (i.e., simple subtraction in bat-and-ball problems; stereotype in base-rate problems; loss aversion in risky-choice problems). When participants gave incorrect explanations for correct fast or slow responses, these explanations often referred to intuition, guessing, or were idiosyncratic (i.e., from the category "Other"). As previously mentioned, in the risky choice problems in Study 3, many of the incorrect explanations for correct responses referred to a preference to take the bet without pointing clearly to the expected value ("Gambler" category).

In an exploratory analysis, we further examined the level of explicitness of the explanations (e.g., an unspecified generic explanation: "I just felt it", as opposed to a more specific explicit explanation: "It's 10 cents because $1.10 - $1 is 10 cents"), after fast and slow trials. However,

we did not identify any clear pattern across the different tasks (see Figure S6 in the Supplementary Material section A).

Finally, we also considered participants' confidence in their fast and slow responses and the link between confidence and different types of explanation. Consistent with previous studies (e.g., Bago & De Neys, 2019; Shynkaruk & Thompson, 2006; Thompson et al., 2011), Figure S7 in the Supplementary Material section A shows that participants consistently have higher confidence in their slow, deliberative responses than in their fast, intuitive responses. However, we did not identify further clear trends in response confidence related to explanation explicitness or correctness (see Figures S8 and S9 in the Supplementary Material section A). Note, however, that we had limited statistical power within each study–especially when splitting by response accuracy–which restricted our ability to reliably detect interactions between response stage, confidence, and explanation explicitness or correctness.

4. Discussion

In this set of studies, we directly tested the justificatory role of deliberation by examining whether individuals can justify their responses to various reasoning and decision-making problems when these were given intuitively or after deliberation. Results show that individuals provided more correct explanations for their deliberative responses compared to their intuitive ones. Most importantly, intuitive correct responses were less likely to be properly justified than deliberative correct responses. These results give empirical support to the hypothesis that deliberative reasoning allows not only to correct erroneous intuitions but also to properly justify sound ones.

Concerning response accuracy, consistent with previous studies, we found that reasoners struggle to solve conflict problems that cue a heuristic answer conflicting with the correct logical answer (Bourgeois-Gironde & Van Der Henst, 2009; De Neys et al., 2013; Frederick, 2005; Raoelison & De Neys, 2019). Although response accuracy remained low in both blocks, it was consistently higher in the final deliberative block in which participants could deliberate on the problems. This highlights that deliberation can indeed serve a corrective function when intuitive responses are incorrect. It is worth noting that we observed lower non-correction rates than in previous studies. Especially in Study 1 (bat-and-ball problems) and 2 (base-rate problems), deliberative reasoning was needed to correct the initial intuitive response most of

the time. These results contrast with previous studies showing that the majority of participants who respond correctly after deliberation are already correct in their earlier intuitive response (Bago & De Neys, 2017, 2019). Contrary to our set of studies, this previous work asked participants to successively provide an intuitive and a deliberate response in every trial (two-response paradigm, Thompson et al., 2011). Our results may suggest that the paradigm could play a role in the amount of non-corrective correct responses given by participants–that is, the amount of correct intuitive responses compared to all correct responses. Consistent with Meyer and Frederick (2023)–but contrary to results from Raoelison and De Neys (2019)–this suggests that the intervening deliberation in the two-response paradigm might boost the accuracy in later intuitive trials.

Furthermore, the combination of being asked to generate one's response and having to justify it may also boost deliberation (Bago & De Neys, 2019; Evans, 2019; Isler et al., 2020). Recent work suggests that putting participants in argumentative settings can improve their reasoning outcomes (Claidière et al., 2017; Mercier et al., 2017). Asking participants to justify their responses might mimic an argumentative setting in which participants have to convince a third party of their response, which could incentivize them to engage in more thorough thinking. As we did not observe an increase in correct intuiting compared to other studies that used similar problems (Bago & De Neys, 2019; Boissin et al., 2021), our results suggest that such an effect might happen specifically at the level of deliberative thinking. This hypothesis, however, would require more direct testing. Future work could systematically manipulate the presence or absence of justification demands to examine their specific impact on intuitive versus deliberative responses. Investigating whether and how argumentative framing selectively enhances deliberation–rather than intuition–represents a promising direction for further research. Implementing a think-aloud protocol (see e.g., Byrd et al., 2023) during the deliberative response stage could also provide valuable insight into participants' reasoning processes and shed light on the specific mechanisms at play during deliberation.

Concerning the critical justification results, it is important to stress once again that while our results showed that sound justification was more likely after deliberative responses than intuitive ones, correct explanations for intuitive responses were not completely absent. Consequently, the claim that deliberative reasoning allows justification is not a categorical one; proper explanations can sometimes be given for decisions made intuitively. In particular, simpler problems seem not only easier to solve, but solutions appear easier to justify as well. This point is illustrated by the exploratory analysis comparing easy and hard risky choice

problems: participants gave more correct explanations for easy problems than for hard ones. Further evidence comes from the exploratory examination of explanations for no-conflict problems. Not only were the (intuitive) accuracy rates near ceiling on these problems, but results also indicated that people had little trouble properly justifying both their correct deliberate *and* intuitive answers. That is, people typically referred to the heuristic answer (i.e., simple subtraction on bat-and-ball problems; stereotype on base-rate problems; win and loss probabilities on risky-choice problems) on these no-conflict problems which–by definition–is also correct on these problems. This is perhaps not surprising given the nature of the easy control problems, but it underscores the more general point that giving a proper explicit reason to justify an answer does not require deliberation per se. People can readily justify intuitive heuristic answers. It seems to be the specific justification of the logico-mathematical principles that are at play in challenging conflict problems that typically will require deliberation.

While we may often tend to act spontaneously based on intuitions, we undoubtedly also rely on (deliberative) reasoning to make decisions. However, we might engage in deliberation not necessarily to correct our intuitions, but also to find reasons to justify them. Reviewing the evidence for both functions, Evans (2019) already speculated that justification or rationalization might be the primary function of deliberative reasoning, rather than correction or decision (see also Haidt, 2001 for a similar perspective in the context of moral reasoning). Recent argumentative theories of reasoning also strongly emphasized the primacy of this justificatory function (Mercier & Sperber, 2017). Our results provide empirical support for the justificatory function of deliberation. Although intuitions can yield correct, logical responses in classic reasoning problems, it is less likely that they enable one to clearly explain why the response is correct. On the contrary, deliberative reasoning allows individuals to reflect on their intuitive responses to justify them. Thus, rather than being a corrective mechanism *per se,* deliberative reasoning also allows individuals to justify their intuitions and decisions. However, we do not claim that deliberation never serves to correct erroneous intuitions. As previously noted, some of our results (e.g., the presence of "01" cases) do support a corrective role of deliberation. While our data does not allow us to determine which function takes precedence, they show that beyond correction, deliberation also plays a key role in enabling individuals to justify their responses. Importantly, this justificatory role does not exclude the possibility of correction: the specific analysis of "01" response patterns showed that even when deliberation corrects an erroneous intuition, it simultaneously enables to explicitly justify the corrected response. Disentangling more precisely the relationship between these functions–whether in terms of

primacy, overlap, or mutual exclusivity– as well as exploring other potential roles of deliberation (see De Neys, 2025), remains an important avenue for future research.

We should also stress that the current results do not inform us about the nature of the process that allows a reasoner to generate a correct intuitive response. That is, it has been suggested that the computation of correct intuitive responses during deductive reasoning may rely on superficial features that align with the logical status of a conclusion–though these features may have no intrinsic or epistemological link to logical validity–rather than on proper "logical" knowledge per se (Ghasemi et al., 2022; Hayes et al., 2022; Mekik et al., 2025; Meyer-Grant et al., 2023). In this sense, intuitive sound reasoning would serve to calculate a proxy of logical reasoning but not actual logical reasoning (De Neys, 2023). Clearly, whatever the underlying nature of a correct intuitive response may be, given that the process will remain intuitive and thus non-transparent, one would not expect reasoners to manage to explicate their underlying reasoning. Hence, the current results are orthogonal to this broader issue and should not be used to support either one account.

To avoid confusion, note that in the literature the role of deliberation in justification or so-called "rationalization" has often focused on the justification of "incorrect" intuitions (see Cushman, 2020; De Neys, 2025; for discussion). This phenomenon likely emerged in our data as well: participants who responded incorrectly on both fast and slow trials may have generated explicit reasons to support their erroneous answers during deliberation. Some may also have used deliberation not merely for post hoc rationalization, but as a mean to actively construct the heuristic response through flawed reasoning. Nonetheless, this paper does not aim to examine rationalizations per se, nor to disentangle them from instances where deliberation fails to override misleading intuitions. What the present work clearly highlights is that a similar deliberate justification process may be equally important for *sound* intuitions. Intuitive or System 1 processes are typically less cognitively transparent, and deliberation appears especially useful for making them more explicit.

Acknowledgements

This research was funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency (HaDEA). Neither the European Union nor the granting authority can be held responsible for them. Grant Agreement no. 101120763 – TANGO.

Declaration of interest

The authors report no conflict of interest.

Research data availability

The data, material, analysis scripts, and pre-registrations are available on OSF at https://osf.io/pm3je/.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412. https://doi.org/10.1016/j.jml.2007.12.005

Bago, B., & De Neys, W. (2017). Fast logic? : Examining the time course assumption of dual process theory. *Cognition*, *158*, 90-109. https://doi.org/10.1016/j.cognition.2016.10.014

Bago, B., & De Neys, W. (2019). The Smart System 1 : Evidence for the intuitive nature of correct responding on the bat-and-ball problem. *Thinking and Reasoning*, *25*(3), 257-299. https://doi.org/10.1080/13546783.2018.1507949

Barbey, A. K., & Sloman, S. A. (2007). Base-rate respect : From ecological rationality to dual processes. *Behavioral and Brain Sciences*, *30*(3), 241-254. https://doi.org/10.1017/S0140525X07001653

Boissin, E., Caparos, S., Raoelison, M., & De Neys, W. (2021). From bias to sound intuiting : Boosting correct intuitive reasoning. *Cognition*, *211*, 104645. https://doi.org/10.1016/j.cognition.2021.104645

Boissin, E., Caparos, S., Voudouri, A., & de Neys, W. (2022). Debiasing system 1 : Training favours logical over stereotypical intuiting. *Judgment and Decision Making*, *17*(4), 646-690.

Bonnefon, J.-F. (2013). New ambitions for a new paradigm : Putting the psychology of reasoning at the service of humanity. *Thinking & Reasoning*, *19*(3-4), 381-398. https://doi.org/10.1080/13546783.2013.774294

Bonnefon, J.-F. (2016). The Pros and Cons of Identifying Critical Thinking with System 2 Processing. *Topoi*, *37*(1), 113-119. https://doi.org/10.1007/s11245-016-9375-2

Bourgeois-Gironde, S., & Van Der Henst, J.-B. (2009). How to open the door to System 2 : Debiasing the Bat-and-Ball problem. *Rational Animals, Irrational Humans*, 235-252.

Byrd, N., Joseph, B., Gongora, G., & Sirota, M. (2023). Tell Us What You Really Think : A Think Aloud Protocol Analysis of the Verbal Cognitive Reflection Test. *Journal of Intelligence*, *11*(4), Article 4. https://doi.org/10.3390/jintelligence11040076

Camerer, C. (2005). Three Cheers—Psychological, Theoretical, Empirical—For Loss Aversion. *Journal of Marketing Research*, *42*(2), 129-133. https://doi.org/10.1509/jmkr.42.2.129.62286

Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers : Consequences and solutions for behavioral researchers. *Behavior Research Methods*, *46*(1), 112-130. https://doi.org/10.3758/s13428-013-0365-7

Claidière, N., Trouche, E., & Mercier, H. (2017). Argumentation and the diffusion of counter-intuitive beliefs. *Journal of Experimental Psychology: General*, *146*, 1052-1066. https://doi.org/10.1037/xge0000323

Cushman, F. (2020). Rationalization is rational. *Behavioral and Brain Sciences*, *43*, e28. https://doi.org/10.1017/S0140525X19001730

De Neys, W. (2014). Conflict detection, dual processes, and logical intuitions : Some clarifications. *Thinking & Reasoning*, *20*(2), 169-187. https://doi.org/10.1080/13546783.2013.854725

De Neys, W. (2023). Advancing theorizing about fast-and-slow thinking. *Behavioral and Brain Sciences*, *46*, 1-71. https://doi.org/10.1017/S0140525X2200142X

De Neys, W. (2025). Defining deliberation for dual-process models of reasoning. *Nat Rev Psychol* (2025). https://doi.org/10.1038/s44159-025-00466-6

De Neys, W., Rossi, S., & Houdé, O. (2013). Bats, balls, and substitution sensitivity : Cognitive misers are no happy fools. *Psychonomic Bulletin & Review*, *20*(2), 269-273. https://doi.org/10.3758/s13423-013-0384-5

Evans, J. S. B. T. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology*, *59*(1), 255-278. https://doi.org/10.1146/annurev.psych.59.103006.093629

Evans, J. S. B. T. (2019). Reflections on reflection : The nature and function of type 2 processes in dual-process theories of reasoning. *Thinking & Reasoning*, *25*(4), 383-415. https://doi.org/10.1080/13546783.2019.1623071

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition : Advancing the Debate. *Perspectives on Psychological Science*, *8*(3), 223-241. https://doi.org/10.1177/1745691612460685

Franiatte, N., Boissin, E., Delmas, A., & De Neys, W. (2024). Boosting debiasing : Impact of repeated training on reasoning. *Learning and Instruction*, *89*, 101845. https://doi.org/10.1016/j.learninstruc.2023.101845

Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, *19*(4), 25-42. https://doi.org/10.1257/089533005775196732

Gal, D., & Rucker, D. D. (2018). The Loss of Loss Aversion : Will It Loom Larger Than Its Gain? *Journal of Consumer Psychology*, *28*(3), 497-516. https://doi.org/10.1002/jcpy.1047

Ghasemi, O., Handley, S., Howarth, S., Newman, I. R., & Thompson, V. A. (2022). Logical intuition is not really about logic. *Journal of Experimental Psychology: General*, *151*(9), 2009-2028. https://doi.org/10.1037/xge0001179

Gigerenzer, G., Hell, W., & Blank, H. (1988). Presentation and Content : The Use of Base Rates as a Continuous Variable. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 513-525.

Haidt, J. (2001). The emotional dog and its rational tail : A social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814-834. https://doi.org/10.1037/0033-295X.108.4.814

Haigh, M. (2016). Has the Standard Cognitive Reflection Test Become a Victim of Its Own Success? *Advances in Cognitive Psychology*, *12*(3), 145-149. https://doi.org/10.5709/acp-0193-5

Hayes, B. K., Stephens, R. G., Lee, M. D., Dunn, J. C., Kaluve, A., Choi-Christou, J., & Cruz, N. (2022). Always look on the bright side of logic? Testing explanations of intuitive sensitivity to logic in perceptual tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(11), 1598-1617. https://doi.org/10.1037/xlm0001105

Isler, O., Yilmaz, O., & Dogruyol, B. (2020). Activating reflective thinking with decision justification and debiasing training. *Judgment and Decision Making*, *15*(6), 926-938. https://doi.org/10.1017/S1930297500008147

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kahneman, D., & Frederick, S. (2005). A Model of Heuristic Judgment. In K. Holyoak & B. Morrison (Éds.), *The Cambridge Handbook of Thinking and Reasoning* (p. 267-293). Cambridge University Press.

Kahneman, D., & Tversky, A. (1979). Prospect Theory : An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263-291. https://doi.org/10.2307/1914185

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, *39*(4), 341-350. https://doi.org/10.1037/0003-066X.39.4.341

Keysar, B., Hayakawa, S. L., & An, S. G. (2012). The Foreign-Language Effect : Thinking in a Foreign Tongue Reduces Decision Biases. *Psychological Science*, *23*(6), 661-668. https://doi.org/10.1177/0956797611432178

Markovits, H., de Chantal, P.-L., Brisson, J., Dubé, É., Thompson, V., & Newman, I. (2021). Reasoning strategies predict use of very fast logical reasoning. *Memory & Cognition*, *49*(3), 532-543. https://doi.org/10.3758/s13421-020-01108-3

Markovits, H., de Chantal, P.-L., Brisson, J., & Gagnon-St-Pierre, É. (2019). The development of fast and slow inferential responding : Evidence for a parallel development of rule-based and belief-based intuitions. *Memory & Cognition*, *47*(6), 1188-1200. https://doi.org/10.3758/s13421-019-00927-3

Mega, L. F., & Volz, K. G. (2014). Thinking about thinking : Implications of the introspective error for default-interventionist type models of dual processes. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00864

Mekik, C., Vivier, O., & Markovits, H. (2025). A "logical intuition" based on semantic associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0001468

Mercier, H., Boudry, M., Paglieri, F., & Trouche, E. (2017). Natural-Born Arguers : Teaching How to Make the Best of Our Reasoning Abilities. *Educational Psychologist*, *52*(1), 1-16. https://doi.org/10.1080/00461520.2016.1207537

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*(2), 57-74. https://doi.org/10.1017/S0140525X10000968

Mercier, H., & Sperber, D. (2017). The Enigma of Reason. In *The Enigma of Reason*. Harvard University Press. https://doi.org/10.4159/9780674977860

Meteyard, L., & Davies, R. A. I. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, *112*, 104092. https://doi.org/10.1016/j.jml.2020.104092

Meyer, A., & Frederick, S. (2023). The formation and revision of intuitions. *Cognition*, *240*, 105380. https://doi.org/10.1016/j.cognition.2023.105380

Meyer-Grant, C. G., Cruz, N., Singmann, H., Winiger, S., Goswami, S., Hayes, B. K., & Klauer, K. C. (2023). Are logical intuitions only make-believe? Reexamining the logic-liking effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *49*(8), 1280-1305. https://doi.org/10.1037/xlm0001152

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*, 1154-1170. https://doi.org/10.1037/xlm0000372

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. https://doi.org/10.1016/j.cogpsych.2015.05.001

Raoelison, M., Boissin, E., Borst, G., & De Neys, W. (2021). From slow to fast logic : The development of logical intuitions. *Thinking & Reasoning*, *27*(4), 599-622. https://doi.org/10.1080/13546783.2021.1885488

Raoelison, M., & De Neys, W. (2019). Do we de-bias ourselves? : The impact of repeated presentation on the bat-and-ball problem. *Judgment and Decision Making*, *14*(2), 170-178. https://doi.org/10.1017/S1930297500003405

Raoelison, M., Keime, M., & De Neys, W. (2021). Think slow, then fast : Does repeated deliberation boost correct intuitive responding? *Memory & Cognition*, *49*(5), 873-883. https://doi.org/10.3758/s13421-021-01140-x

Shynkaruk, J. M., & Thompson, V. A. (2006). Confidence and accuracy in deductive reasoning. *Memory & Cognition*, *34*(3), 619-632. https://doi.org/10.3758/bf03193584

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*(1), 3-22. https://doi.org/10.1037/0033-2909.119.1.3

Stieger, S., & Reips, U.-D. (2016). A limitation of the Cognitive Reflection Test : Familiarity. *PeerJ*, *4*, e2395. https://doi.org/10.7717/peerj.2395

Szaszi, B., Szollosi, A., Palfi, B., & Aczel, B. (2017). The cognitive reflection test revisited : Exploring the ways individuals solve the test. *Thinking & Reasoning*, *23*(3), 207-234. https://doi.org/10.1080/13546783.2017.1292954

Thompson, V. A., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology*, *63*(3), 107-140. https://doi.org/10.1016/j.cogpsych.2011.06.001

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty : Heuristics and Biases. *Science*, *185*(4157), 1124-1131. https://doi.org/10.1126/science.185.4157.1124

Voudouri, A., Białek, M., & De Neys, W. (2024). Fast & slow decisions under risk : Intuition rather than deliberation drives advantageous choices. *Cognition*, *250*, 105837. https://doi.org/10.1016/j.cognition.2024.105837

Supplementary Materials

A. Explanations

Explanation categories

Study 1 (Bat-and-ball problems). The examples from each category were taken from participants' explanations. For simplicity, examples are rephrased as in the classical version of the bat-and-ball problem, except for the "Other" category.

Correct math: People referred to the correct mathematical solution (e.g., "1.10 in total, and 1 more for the bat than the ball. So .05 for the ball + 1.05 for the bat", ".05 + .05 + 1 = 1.10", "(1.10-1.00)/2 = .05".

Incorrect math: Participants referred to a mathematical solution but it was not correct (e.g., "I took the total number of both, and subtracted the highest number from the total to calculate the lowest number", "1.10 minus 1 is 10 cents", "1.10 – 1 = .10").

Unspecified math: Participants referred to mathematical calculation but did not specify the calculation (e.g., "I just did the math", "More time to think so I can do the math", "Addition/subtraction problem").

Hunch: People referred to their gut feeling or intuition (e.g., "No other choice seemed correct", "I'm not sure if my answer is correct, I find it hard to read the question within the time frame so go with what my head says might be right", "Just seemed right").

Guess: Participants referred to guessing. (e.g., "I guessed", "I guessed based on the numbers I saw. There are too many words to read fast enough to get the point of the question", "Took a bit longer to work out so more of a guess").

Previous: Participants referred to previous answers without specifying it (e.g., "I used the same logic with this one to calculate as the last.", "Same as before").

Other: Any explanation that did not fit in other categories (e.g., "The first answer was wrong", "There would be more roses, because they flower more quickly and last longer and take up less space", "Only that lotuses are more rare than roses so I knew it had to be less").

In conflict problems, explanations from the "Correct math" category were considered correct. Note that for no-conflict problems, justifications of the form "1.10 – 1 = .10" (i.e., corresponding to incorrect math explanations for conflict problems) are correct explanations and were coded as such.

Study 2 (Base-rate problems).

Base-rate: The explanation refers to the base-rate (e.g.: "Because the ratio of secretaries to telemarketers is so overwhelming the odds favor it being a secretary"; "Number of secretaries so much greater").

Intuitions: Explanation refers to gut-feelings or intuition (e.g., "This one came straight to my mind"; "Instinct").

Guess: Explanation refers to guessing (e.g., "I got confused by the information and guessed"; "Guessed").

Stereotype: The explanation refers to the stereotype of a group as justifying the answer (e.g.: "Because students are generally louder than librarians who tend to be calm or used to calm environments").

BASE-RATE & Stereotype: When the explanation mentions both the base-rate and the stereotype but emphasizes the former (e.g.: "There is only a 5 in 1000 chance that the person selected is a high school student. So despite the characteristic being less typical of a librarian, the person is more likely to be a librarian.")

Base-rate & STEREOTYPE: When the explanation mentions both the base-rate and the stereotype but emphasizes the latter (e.g.: "Lawyers are paid to put forward a strong argument. However, with there being so few lawyers, then it could be a gardener.")

Base-rate & Stereotype: When the explanation mentions both the base-rate and the stereotype but without clear hierarchy (e.g.: "Aerobics instructors are active and there is more of them in the study").

Other: Any other explanation not fitting in the other categories (e.g., "Not sure"; "Having new data changes my mind").

For conflict problems, explanations were considered correct if they belonged to any category that mentioned the base rate (i.e., categories "Base-rate", "BASE-RATE &

Stereotype", "Base-rate & STEREOTYPE", and "Base-rate & Stereotype"). For no-conflict problems, explanations were considered correct if they belonged to any category that mentioned either the base-rate or the stereotype (i.e., categories "Base-rate", "BASE-RATE & Stereotype", "Base-rate & STEREOTYPE", Base-rate & Stereotype", and "Stereotype").

Study 3 (Risky-choice problems).

Expected value: The explanation refers to the (positive) expected value of the bet, considering both the amounts and the probabilities and the relationship between them (e.g., "Although you wouldn't win often on this, when you do, overall the amount to bet is worth it. 95/100 times you'd lose the dollar, but then 5/100 times you'd win a lot more than you lost. Overall it is good value."; "The loss chance was high, but the gain was substantial compared to minimal amount lost".

Gambler: The explanation does not explicitly refer both to the values and the probabilities and their relationship. Rather, it shows a tendency towards taking the risk (e.g., "Yeah, will take my chances."; "It's a low amount to lose").

Loss aversion: The explanation refers either to the amount(s) (e.g., "A lot of money to lose, not a lot of money to win") or to the probabilities (e.g., "High chance to lose, low chance to win") -but not to both- to justify not taking the bet.

Probabilities and values: The explanation refers both to the win/loss amounts and their probabilities, either with words or with numbers. In conflict problems, this concerns explanations for not taking the bet (e.g., "Insufficient reward for such a remote chance of winning."; "Not worth the 90% chance to lose $10"). In no-conflict problems, this concerns explanations for taking the bet (e.g., "High probability of winning and large payoff").

Intuition: The explanation refers to gut feelings or intuition (e.g., "This was an instinctive choice, but the odds looked really good"; "This was an instinct of the money that could be won").

Guess: The explanation refers to guessing (e.g., "Just guess"; "I didn't have enough time to think").

Other: The explanation cannot be coded into the other categories or is not understandable (e.g., "I have bad luck"; "Could process the information without time pressure").

Specifically for the no-conflict problems we also distinguished two additional categories:

Probability: In no-conflict problems, the explanation refers to the fact that the probability of winning is higher, regardless of the amounts (e.g., "Because there were higher chances I could win."; "Very high chance of winning').

Values: In no-conflict problems, the explanation refers to the amounts and not to the probabilities (e.g., "Not much money to lose and lots to gain"; "The reward was greater than the loss").

Note that these two no-conflict categories (i.e., Probability; Values) would correspond to the Loss aversion category in conflict problems, only in no-conflict problems the probability of winning is much higher than that of losing *and* the amount to win is high (expected value is high), so these categories allow a more fine-grained level of detail.

Explanations were considered correct in conflict problems when belonging to the "Expected value" category. In no-conflict problems, explanations were considered correct when belonging to the categories "Probability", "Values", "Probabilities and values" and "Expected value".

Explanation Accuracy

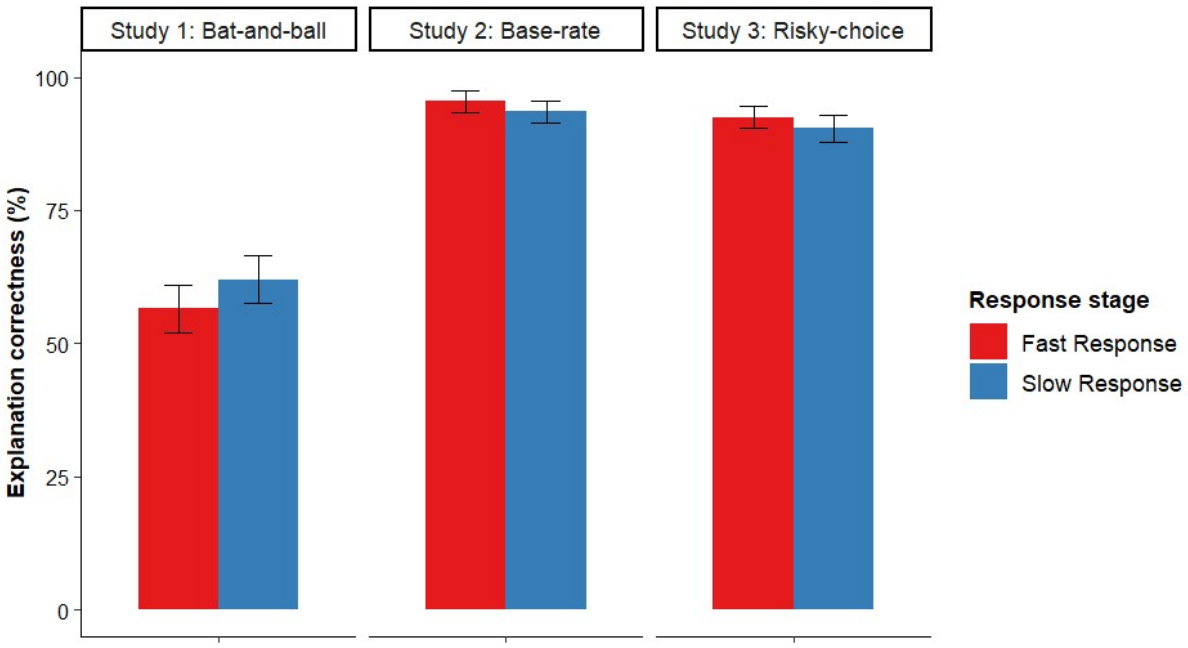

Figure S1. Proportion of correct explanations by response stage in no-conflict problems, at each task. Error bars represent the Standard Error of the Mean.

Note that the lower proportion of correct explanations for no-conflict problems observed in Study 1 compared to Studies 2 and 3 is partly explained by the significant number of explanations in the "unspecified math" category (e.g., "did the math") and "previous" category (e.g., "same logic as before"), in which participants did not specify their actual calculations (see Figure S5). This implies that participants could have known the correct explanation but we simply could not code their answer.

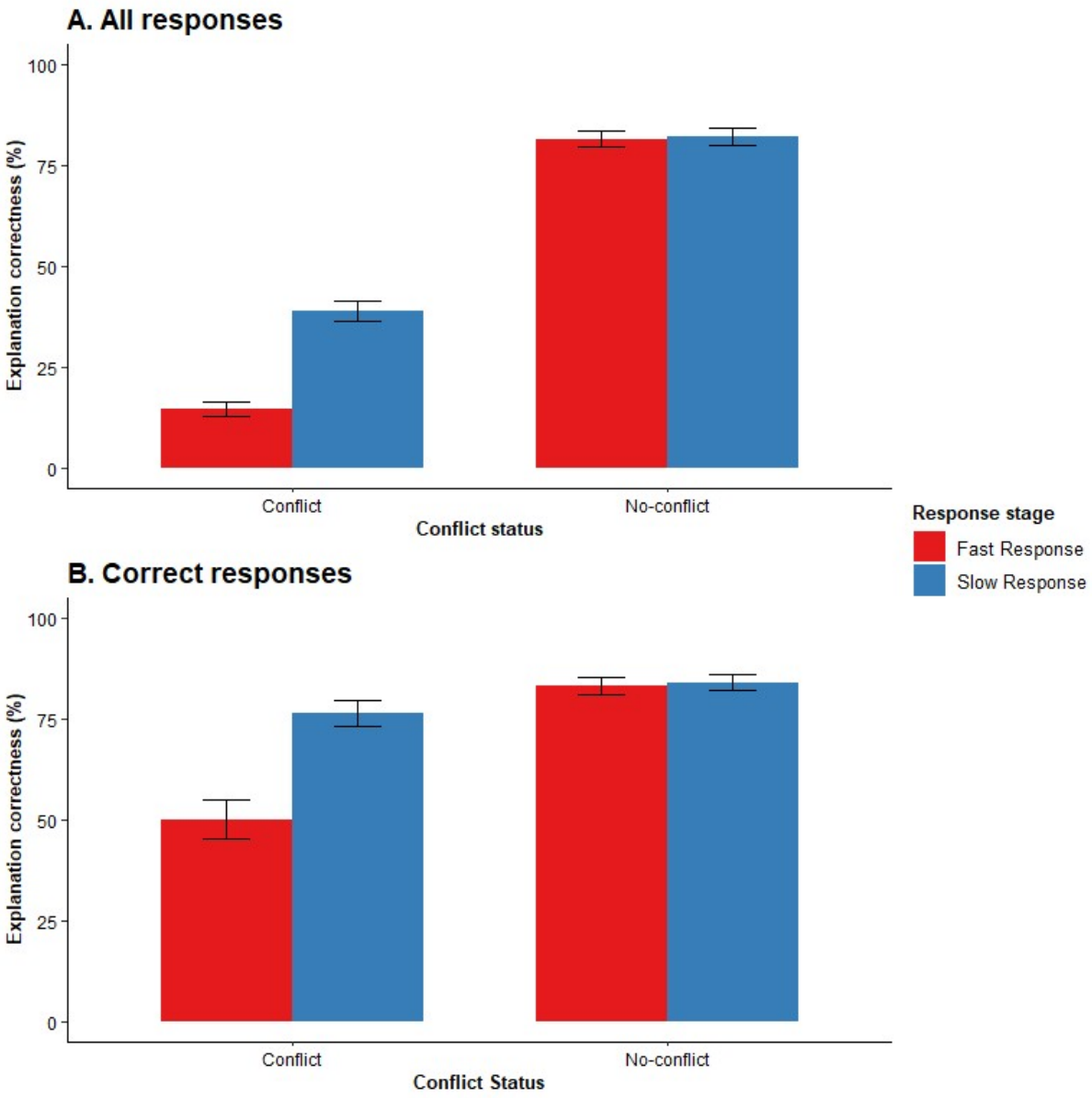

Figure S2. Mean proportion of correct explanations after fast and slow responses overall (A) and for correct responses only (B), in conflict and no-conflict problems all tasks combined. Error bars represent the Standard Error of the Mean.

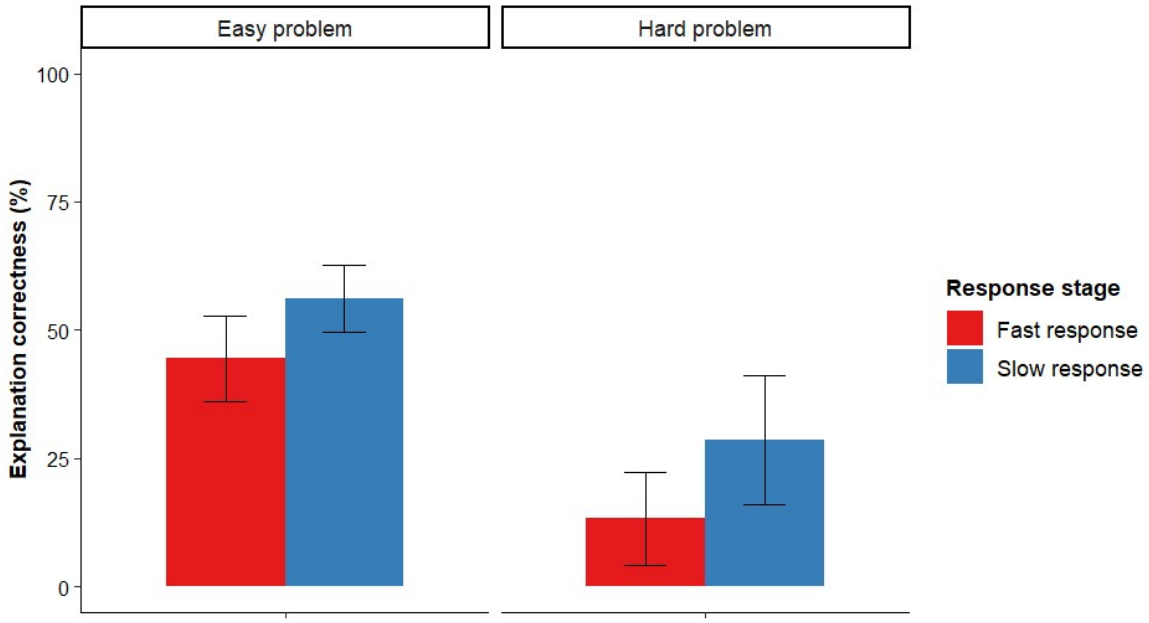

Figure S3. Proportion of correct explanations by response stage and problem difficulty for correct responses in risky-choice conflict problems (Study 3). Error bars represent the Standard Error of the Mean.

Explanation categories distribution

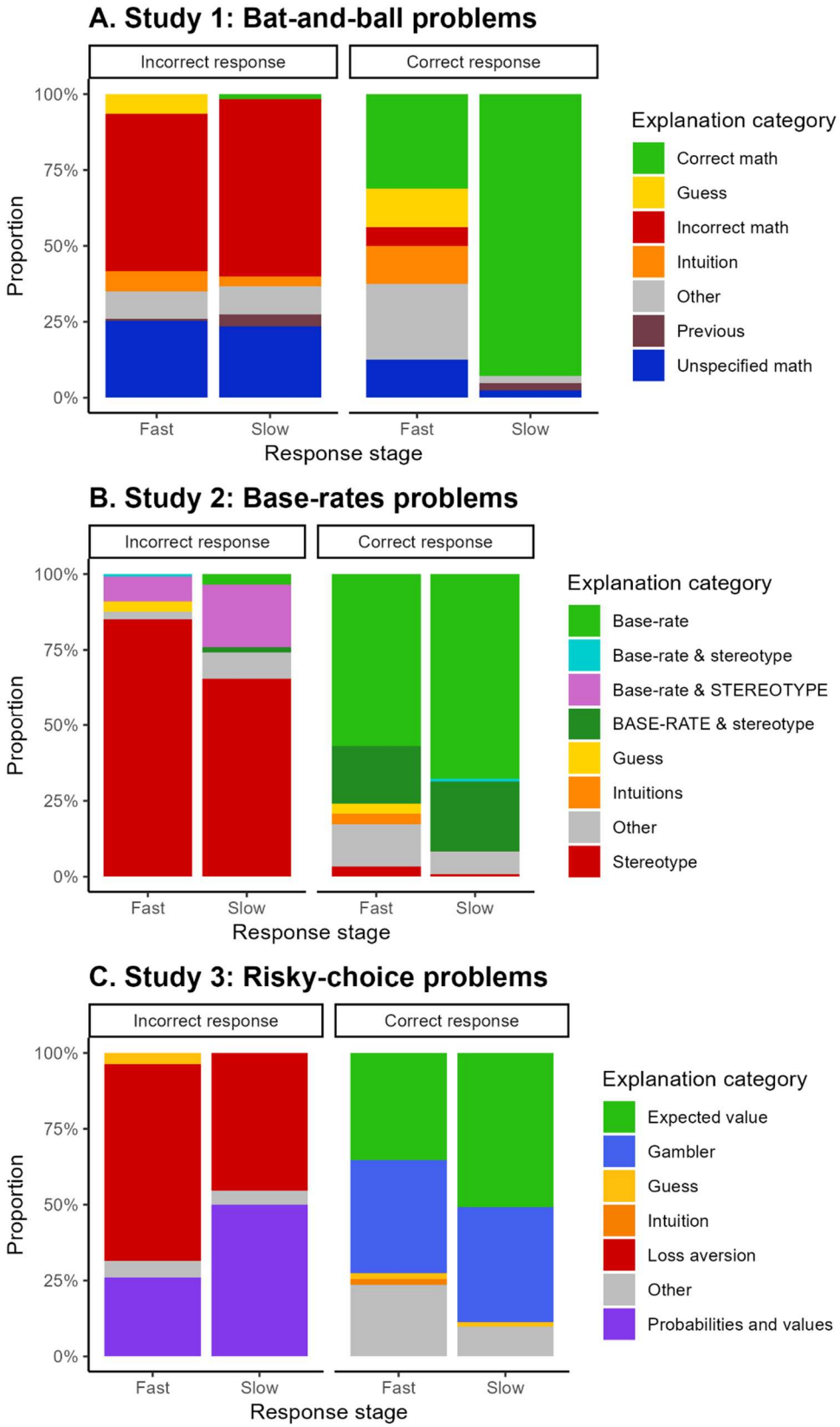

Figure S4. Proportion of each explanation category by response stage and accuracy, for conflict problems, at each task.

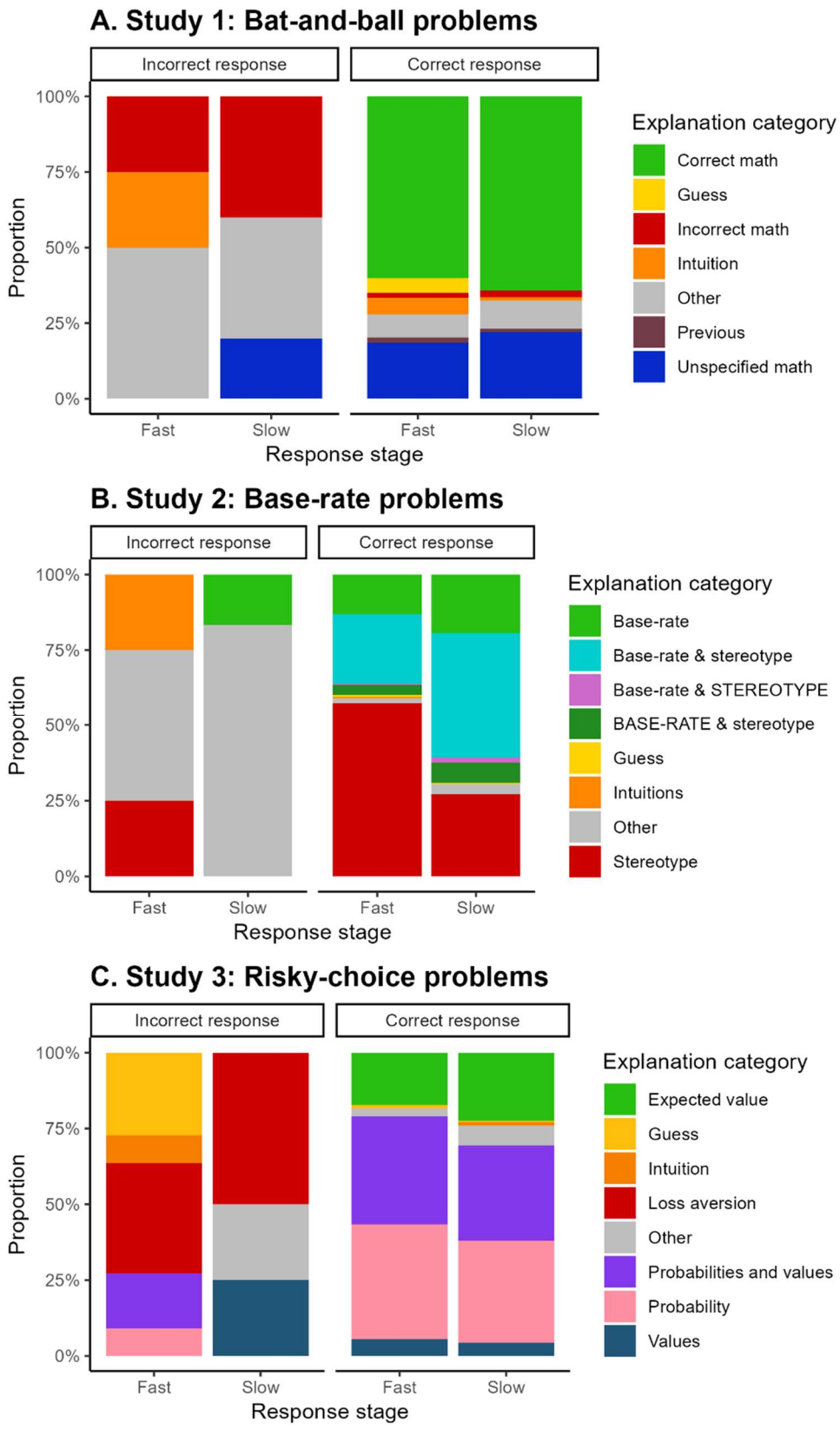

Figure S5. Proportion of each explanation category by response stage and accuracy, for no-conflict problems, at each task.

Explanation explicitness

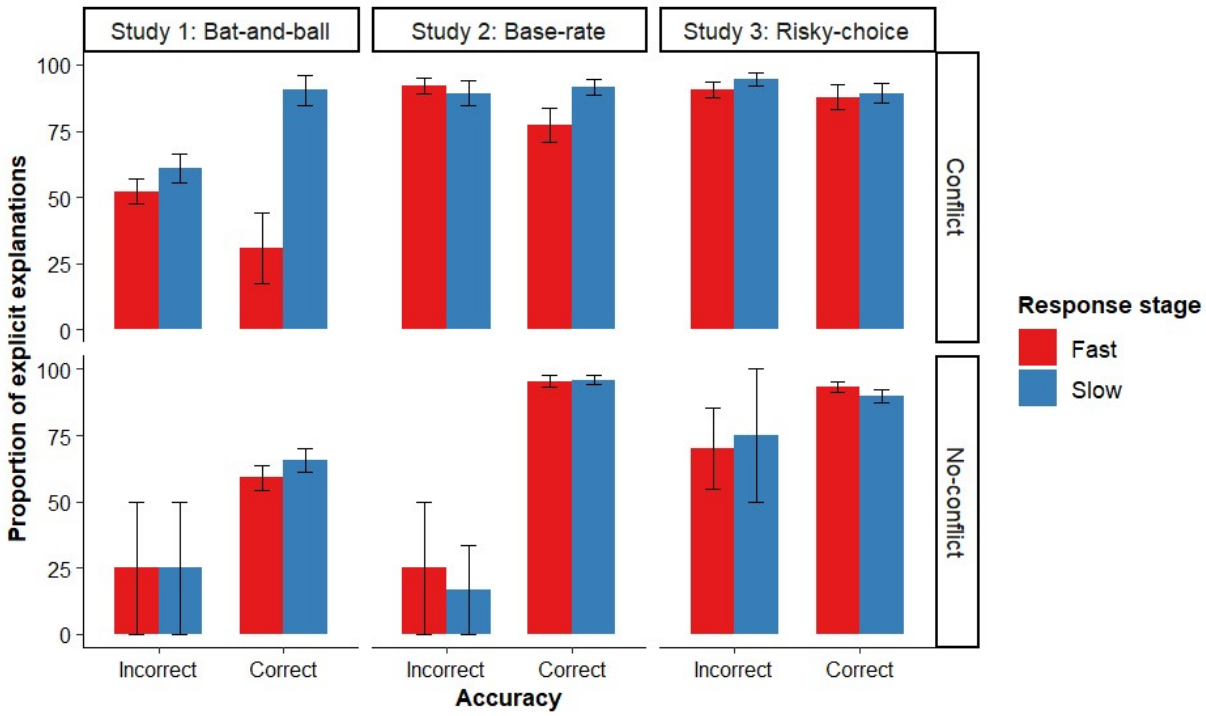

Figure S6: Proportion (%) of explicit explanations by response stage and accuracy, for conflict and no-conflict problems at each task. Error bars represent the Standard Error of the Mean.

In an exploratory analysis we examined the level of explicitness of the explanations (e.g., an unspecified generic explanation ("I just felt it") as opposed to a more specific explicit explanation ("it's 10 cents because $1.10 - $1 is 10 cents") after fast and slow trials. In Study 1 with bat-and-ball problems, the explanations from the categories "Correct math" and "Incorrect math" were labelled as explicit. In Study 2 with base-rate problems, the explanations from the categories "Base-rate", "Stereotype", "BASE-RATE & Stereotype", "Base-rate & STEREOTYPE" and "Base-rate & Stereotype" were labelled as explicit. In Study 3 with risky-choice problems, the explanations from the categories "Expected value", "Gambler", "Probabilities and values", "Probabilities", "Values", and "Loss aversion" were labelled as explicit.

Confidence

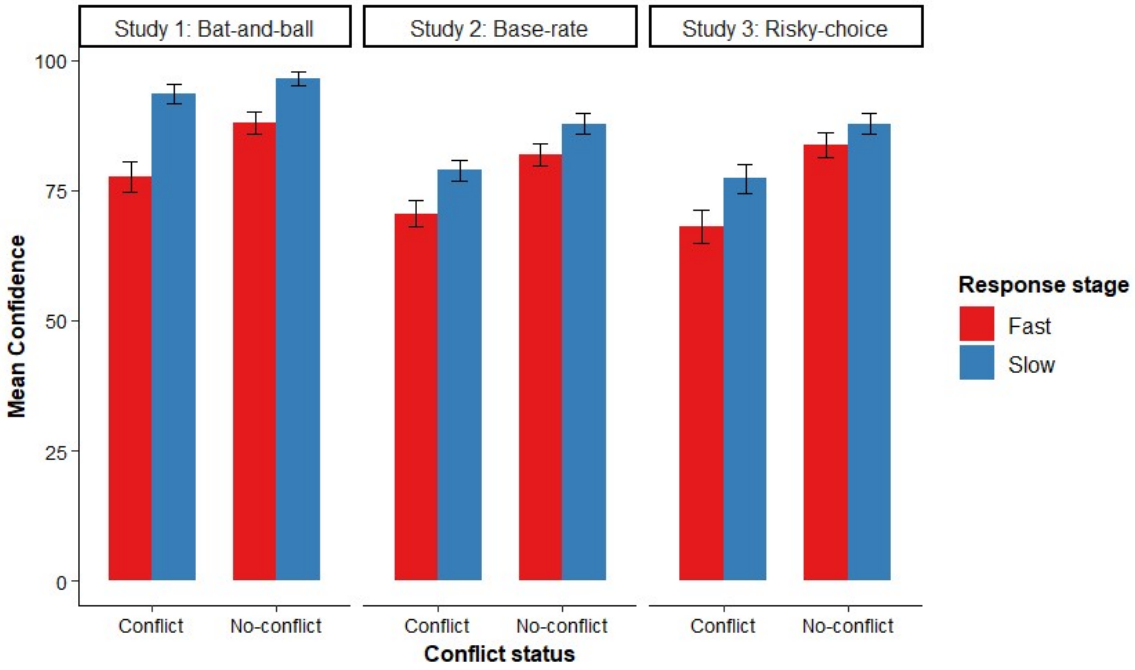

Figure S7: Mean confidence by response stage for conflict and no-conflict problems, at each task. Error bars represent the Standard Error of the Mean.

Explanation and confidence

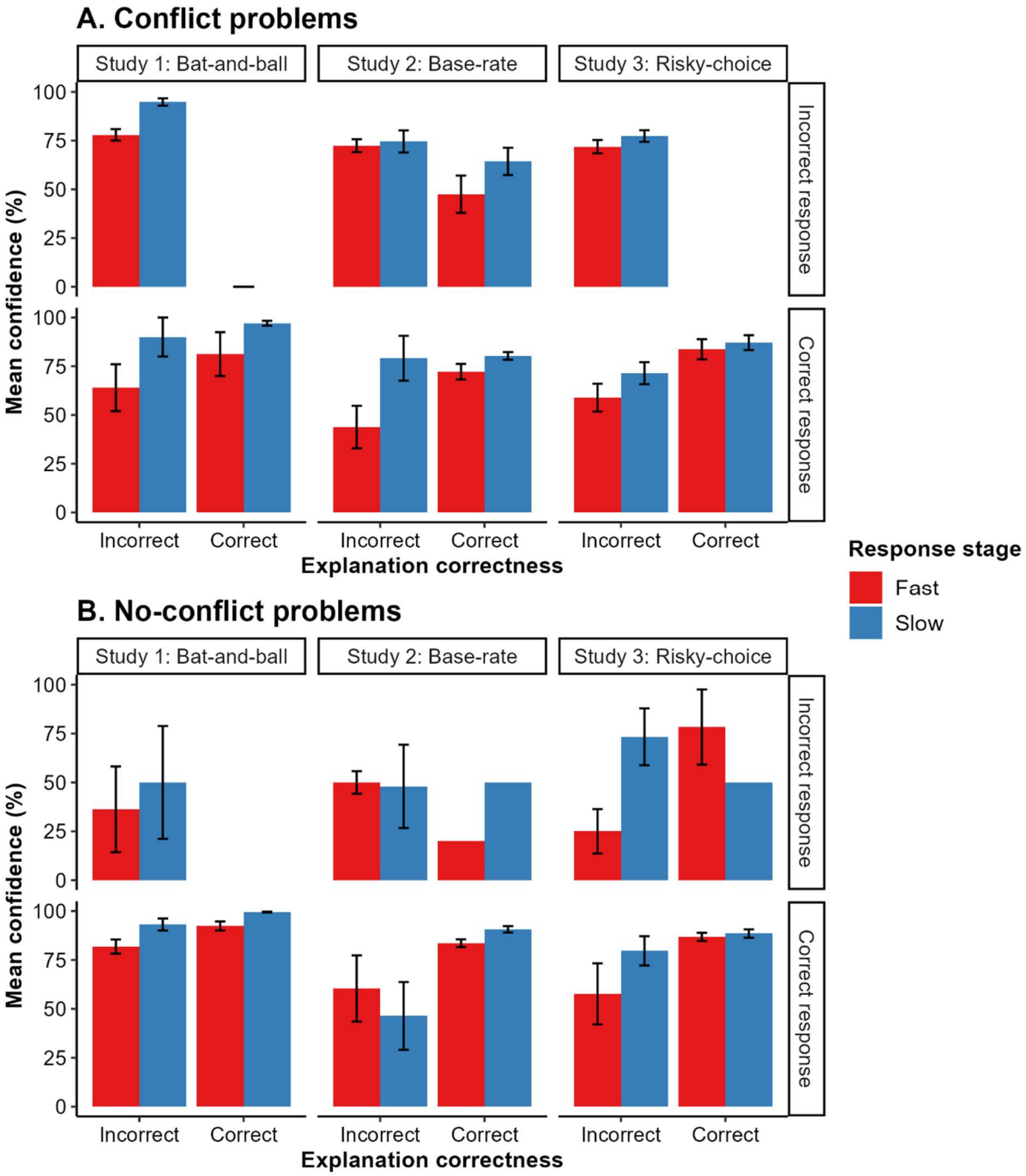

Figure S8. Mean confidence per response accuracy and explanation correctness in fast and slow trials, for conflict (A) and no-conflict problems (B), at each task. Error bars represent the Standard Error of the Mean.

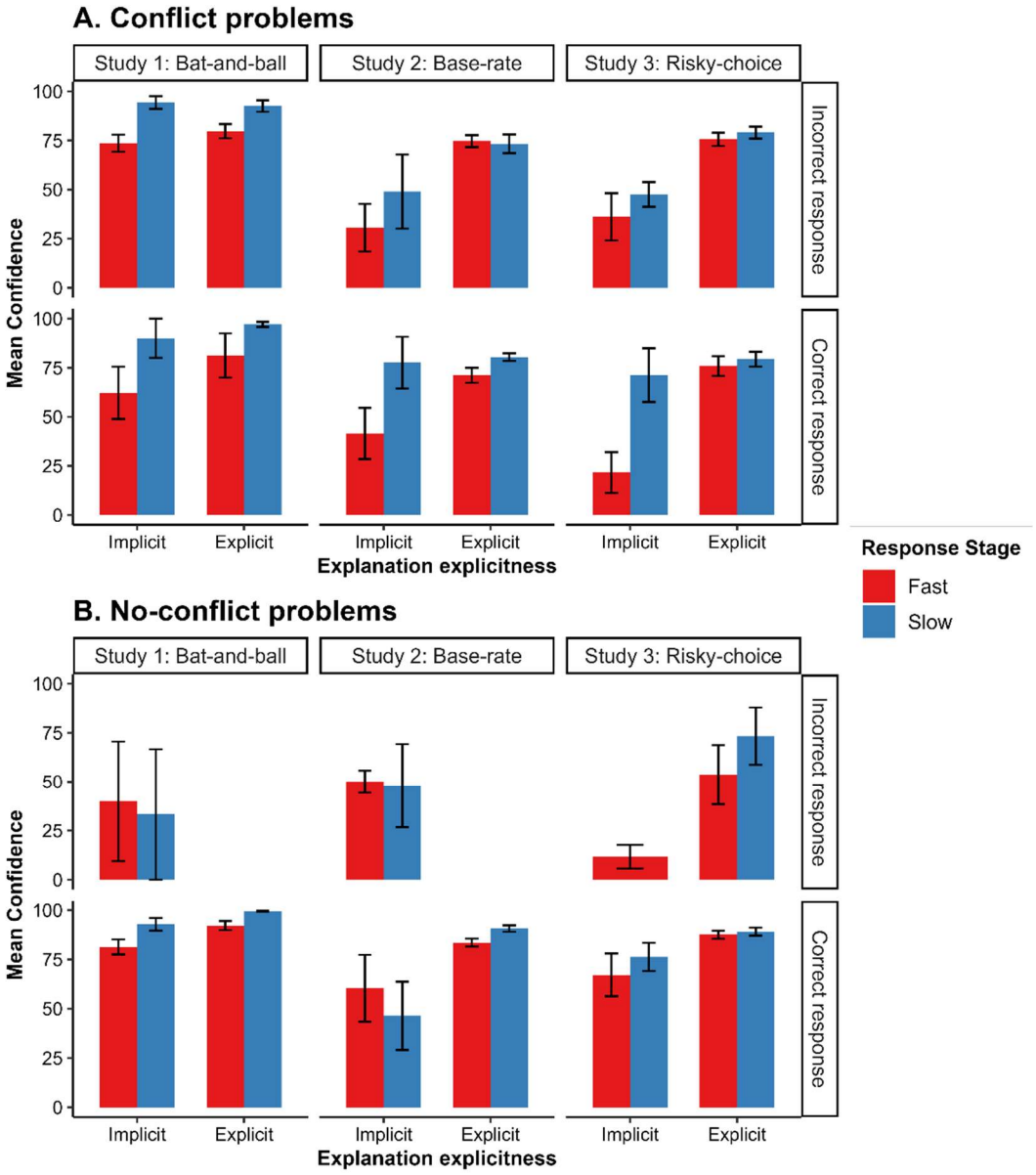

Figure S9. Mean confidence by response accuracy and explanation explicitness in fast and slow trials, for conflict (A) and no-conflict problems (B), at each task. Error bars represent the Standard Error of the Mean.

B. Mixed-effects models

Response Accuracy

Study 1 (Bat-and-ball problems)

Table S1. Generalized mixed-effects logistic regression model of response accuracy based on response stage in conflict bat-and-ball problems (Study 1).

Accuracy ~ 1 + response stage + (1 | subject)

Predictor	Estimate	Std. Error	z value	p-value
(Intercept) [fast]	-12.73	1.85	-6.87	**< .001**
Response stage [slow]	4.93	1.12	4.41	**< .001**

Study 2 (Base-rate problems)

Table S2. Generalized mixed-effects logistic regression model of response accuracy based on response stage in conflict base-rate problems (Study 2).

Accuracy ~ 1 + response stage + (1 | subject) + (1 | item)

Predictor	Estimate	Std. Error	z value	p-value
(Intercept) [fast]	-1.03	.34	-3.04	**< .01**
Response stage [slow]	2.12	.32	6.58	**< .001**

Study 3 (Risky-choice problems)

Table S3. Generalized mixed-effects logistic regression model of response accuracy based on response stage in conflict risky-choice problems (Study 3).

Accuracy ~ 1 + response stage + (1 | subject)

Predictor	Estimate	Std. Error	z value	p-value
(Intercept) [fast]	- 1.15	.92	-1.24	.21
Response stage [slow]	.83	.29	2.85	**< .01**

Explanations

Study 1 (Bat-and-ball problems)

Table S4. Generalized mixed-effects logistic regression model of explanation correctness based on response stage in conflict bat-and-ball problems (Study 1), for all responses.

Explanation correctness ~ 1 + response stage + (1 | subject) + (1 | item)

Predictor	Estimate	Std. Error	z value	p-value
(Intercept) [fast]	- 23.47	3.59	-6.54	**< .001**
Response stage [slow]	13.10	2.81	4.67	**< .001**

Table S5. Generalized mixed-effects logistic regression model of explanation correctness based on response stage in conflict bat-and-ball problems (Study 1) after removing participants familiar with the classical bat-and-ball problem, for all responses.

Explanation correctness ~ 1 + response stage + (1 | subject)

Predictor	Estimate	Std. Error	z value	p-value
(Intercept) [fast]	- 28.29	8.91	-3.17	**< .01**
Response stage [slow]	15.85	7.80	2.03	**< .05**

Table S6. Generalized mixed-effects logistic regression model of explanation correctness based on response stage in conflict bat-and-ball problems (Study 1), for correct responses.

Explanation correctness ~ 1 + response stage + (1 | subject)

Predictor	Estimate	Std. Error	z value	p-value
(Intercept) [fast]	- 8.08	2.98	-2.72	**< .01**
Response stage [slow]	17.88	5.13	3.49	**< .001**

Table S7. Generalized mixed-effects logistic regression model of explanation correctness based on response stage in conflict bat-and-ball problems (Study 1) after removing participants familiar with the classical bat-and-ball problem, for correct responses.

Explanation correctness ~ 1 + response stage + (1 | subject)

Predictor	Estimate	Std. Error	z value	p-value
(Intercept) [fast]	-11.85	6.25	-1.90	.058
Response stage [slow]	24.93	9.75	2.56	**< .05**

Study 2 (Base-rate problems)

Table S8. Generalized mixed-effects logistic regression model of explanation correctness based on response stage in conflict base-rate problems (Study 2), for all responses.

Explanation correctness ~ response stage + (1 | subject) + (1 | item)

Predictor	Estimate	Std. Error	z value	p-value
(Intercept) [fast]	- 2.06	.72	-2.85	**< .01**
Response stage [slow]	4.25	.73	5.83	**< .001**

Table S9. Generalized mixed-effects logistic regression model of explanation correctness based on response stage in conflict base-rate problems (Study 2), for correct responses.

Explanation correctness ~ response stage + (1 | subject) + (1 | item)

Predictor	Estimate	Std. Error	z value	p-value
(Intercept) [fast]	11.70	2.10	5.56	**< .001**
Response stage [slow]	12.99	3.49	3.73	**< .001**

Study 3 (Risky-choice problems)

Table S10. Generalized mixed-effects logistic regression model of explanation correctness based on response stage in conflict risky-choice problems (Study 3), for all responses.

Explanation correctness ~ response stage + (1 | subject) + (1 | item)

Predictor	Estimate	Std. Error	z value	p-value
(Intercept) [fast]	- 3.44	1.24	-2.78	**< .01**
Response stage [slow]	1.28	.42	3.06	**< .01**

Table S11. Generalized mixed-effects logistic regression model of explanation correctness based on response stage in conflict risky-choice problems (Study 3), for correct responses.

Explanation correctness ~ response stage + (1 | subject) + (1 | item)

Predictor	Estimate	Std. Error	z value	p-value
(Intercept) [fast]	- 1.10	.65	-1.68	.09
Response stage [slow]	.72	.45	1.60	.11

Table S12. Generalized mixed-effects logistic regression model of explanation correctness based on response stage in conflict risky-choice problems (Study 3) when including justifications of the "Gambler" category as correct, for all responses.

Explanation correctness ~ response stage + (1 | subject) + (1 | item)

Predictor	Estimate	Std. Error	z value	p-value
(Intercept) [fast]	-2.05	1.16	-1.77	.08
Response stage [slow]	1.33	.36	3.70	**< .001**

Table S13. Generalized mixed-effects logistic regression model of explanation correctness based on response stage in conflict risky-choice problems (Study 3) when including justifications of the "Gambler" category as correct, for correct responses.

Explanation correctness ~ response stage + (1 | subject) + (1 | item)

Predictor	Estimate	Std. Error	z value	p-value
(Intercept) [fast]	.87	.59	1.46	.14
Response stage [slow]	1.09	.52	2.08	**< .05**

Table S14. Generalized mixed-effects logistic regression model of explanation correctness based on response stage and problem difficulty in conflict risky-choice problems (Study 3), for correct responses.

Explanation correctness ~ response stage + problem difficulty (1 | subject)

Predictor	Estimate	Std. Error	z value	p-value
(Intercept) [fast]	-.39	.40	-.97	.33
Response stage [slow]	.71	.46	1.52	.13
Problem difficulty [hard]	-1.65	.63	-2.61	**< .01**

Confidence

Study 1 (Bat-and-ball problems)

Table S15. Linear mixed-effects regression model of confidence based on response stage and accuracy in conflict bat-and-ball problems (Study 1), for all responses.

Confidence ~ 1 + response stage + accuracy + (1 | subject)

Predictor	Estimate	Std. Error	t value	p-value
(Intercept) [fast]	77.73	2.24	34.66	**< .001**
Response stage [slow]	15.23	2.23	6.84	**< .001**
Accuracy [correct]	3.15	3.90	.81	.42

Study 2 (Base-rate problems)

Table S16. Linear mixed-effects regression model of confidence based on response stage and accuracy in conflict base-rate problems (Study 2), for all responses.

Confidence ~ 1 + response stage + accuracy + (1 | subject)

Predictor	Estimate	Std. Error	t value	p-value
(Intercept) [fast]	69.93	2.45	28.53	**< .001**
Response stage [slow]	7.60	2.34	3.24	**< .01**
Accuracy [correct]	1.59	2.78	.57	.57

Study 3 (Risky-choice problems)

Table S17. Linear mixed-effects regression model of confidence based on response stage and accuracy in conflict risky-choice problems (Study 3), for all responses.

Confidence ~ 1 + response stage + accuracy + (1 | subject) + (1 | item)

Predictor	Estimate	Std. Error	t value	p-value
(Intercept) [fast]	68.08	3.78	18.03	**< .001**
Response stage [slow]	9.26	2.59	3.57	**< .001**
Accuracy [correct]	.36	3.44	.11	.92

C. Accuracy

Response Accuracy

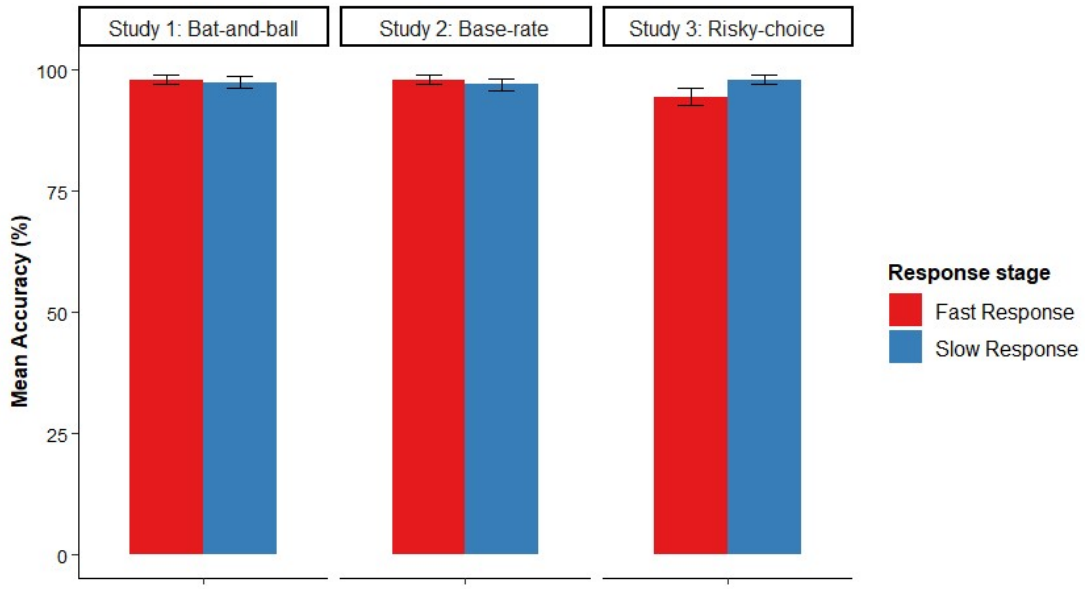

Figure S10. Mean accuracy (%) in fast and slow trials for no-conflict problems, at each task. Error bars represent the Standard Error of the Mean.

Direction of change

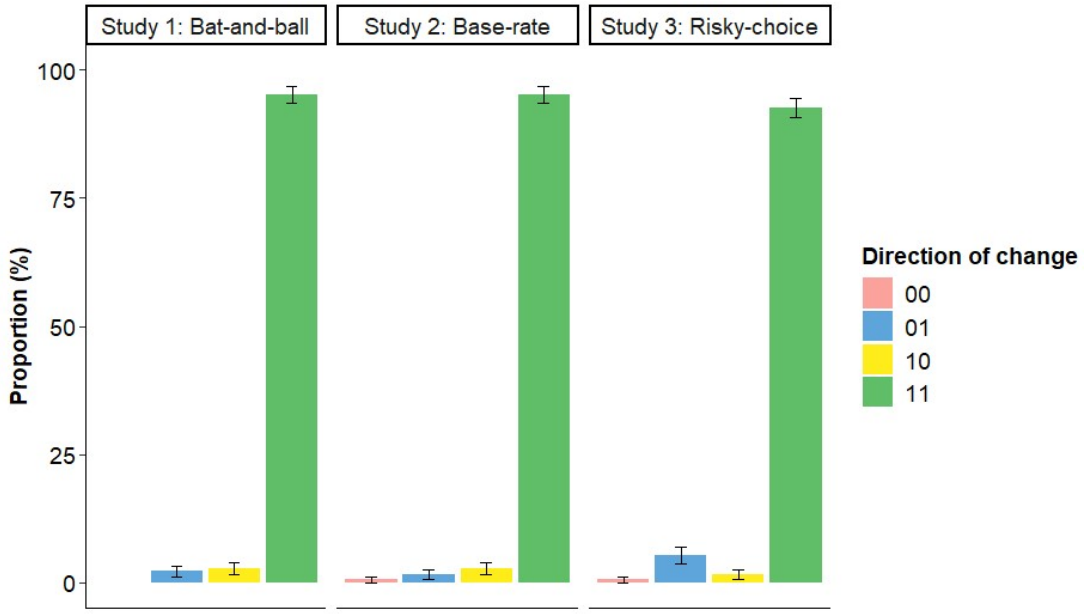

Figure S11. Proportion of each direction of change category for no-conflict problems, at each task. Error bars represent the Standard Error of the Mean.