Preprint - Memory & Cognition, 2024

Easy-fix attentional focus manipulation boosts the intuitive and deliberate use of base-rate information

Esther Boissin[1*], Serge Caparos[2,3], John Abi Hana[4], Cyann Bernard[5], Wim De Neys[4]

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

[2] Université Paris 8, DysCo lab, Saint-Denis, France

[3] Institut Universitaire de France, Paris, France

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

[5] Université Paris Cité, CNRS, Integrative Neuroscience and Cognition Center, Paris, France

*Corresponding author:
Esther BOISSIN
boissinesther@gmail.com,
Department of Psychology,
Cornell University,
211 Tower Rd,
Ithaca, NY,
United States,
14850

Abstract

In addressing human reasoning biases, "easy-fix" attentional focus interventions have shown that we can prompt reasoners to align responses with logico-mathematical principles. The current study aimed to test the impact of such interventions on both intuitive and deliberate responses on base-rate items. Using a two-response paradigm, participants provided initial intuitive responses under time constraints and cognitive load, followed by deliberate responses. During the intervention, we used attentional focus manipulations with base-rate items that aimed to redirect participants' attention toward the 'logical' base-rate cue (i.e., the logical intervention) or toward the 'heuristic' descriptive cue (i.e., the heuristic intervention). The results indicate that the logical intervention led to improved alignment with logico-mathematical principles in both intuitive and deliberate responses, albeit with a modest effect size. Conversely, the heuristic intervention had no discernible impact on accuracy. This indicates that our attentional focus manipulation is more effective at getting reasoners to respect rather than to override base-rates.

Keywords: heuristics & biases; debiasing; easy fix; intuitions

Introduction

Decades of research in decision-making have revealed our propensity for biased reasoning, often influenced by intuitive thinking. For example, when presented with a scenario with 995 females and 5 males, where a randomly selected person is described as liking motorcycles and hunting, many individuals tend to intuitively assume the person is male, guided by stereotypical associations. If we solely consider the description, that assumption may appear reasonable. In general, there might be more males than females who like motorcycles and hunting. However, some females also have these interests, and there were far more females (995) than males (only 5) in the sample. The base-rate information should, therefore, push the scale to the "female" side. Yet, untrained individuals typically neglect the base-rate principle and opt for the intuitive response prompted by their stereotypical prior beliefs (e.g., Kahneman, 2011).

Intuitive or so-called "heuristic" thinking can be useful because it is fast and effortless and can often provide valid problem solutions, however it can also cue responses that conflict with more logical or probabilistic principles. In the framework of

dual-process theories, responses aligning with logico-mathematical principles are traditionally considered as requiring the engagement of the deliberative system to correct the first erroneous intuition (Evans & Stanovich, 2013; Kahneman, 2011; Sloman, 1996). Unlike its intuitive counterpart, the deliberative system requires the commitment of time and cognitive resources to operate. However, since most people tend to minimize demanding computations (Kahneman, 2011), they make extensive use of their intuitions (Evans, 2010; Kahneman, 2011; Stanovich & West, 2000). Consequently, the rate of biased response is substantial, and the issue of remediating people's biased thinking is currently a matter of great interest (e.g., Boissin et al., 2021, 2022, 2023b; Bourgeois-Gironde & Van der Henst, 2009; Claidière et al., 2017; Franiatte et al., 2024; Hoover & Healy, 2017, 2021; Isler et al., 2020; Isler & Yilmaz, 2022; Mata, 2020; Mata et al., 2013; Morewedge et al., 2015; Trouche et al., 2014).

Such "debiasing" work has already shown that a short, single-shot explanation about the intuitive bias and the correct solution strategy often helps reasoners to solve structurally similar problems afterwards. Specifically, in a series of recent studies, participants received explanations regarding the correct response. These brief explanations included both an explicit description of the logical principles and how to apply them, and an explicit statement that clarified the heuristic response was incorrect (see Boissin et al., 2022; Franiatte et al., 2024, for a detailed description of the intervention and in Supplementary Material Section A for an example.)

Consequently, explaining the correct solution to reasoners results in accurate subsequent intuitive responses (Boissin et al., 2021, 2022, 2023a; Franiatte et al., 2024). These studies used the two-response paradigm (Thompson et al., 2011) to differentiate intuitive from deliberate responses. In this paradigm, participants are asked to give two consecutive responses to a reasoning problem. First, they respond as fast as possible with the initial intuitive hunch that comes to mind. Next, they can take all the time they need to reflect on the problem and give a final deliberate response. To make maximally sure that the initial response is generated intuitively, the response needs to be given under time pressure and/or cognitive load (i.e., resources that are critical to engage in deliberation, Bago & De Neys, 2017). Overall, the two-response results showed that debias training typically already boost correct responses in the initial, intuitive response stage (Boissin et al., 2021, 2022, 2023a; Franiatte et al., 2024).

These debiasing approaches show great promise as they significantly increase performance at the intuitive level. They have the added advantage of being easily scalable, requiring no teacher intervention, and can be completed within a brief 5 to 10-minutes timeframe (Boissin et al., 2021, 2022; Franiatte et al., 2024), even for individuals with limited education (Boissin et al., 2024). However, despite the brevity of these single-shot interventions (Boissin et al., 2021, 2022, 2023a; Franiatte et al., 2024), they may require specific conditions for optimal effectiveness. A recent study indicates that restricting the time allocated and the engagement of cognitive resources during the explanation phase reduces the intervention effect on both intuitive and deliberate responses (Boissin et al., 2023a). This highlights the need for a minimal investment of time and cognitive resources to process the explicit instructions when being provided with correct solution explanations during training. Nonetheless, in some situations, we may lack the necessary time or cognitive resources to allocate to such interventions. For instance, noisy environments like busy classrooms or multitasking scenarios can compromise the effectiveness of intervention aiming at providing explanations about the correct solution. Thus, it is crucial to explore whether low-cognitive-demand interventions can also be successful in rewiring our intuitions.

Prior studies have shown that some debiasing approaches, often characterized by reduced cognitive demand and implicitness, can effectively boost correct responding. These interventions, frequently termed as ‘easy fix’, have relied on manipulating participants’ attention to improve their performance on subsequent reasoning tasks (Hoover & Healy, 2017, 2021; Isler et al., 2020; Isler & Yilmaz, 2022; Mata, 2020; Mata et al., 2013, 2014). These ‘easy fix’ interventions do not require the intensive engagement typical of more explicit direct explanation instruction. For example, Mata (2020) simply emphasized the critical premise of the Cognitive Reflection Test (CRT; Frederick, 2005) by using underlined font, which led to increased accuracy. In the same vein, Hoover and Healy (2017) showed that solving algebra equations before solving a reasoning task boosts subsequent performance.

While it is generally assumed that such attentional manipulations promote deliberative reasoning and correction of intuitive errors (Isler et al., 2020; Moutier et al., 2006), this has never been directly tested. In the current paper, we address this issue by testing the nature of an ‘easy fix’ attentional manipulation approach and more specifically whether it boosts deliberative or intuitive reasoning.

In the present study, we aimed to debias base-rate reasoning using an implicit attentional manipulation intervention with base-rate items. Base-rate problems typically contain a descriptive stereotypical cue (e.g., “X likes motorcycles and hunting”) that prompts a heuristic response (“X is more likely a man”), and a base-rate cue (e.g., “X is drawn from a sample consisting of 995 women and 5 men”) that prompts a conflicting logico-mathematical response (“X is more likely a woman”). We attempted to get participants to rely on base-rate information (rather than stereotypical information) by exposing them to “neutralized” problems. In these problems, we replaced the stereotypical description with a neutral one (e.g., “X has a mouth”), forcing participants to rely on the base-rate information to make their decision. A descriptive-neutralized problem would thus read: *“*This study contains 995 women and 5 men. Person ‘X’ has a mouth. Is Person ‘X’ more likely to be a woman or a man?*”*. We assumed that exposed participants to a set of such descriptive-neutralized problems would shift their attentional focus to the point where they would be more likely to rely on base-rate information in typical problems which contain both stereotypical and base-rate information. We name this intervention the “logical intervention”, which was received by half of the participants.

In addition, we also introduced a heuristic intervention, using base-rate-neutralized items. This was strategically employed to examine whether an inverse effect—a reinforcement of intuitive, heuristic-driven responses—might emerge with an attentional intervention using problems in which we neutralized the base-rate information (e.g., “This study contains 500 women and 500 men. Person ‘X’ likes motorcycles and hunting. Is Person ‘X’ more likely to be a woman or a man?”). We wanted to test whether this manipulation would shift participants attention to the descriptive information and potentially decrease reliance on base-rate information in typical problems afterwards. This approach allowed us to explore the distinct effects of these contrasting interventions on participants' reasoning.

To distinguish whether the attentional manipulation affects intuitive or deliberative responding, we used the two-response paradigm before and after the intervention and compared interventions effects between the two groups.

Method

Participants. Participants were recruited online, using the Prolific Academic website (http://www.prolific.ac). Participants had to be native English speakers from Canada, Australia, New Zealand, the United States of America, or the United Kingdom to take part. The same sample size as Boissin et al. (2022) was selected. In total, 101 individuals participated (50 females, M = 37.4 years, SD = 13.0), 52 participants were randomly assigned to the logical intervention group and 49 to the heuristic intervention group. In total, 38 participants had secondary school as their highest level of education, and 63 reported a university degree. We compensated participants for their time at the rate of £5 per hour.

Materials. The study consisted of three blocks presented in the following order: a pre-intervention, an intervention, and a post-intervention block. In total, each participant had to solve six problems during the pre-intervention block, namely, three conflict and three no-conflict problems (see below), and again the same number of problems during the post-intervention block. Additionally, participants solved 10 items during the intervention. All the problems are presented in the Supplementary Material, Section B.

Base rate problems. Base-rate problems were taken from Bago and De Neys (2017). Participants always received a description of the composition of a sample (e.g., “This study contained I.T. engineers and professional boxers”), base rate information (e.g., “There were 995 engineers and 5 professional boxers”) and a description that was designed to cue a stereotypical association (e.g., “This person is strong”). Participants' task was to indicate to which group the person most likely belonged. The task instructions stressed that the person was drawn randomly from the specified sample.

The problem presentation format was based on Pennycook et al.'s (2014) rapid-response paradigm. The base rates and descriptive information were presented serially and the amount of text that was presented on screen was minimized. First, participants received the names of the two groups in the sample (e.g., “This study contains businessmen and firemen”). Next, under the first sentence (which remained on the screen) we presented the descriptive information (e.g., Person ‘K’ is brave). The descriptive information specified a neutral name (‘Person K’) and a single word

personality trait (e.g., “brave”) that was designed to trigger the stereotypical association. Finally, participants received the base rate probabilities. As in Pennycook et al., base rates varied between 995/5, 996/4, and 997/3. The following illustrates the full problem format:

> *This study contains businessmen and firemen.*
> *Person ‘K’ is brave.*
> *There are 996 businessmen and 4 firemen.*
> *Is Person ‘K’ more likely to be:*
> *o A businessman*
> *o A fireman*

Pennycook et al. (2014) pre-tested the material to make sure that words that were selected to cue a stereotypical association consistently did so but avoided extremely diagnostic cues. As Bago and De Neys (2017) clarified, the importance of such a non-extreme and moderate association is not trivial. Note that we label the response that is in line with the base rates as the correct response. Critics of the base rate task (e.g., Gigerenzer et al., 1988; see also Barbey & Sloman, 2007) have long pointed out that if reasoners adopt a Bayesian approach and combine the base rate probabilities with the stereotypical description, this can lead to interpretative complications when the description is extremely diagnostic. For example, imagine that we have an item with males and females as the two groups and give the description that Person ‘A’ is ‘pregnant’. Now, in this case, one would always need to conclude that Person ‘A’ is a woman, regardless of the base rates. The more moderate descriptions (such as ‘kind’ or ‘funny’) help to avoid this potential problem. In addition, the extreme base rates (i.e., 997/3, 996/4, 995/5) that were used in the current study further help to guarantee that even a very approximate Bayesian reasoner would need to pick the response cued by the base-rates (see De Neys, 2014).

Note that Pennycook et al. (2014) created the rapid-response base-rate format with a single word personality trait to reduce reading time (variability) and optimize latency measurement. They showed that the single-word format did not affect accuracy results: people were as biased with their single-word associations as with lengthier descriptions.

In each block, we presented three critical “conflict” items, and three control “no-conflict” items. In the conflict items, the base rate probabilities and the stereotypical

information cued conflicting responses (see example above). In the no-conflict items, they both cued the same response (i.e., the description triggered a stereotypical trait of a member of the largest group). The following is an example of a no-conflict problem:

> *This study contains businessmen and firemen.*
> *Person 'K' is brave.*
> *There are 996 firemen and 4 businessmen.*
> *Is Person 'K' more likely to be:*
> *o A fireman*
> *o A businessman*

These control 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 high accuracy (Bago & De Neys, 2020).

Two sets of 12 unique items (6 pre-intervention and 6 post-intervention block items) were used for counterbalancing purposes. For each block, the conflict problems in one set were the no-conflict problems in the other, and vice-versa (i.e., the base-rates were reversed). Participants were randomly assigned to one of the two sets. Consequently, none of the pre- and post-intervention problem contents was repeated within-subjects (i.e., participants saw a total of 12 different items with a unique stereotypical association).

The presentation order of conflict and no-conflict items was pseudo-randomized both before and after the intervention. Specifically, the first item in the sequence was consistently a conflict item, while the subsequent items were presented in a random order.

Due to a programming error, 23 participants in the heuristic intervention group and 25 of the logical intervention one solved 4 conflict items and 2 no-conflict items after the intervention while 26 participants in the heuristic intervention group and 27 in the logical intervention one solved 2 conflict items and 4 no-conflict items. However, this discrepancy was accounted for in the analyses.

Two-response format. For both the pre- and post-intervention blocks, participants responded to each problem using a two-response procedure, where they first provided a 'fast' answer, directly followed by a second 'slow' answer (Thompson et al., 2011). This method allowed us to capture both an initial 'intuitive' response, and

then a final 'deliberate' one. To minimize the possibility that deliberation was involved in producing the initial 'fast' response, participants had to provide their initial answer within a strict time limit while performing a concurrent cognitive load task (see Bago & De Neys, 2017, 2020). The load task was based on the dot memorization task (Miyake et al., 2001) given that it had been successfully used to burden executive resources during reasoning tasks (e.g., De Neys, 2006; Franssens & De Neys, 2009). Participants had to memorize a complex visual pattern (i.e., 4 crosses in a 3x3 grid) that was presented briefly before each reasoning problem. After their initial (intuitive) response to the problem, participants were shown four different patterns (i.e., with different matrices of crosses) and had to identify the one that they had memorized (see De Neys, 2006, for more details).

For all base-rate problems, a time limit of 3 seconds was chosen for the initial response, based on previous pre-testing that indicated it amounted to the time needed to read the preambles, move the mouse, and click on a response option (Bago & De Neys, 2017, 2020; Raoelison et al., 2021). The time limit and cognitive load were applied only for the initial response, and not for the final one (see below).

Intervention block. During the intervention block, participants solved ten additional base-rate-like problems without any cognitive or time constraints. The presentation order for these 10 items was randomized, and they were presented serially. Participants were randomly assigned to one of the two intervention groups: the logical intervention group and the heuristic one. These interventions were specifically designed to draw reasoner's attention either towards the base-rate cue in the logical intervention or towards the stereotypical description in the heuristic intervention.

The logical intervention consisted in the presentation of base-rate-like items where the description was designed to avoid eliciting any stereotypical associations. The following is an example of a 'neutralized-description item':

> *This study contains businessmen and firemen.*
> *Person 'K' has two hands.*
> *There are 996 firemen and 4 businessmen.*
> *Is Person 'K' more likely to be:*
> *o A fireman*

o A businessman

In the ‘neutralized-description’ item, the description equally applies to both groups, effectively neutralizing any potential for participants to favor one group over the other by using the description. For instance, both firemen and businessmen are expected to have two hands. Consequently, participants are constrained to rely exclusively upon the base-rate cue to generate their responses. This should lead attentional focus of reasoners toward this specific base-rate information. Repetition of structurally similar problems should encourage consideration of the base-rate cue in subsequent conflict base-rate items, and therefore, induce responses that are more aligned with logico-mathematical considerations.

Conversely, the heuristic intervention consisted in the presentation of 10 ‘base-rate-neutralized’ items. These items were designed to draw participants’ attention toward the descriptive cue at the expense of the base-rate one. The following is an example of a ‘base-rate-neutralized’ item:

This study contains businessmen and firemen.
Person ‘K’ is brave.
There are 500 firemen and 500 businessmen.
Is Person ‘K’ more likely to be:
o A fireman
o A businessman

We used ‘base-rate-neutralized’ item in which the base-rate was equally likely for both groups (e.g., “There are 500 firemen and 500 businessmen”). Given that participants had only the stereotypical description to rely on for their responses, their attention is expected to be directed toward this specific cue. Repetition of structurally similar problems should promote consideration of the descriptive cue in subsequent conflict base-rate items.

Procedure. The experiment was run online using the Qualtrics platform. Participants were instructed that the experiment would take 13-15 minutes and that it demanded their full attention. A general description of the task was presented in which participants were instructed that they would read reasoning problems, for which they would have to provide two consecutive responses. They were told that we were

interested in their very first, initial answer that comes to mind and that—after providing their initial response—they could reflect on the problem and take as much time as they needed to provide a final answer (see Bago & De Neys, 2017, for literal instructions). In order to familiarize themselves with the two-response procedure, they first solved two unrelated practice reasoning problems with a response deadline only. Next, they familiarized themselves with the cognitive load procedure by solving two memorization trials and, finally, they solved the same two reasoning problems as before with the full two-response procedure (i.e., deadline + load on initial response).

Figure 1 shows a typical base-rate trial, which started with the presentation of a fixation cross for 2000ms, followed by the description of the sample (e.g., “This study contains businessmen and firemen”) for 2000ms, and subsequently, by the visual matrix for the cognitive-load task for 2000ms. Afterwards, the descriptive adjective (e.g., “Person ‘K’ is brave”) was presented for 2000ms followed by the full problem which featured the base-rate information (e.g., “There are 996 businessmen and 4 firemen”) and the answer options. At this point participants had 3000ms to choose a response. After 2000ms the background of the screen turned yellow to warn participants that they only had a short amount of time left to answer. If they had not provided an answer before the time limit, they were given a reminder that it is important to provide an answer within the time limit on subsequent trials. Participants were then asked to enter how confident they were with their response (from 0%, absolutely not confident, to 100%, absolutely confident). Then, they were presented with four visual matrices and had to choose the one that they had previously memorized. They received feedback as to whether their memory response was correct. If the answer was not correct, they were reminded that it was important to perform well on the memory task on subsequent trials. Finally, the same reasoning problem was presented again, and participants were asked to provide a final deliberate answer (with no time limit) and, once again, to indicate their confidence level.

After responding 6 items during the pre-intervention, participants were introduced to one of two interventions (either logical or heuristic intervention). They were presented to specific instructions that stated the item format was different such as they only needed to give one response without any time constraint, load task or reporting confidence. At the end of the intervention, all participants were instructed that they were going back to the two-response format with time constraint, load task and having to report confidence in their response.

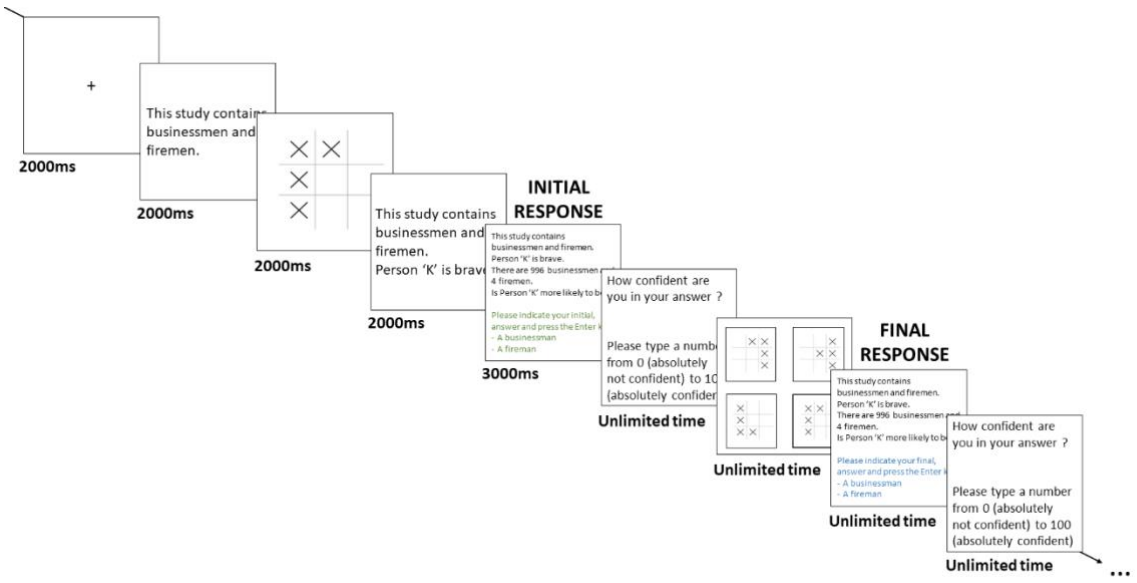

Figure 1. Time course of a complete two-response base-rate item.

Trial exclusion. We discarded trials in which participants failed to provide their initial answer before the deadline (4.7% of all trials) or failed to pick the correct matrix in the load task (11.0% of the remaining trials), and we analysed the remaining 84.8% of all trials. On average, each participant contributed 5.21 (SEM = 0.13) conflict trials out of 6.

Results

Manipulation check. The aim of our study was to test the intuitive or deliberative nature of an intervention that was designed to shift participants' attention, prompting to either a 'logical' response when attention is directed towards a base-rate cue or a heuristic response when attention is directed towards a stereotypical description cue. To determine the effectiveness of our manipulations, we analyzed the participants' reasoning performance during these interventions. Participants in both groups gave a high percentage of expected responses (M logical intervention = 95.2%, SEM = 1.5; M heuristic intervention = 92.2%, SEM = 1.7), suggesting that most participants selected the response prompted by the available cue during the intervention they received[1].

[1] In the logical intervention block, the expected response is the base-rate response while it's the descriptive response in the heuristic intervention block.

Base-rate response accuracy. The following analyses were conducted using R (R Core Team, 2023) with BayesFactor (Morey et al., 2015) and bayestestR (Makowski et al., 2019) packages. We performed Bayesian analyses on generalized linear mixed models on accuracy both for initial—intuitive—and final—deliberate—trials for each group (logical intervention group and heuristic intervention group) with block (pre- vs. post-intervention) as the predictor. We also entered the number of conflict trials in the post-intervention (2 vs. 4) as a predictive factor to control for its effect (analyses without the controlling factor can be found on Supplementary Material Section C. Note that all conclusions are consistent with or without the controlling factor). Participants were entered as random intercepts. For each intervention group and each trial type, we compared the full model, which included the predictor, the controlling factor and the random intercepts, to a null model which included only the controlling factor and the random intercepts. Bayes Factors were reported as BF01 when the null hypothesis (suggesting no difference between the two models, hence no effect of the predictor) was more likely, and as BF10 in favor of the alternative hypothesis (indicating a difference between the models, thus an effect of the predictor). In the results section, we reported both BF01 and BF10 depending on the hypotheses tested. Using one or the other, depending on the context, enhances the relevance and precision of our conclusions. This approach clarifies which hypothesis is supported by the data, avoids any ambiguity, and aligns with best practices in Bayesian analysis. To interpret the resulting BFs, we adopted the classification scale proposed by Jeffreys (1961) as cited in Lee and Wagenmakers (2014); a BF of 1 indicates no evidence, anecdotal evidence between 1 and 3, substantial evidence between 3 and 10, strong evidence between 10 and 30, very strong evidence between 30 and 100 and extreme evidence for BF greater than 100. Consequently, BFs between 1/3 and 3 were interpreted as inconclusive. Note that the readers can also find traditional inferential analyses in Supplementary Material Section D.

First, focusing on accuracy of the final—deliberate—responses in conflict trials. Figure 2 shows that pre-intervention performance for both intervention groups is relatively low (51.0% for the logical intervention group and 54.5% for the heuristic intervention group). The proportion of incorrect final responses before the logical and heuristic interventions was high, at 67.7% and 69.6%, respectively. These proportions are significantly different from the chance level (BF10 for both intervention groups are greater than 100), indicating that participants are not responding randomly but have a

distinct tendency to respond correctly when they can deliberate. This replicates previous findings using the traditional one-response paradigm, where participants are asked to give only one response (Pennycook et al., 2015).

Regarding our main hypothesis, Figure 2 shows a notable pre- to post-intervention improvement among participants who received the logical intervention. Specifically, their accuracy increased from 51.0% (SEM = 6.3) pre-intervention to 65.9% (SEM = 5.9) post-intervention. This pattern provides very strong evidence for the effectiveness of the logical intervention, with a BF10 of 125.20. This suggests that redirecting participants' attention towards logico-mathematical cues significantly improves the likelihood of choosing the correct response in subsequent trials. Conversely, participants who received the heuristic intervention showed similar pre- and post-intervention final-response accuracies (pre-intervention: M = 54.5%, SEM = 6.3; post-intervention: M = 58.2%, SEM = 6.6), reflected by a BF01 of 2.05. This suggests that there is no strong evidence that driving attention towards heuristic cues significantly altered participants' performance in the final, deliberate responses of the participants.

With respect to the accuracy of initial—intuitive—responses in conflict trials, a similar pattern emerged. Figure 2 shows that performance before the intervention was low, with the proportion of incorrect initial intuitive responses before the logical and heuristic interventions being high at 70.2% and 68.0%, respectively. These proportions are different from the chance level (BF10 for both intervention groups are greater than 100), indicating that participants are not responding randomly but have a distinct tendency to rely on stereotypical responses during the intuitive stage. Regarding our main hypothesis, the logical intervention group showed a post-intervention increase in accuracy (M pre-intervention = 30.4%, SEM = 6.0 and M post-intervention = 41.3%, SEM = 6.0). The high BF (BF10 = 47.71) strongly corroborates this finding, suggesting that even at an intuitive stage, the logical attentional intervention boosted sound intuitive reasoning. Conversely, in alignment with final-response accuracies, post-intervention accuracies little varied from pre- to post-intervention in the heuristic intervention group (pre-intervention: M = 33.3%, SEM = 6.0; post-intervention: M = 37.3%, SEM = 5.9), as indicated by a BF01 of 3.43.

In addition, we analysed the performance for the no-conflict control problems. We observed that performance was consistently at ceiling, with grand means of 96.9% (SEM = 0.8) for initial accuracy, and 98.2% (SEM = 0.6) for final accuracy (See

Supplementary Material Section E). In line with previous studies (Bago & De Neys, 2020; Pennycook et al., 2015; Raoelison et al., 2021), participants' high accuracy rates on the no-conflict problems indicated that they were paying attention to the task and refrained from random guessing.

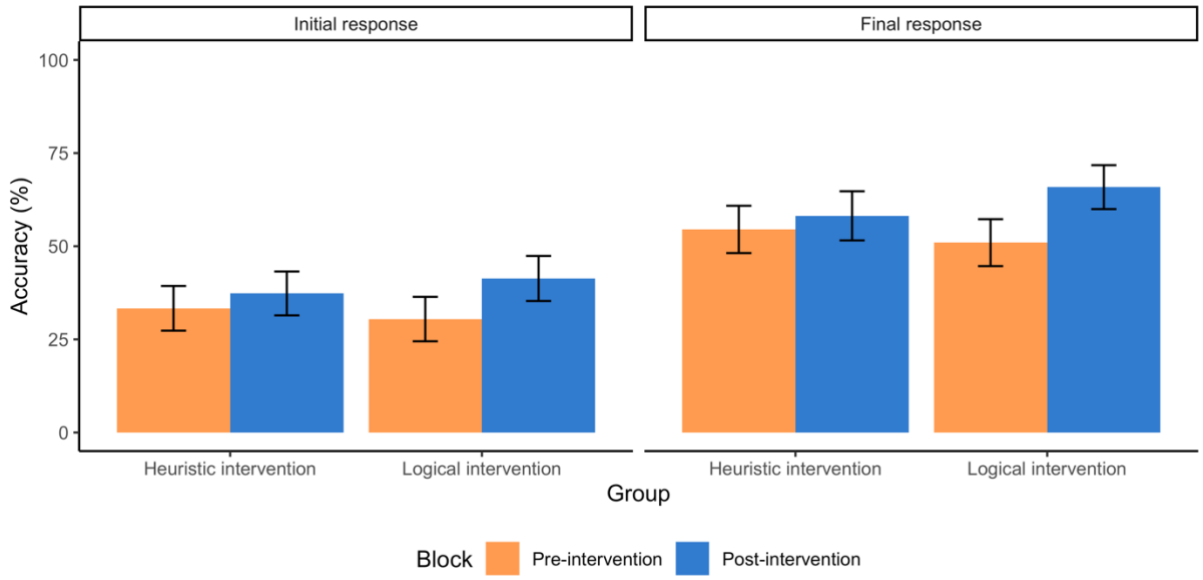

Figure 2. Average initial and final accuracy on conflict problems before and after the heuristic and logical interventions. Error bars represent standard error of the mean (SEM).

Direction of change. To gain some deeper insight into how people changed (or did not change) their response after deliberation, we conducted a direction of change analysis (Bago & De Neys, 2017, 2020). More specifically, for each problem, people could give a correct ('1') or incorrect ('0') response at each of the two response stages (i.e., initial and final). Hence, this can result in four different types of response patterns on any single problem ("00" pattern, incorrect response at both stages; "11" pattern, correct response at both stages; "01" pattern, initial incorrect and final correct response; "10" pattern, initial correct and final incorrect response). This helps us determine if the interventions lead to a significant shift in response patterns. For example, a reduction in the "00" pattern (incorrect response at both stages) and an increase in the "11" pattern (correct response at both stages) would indicate that the intervention is effective in improving both intuitive and deliberate responses. Similarly, an increase in the "00" pattern (initial incorrect and final incorrect response) would suggest that the intervention hampers participants performance both at the intuitive and deliberative stages.

We conducted Bayesian mixed-effects regression analyses on the proportion of patterns (00, 01, and 11) for each intervention group, with block (pre- vs. post-intervention) and number of conflict trials after the intervention (2 vs. 4) as fixed factors and participants as random intercepts (analyses without the controlling factor can be found on Supplementary Material Section C. Note that all conclusions are consistent with or without the controlling factor. Inferential analyses can also be found in the Supplementary Material Section D.)

Figure 3 shows the average proportion of each type of response pattern in conflict trials, before and after the intervention for both intervention groups. As Figure 3 shows, in the logical intervention group, there is a drop in “00” patterns of responses (-17.3 points, $BF10^2 > 100$) that is accompanied by an increase of “01” patterns (+6.4 points, BF10 = 1.26) and of “11” patterns (+8.5 points, BF10 > 100). After the logical intervention, participants tend to produce more “11” response patterns than “01” (BF10 = 1.26). This finding lends support to the hypothesis that the logical intervention boosts subsequent intuitive and deliberate conflict responses rather than merely correcting erroneous intuitive responses. In contrast, the heuristic intervention shows a decrease in “00” patterns (-5.1 points, BF10 = 26.07), no change in “01” patterns (+1 point, the BF in favor of the null hypothesis is BF01 = 6.76), and a slight increase in “11” patterns (+ 2.6 points, BF10 = 2.17). Hence, there is no evidence that correct responding decreased after the heuristic intervention.

[2] Note that the indices ‘10’ or ‘01’ associated with these Bayes Factors (BFs) indicate whether the BFs support the alternative hypothesis (the effect of our main factor is more likely) or the null hypothesis (the effect of our main factor is less likely). These indices are not related to the ‘01’ or ‘10’ terminology used in the direction of change framework.

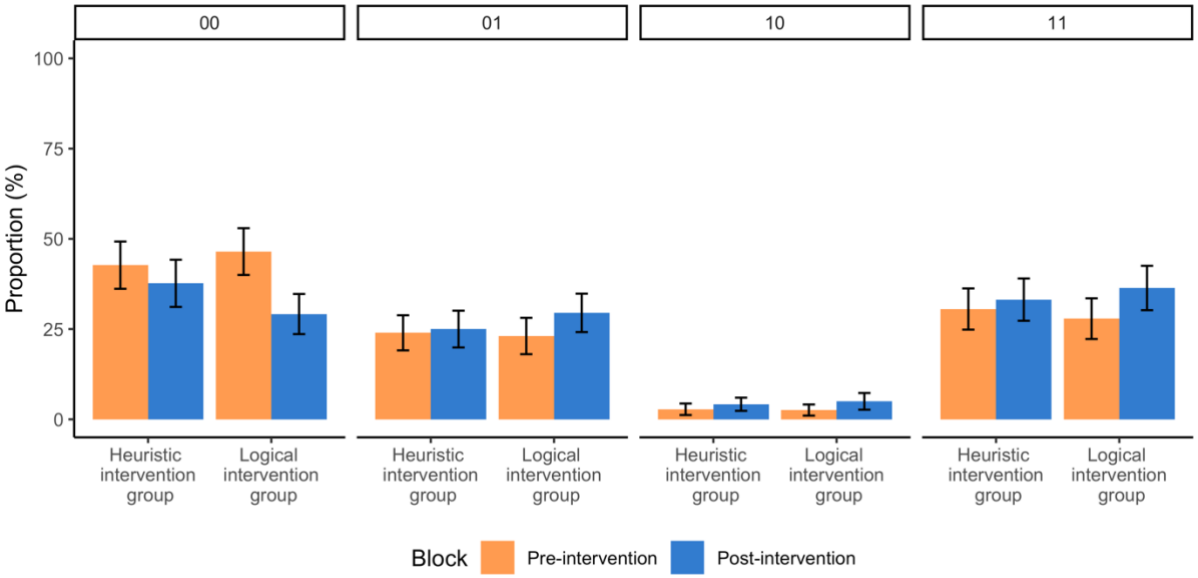

Figure 3. Average proportion of each direction of change (i.e., 00, 01, 10 and 11 response patterns) for the conflict problems according to block (pre- vs. post-intervention) and group (logical intervention vs. heuristic intervention group). Errors bars represent standard errors of the mean (SEM).

Individual training effect classification. To explore further how participants benefited from the intervention (or not), we classified reasoners according to an individual-level accuracy analysis for each participant, examining performance for each conflict problem, from start to end of the experiment. We created four different categories distinguishing participants who improved after the intervention from those who did not improve and from those who already showed accurate reasoning performance before the intervention. These categories were defined by the proportion of each response pattern (i.e., “00”, “01”, “11”, “10”) in each block (i.e., pre-, and post-intervention).

Classification was based on the predominant patterns in a block, except for cases where participants provided an equal number of them. In such cases, the response-pattern for the last conflict problem determined the classification. Therefore, a block was classified as a “00” block if the majority of its conflict problems had “00” response patterns, and similarly for the other types of response patterns (i.e., “01”, “10”, “11”). The change (or lack thereof) in block type from pre- to post-intervention determined each participant’s training classification. For instance, a ‘00-00’ classification indicated a majority of “00” response patterns both before and after the intervention, while a ‘00-11’ classification indicated a shift from “00” patterns (i.e.,

people gave a majority of incorrect responses both at the intuitive and the deliberative stages) pre-intervention to “11” patterns (i.e., people gave a majority of correct responses both at the intuitive and the deliberative stages) post-intervention. Based on these classifications, we established four distinct categories for the participants. Participants categorized as '11-11' (i.e., people gave a majority of correct responses both at the intuitive and the deliberative stages before and after the intervention) and '01-01’ (i.e., people gave a majority of correct responses only at the deliberative stages before and after the intervention) were deemed “correct” reasoners, consistently giving correct responses even without any intervention. Participants classified as '00-11' (i.e., people gave a majority of incorrect responses before the intervention and correct responses both at the intuitive and the deliberative stages after the intervention), '00-01' (i.e., people gave a majority of incorrect responses before the intervention and correct responses only at the deliberative stage after the intervention), and ’01-11’ (i.e., people gave a majority of correct responses only at the deliberative stage before the intervention and a majority of correct responses both at the intuitive and deliberative stages after the intervention) were labeled “improved” reasoners, showing better (intuitive or deliberate) performance after the intervention. Participants consistently providing “00” response patterns were labeled as “biased” reasoners (i.e., people gave a majority of incorrect responses both before and after the intervention). All others, classified as ’01-00’ (i.e., people gave a majority of correct responses only at the deliberative stage before the intervention and a majority of incorrect responses after the intervention), ’11-01’ (i.e., people gave a majority of correct responses both at the intuitive and the deliberative stages before the intervention and a majority of correct responses only at the deliberate stage after the intervention), or ’11-00’ (i.e., people gave a majority of correct responses both at the intuitive and the deliberative stages before the intervention and a majority of incorrect responses after the intervention) and demonstrating decreased performance, were grouped as "decreased".

This classification is helpful as it provides detailed insights into the specific impact of the intervention at the individual (vs. mere group) level. By identifying whether initially biased participants become genuinely intuitive reasoners post-intervention, similar to findings in explicit training (Boissin et al., 2022), or whether the intervention primarily enhances their deliberative responses, we can better understand the nature of the logical intervention’s effectiveness. Table 1 shows that in the logical intervention group, 27% of participants were classified as “biased”, 40% as “correct” (among them,

67% were '11-11' type), 12% as “decreased”, and 21% as “improved” reasoners. Notably, among the improved reasoners, 73% were of the ’00-01’ type and 27% were of the ’01-11’ type. This indicates that those improved respondents who finished by giving predominantly ‘11’ responses after the intervention were participants who were predominantly giving ‘01’ responses before the intervention. This suggests that the logical intervention boosts correct intuitive responding for those who already have correct deliberate responses. Whereas participants who were biased before the intervention benefited from the training at the deliberative stage, rather than at the intuitive one.

In contrast, participants in the heuristic intervention group mostly remained biased. Specifically, 13% of the participants were classified as “improved” reasoners and 36% as “biased” as shown in Table 1. Interestingly, 38% of the participants were classified as “correct” and only 13% as “decreased”, equaling the proportions found in the logical intervention group—further underlining the ineffectiveness of the heuristic intervention to decrease performance.

Table 1. Individual training effect classification proportion both for the heuristic intervention and the logical intervention groups according to the pre-post classification.

		Heuristic intervention group (n = 47)	Logical intervention group (n = 50)
Type of reasoners	Type of classification		
Biased	***Total***	**36% (n = 17)**	**27% (n = 14)**
	00-00	100**% (n = 17)**	100**% (n = 14)**
Improved	***Total***	***13% (n = 6)***	***21% (n = 11)***
	00-01	33% (n = 2)	73% (n =8)
	01-11	67% (n = 4)	27% (n = 3)
Correct	***Total***	***38% (n = 18)***	***40% (n = 21)***
	01-01	33% (n = 6)	33% (n = 7)
	11-11	67% (n = 12)	67% (n = 14)
Decreased	***Total***	***13% (n = 6)***	***12% (n = 4)***
	01-00	33% (n = 2)	33% (n = 2)
	11-01	50% (n = 3)	33% (n = 2)

11-00	17% (n = 1)	-

Conflict detection. Previous studies have shown that, despite giving an incorrect response, reasoners often show some conflict or error sensitivity—as expressed for example in decreased confidence in their erroneous conflict trial responses (e.g., see De Neys, 2022 for review). In the present study, we explored whether the logical intervention affected biased reasoners' ability to detect conflict in base-rate problems. That is, although the intervention might not have succeeded in getting some biased people to reason accurately, it might have helped them to better detect that their answer was incorrect. We used the conflict detection index introduced in the study of De Neys et al. (2011), which contrasts confidence ratings for no-conflict trials that yielded a correct response to confidence ratings for conflict trials that yielded an incorrect response. We compared the conflict-detection index before and after the logical intervention. A higher difference value implies a larger confidence decrease when solving conflict items, which is believed to reflect a more pronounced conflict experience (Bago & De Neys, 2017, 2019; Pennycook et al., 2015). We primarily focused on initial response conflict detection since it gives a purer measure of intuitively experienced conflict (e.g., Bago & De Neys, 2017; Voudouri et al., 2023).

We conducted Bayesian mixed-effects regression analyses on conflict-detection index for initial trials, with block (pre- vs. post-intervention) and number of conflict trials (2 vs. 4) as fixed factors and participants as random intercepts. We found a relatively consistent level of conflict detection before (M = 9.9, SEM = 4.3) and after the logical intervention (M = 10.6, SEM = 4.7) with a moderate BF01 = 3.84. This suggests that directing participants' attention toward logico-mathematical cue does not boost their ability to detect that their heuristic-based conflict solution is not the correct solution.

For completeness, we also compared the conflict-detection indexes of participants in the heuristic intervention and also found it was fairly consistent before (M = 7.6, SEM = 3.4) and after (M = 5.6, SEM = 3.5) the intervention, with a moderate BF01 = 2.92.

Discussion

The present study aimed to investigate the effectiveness of an 'easy fix' intervention designed to reduce biased reasoning. Specifically, we explored whether an attentional intervention could improve intuitive and/or deliberative reasoning performance, focusing on base-rate problems. Consequently, half of participants were introduced to a logical attentional intervention that aimed at implicitly directing attention toward base-rate information and was expected to lead to higher accuracy. To test whether a heuristic attentional intervention would generate the opposite effect, namely, decreasing performance, the other half of participants were presented with a heuristic intervention that implicitly emphasized reliance on stereotypical descriptions.

Results showed that the logical intervention proved to be effective at boosting reasoning performance. This suggests that shifting attention to base-rate cues can indeed improve the accuracy of reasoning in individuals prone to heuristic biases. Interestingly, the logical intervention appeared to promote both initial intuitive and final deliberate correct responses (albeit for different individuals, see further).

Hence, similar to more explicit direct explanation interventions (Boissin et al., 2022; Franiatte et al., 2024), our implicit attentional manipulation succeeded in improving sound intuition and deliberation. Nevertheless, the magnitude of the logical intervention effect, while present, was relatively modest, especially when compared to the more substantial enhancements reported in explicit interventions which aim to provide the rationale behind the logical solution such as those by Boissin and al. (2022) and Franiatte and al. (2024). In these studies, the increase amounted up to 33 points for final responses and 48 points for initial responses, in contrast to the current study, which showed an increase of 11 and 18 points respectively for initial and final responses. This contrast highlights a well-acknowledged challenge inherent to the debiasing of intuitive beliefs, often cited as difficult in the literature (Goel, 2022; Kaplan et al., 2016). Our findings are consistent with this idea, highlighting that implicit interventions may not suffice to effectively replace deeply rooted intuitive biases with more logical responses. In contrast, explicit interventions that provide detailed explanations of the correct solution appear to be more effective in promoting logical reasoning.

In addition, the individual-level classification analysis indicated that participants who benefited from the logical intervention at the intuitive level were already effective

at producing correct deliberate responses prior to the intervention. Conversely, the impact on biased individuals was predominantly observed in their deliberate responses, and less so in their intuitive ones. This suggests that those who were biased before the intervention could only be partially debiased through the attentional manipulation. This highlights that the attentional intervention cannot be considered an ideal solution for the debiasing of intuitive reasoning.

Altogether, these results highlight a significant conundrum in determining the effectiveness of an intervention. While explicit direct interventions have demonstrated significant efficacy in boosting the accuracy of intuitive responses—even among those with pre-intervention biased responses—their success comes at the cost of higher cognitive demands. Conversely, our allegedly less demanding logical intervention, while easier to apply and potentially more scalable, did not yield comparable effects. This dichotomy presents a crucial challenge for future research endeavors, which should aim to strike a balance between interventions that are both effective and cognitively feasible. Addressing this challenge is essential for developing ecological and efficient bias reduction interventions, which can be tailored to diverse contexts and varying cognitive resource availability.

Regarding the heuristic intervention, it yielded unexpected results as it did not significantly alter reasoning accuracy. Participants who received the heuristic intervention did not exhibit a significant decrease in accuracy, indicating that this intervention may not effectively promote stereotypical thinking or biased responses. However, the absence of a significant impact from the heuristic intervention is encouraging. It demonstrates that correct reasoners are not easily swayed. Even when their attention is drawn to heuristic cues, they do not blindly and automatically adhere to them. In other words, the absence of a performance decrease following the heuristic intervention suggests that while we can encourage individuals to respond more logically, we cannot so easily compel them to perform less effectively, even at an intuitive stage.

We speculate that the absence of a significant impact following the heuristic intervention may highlight a fundamental aspect of human reasoning: the resilience of sound reasoning. Individuals who spontaneously align their responses with base-rate principles apparently demonstrate a robust understanding and application of these principles, remaining unaffected by implicit attentional manipulation. Our individual-level classification clearly showed that most reasoners who provided correct responses

before the intervention consistently continued to do so post-intervention, whether relying on intuition or deliberation. This emphasizes that resilience extends across both intuitive and deliberative reasoning. Additionally, the lack of an effect from the heuristic intervention on correct reasoners suggests that they are capable of discerning when a heuristic response deviates from logico-mathematical principles, leading them to avoid such responses. This discernment indicates that correct reasoners reliably avoid erroneous heuristic responses, even when exposed to manipulations meant to specifically elicit such responses. These combined findings emphasize the robust and discerning nature of correct responding, particularly in intuitive reasoning, offering valuable insights into the potential of implicit attentional manipulation approaches.

More broadly, the observed asymmetry in the outcomes of the two attentional focus manipulations presents a considerable advantage for debiasing interventions. This differential effect holds promise, especially for its potential to assist those prone to biases while preserving the performance of sound reasoners. It suggests that interventions can be specifically designed to target and mitigate biased reasoning without compromising the established logical performance of individuals. Nonetheless, we emphasize once again that our reasoning is speculative, and these claims should be interpreted with caution in the absence of direct evidence.

However, our current findings are closely tied to the specific tasks we employed, namely the base-rate items. It is crucial to investigate whether the easy-fix effect can be generalized to other tasks that involve distinct logical principles and heuristics. Additionally, evaluating the impact of these interventions in more ecologically valid contexts is essential, especially where individuals hold entrenched beliefs that are not easily challenged (Goel, 2022; Kaplan et al., 2016).

For example, in the medical field, both patients and healthcare providers often struggle to understand base rates related to disease prevalence and the accuracy of diagnostic tests. This lack of understanding can lead to suboptimal decision-making regarding treatments and screenings, potentially compromising health outcomes (Gigerenzer, 2009). Emphasizing the importance of base-rate information in medical decision-making could therefore play a significant role in improving health outcomes by ensuring more informed choices.

Recent research in the domain of fake news has highlighted the considerable potential of easy-fix interventions in real-life settings. Pennycook et al. (2021) demonstrated that nudging individuals to consider the accuracy of news before sharing

it can significantly reduce the likelihood of spreading fake news on social media. Their studies suggest that prompting individuals to prioritize accuracy can effectively redirect their attentional priorities without increasing the overall amount of deliberation, thus enabling them to intuitively and effortlessly refrain from sharing fake news (Lin et al., 2023).

These findings underscore the potential for easy-fix interventions to enhance decision-making processes across various real-life contexts. Consequently, further research is needed to explore the applicability and effectiveness of these interventions in more naturalistic settings. Such settings may include education, voting, and purchasing decisions, where they could help consumers make more prudent choices, or community safety decisions, where they could encourage better practices regarding personal and public safety.

In summary, while the current study provides valuable insights, future research should aim to extend these findings to a broader range of tasks and more ecologically valid contexts. This will help to better understand the full potential of easy-fix interventions in promoting improved decision-making in everyday life.

To sum-up, our study serves as a proof-of-principle, illustrating the potential of attentional manipulation as a debiasing tool. Nonetheless, this approach requires further validation, refinement, and generalization. By shedding light on the dynamic nature of human reasoning and the potential for attentional manipulations to enhance decision-making, our research contributes to the broader understanding of cognitive biases and their remediation.

Declarations

Funding. This research was supported by a grant from the Agence Nationale de la Recherche (ANR-23-CE28-0004-01 by WDN) and by The Fyssen Foundation (by EB).

Conflicts of interest. The authors report no conflict of interest.

Ethics approval. Ethics approval for this research was obtained from the Research Ethics Committee at the University of Paris Cité.

Consent to participate. All participants provided written informed consent to participate in this study. They were informed about the purpose of the study, the procedures involved, potential risks, and benefits, and were assured of the confidentiality of their data.

Consent for publication. All participants provided written informed consent for the publication of the study's findings. They were made aware that their data would be anonymized and used solely for research purposes.

Availability of data and materials. The design and research questions were preregistered on the AsPredicted website (https://aspredicted.org) and stored on the Open Science Framework (https://osf.io/vdwzx/) where all data and material can be accessed. No specific statistical analyses were preregistered.

Code availability. Not applicable

CRediT author statement. Conceptualization: EB, CB, JAH, SC, WDN. **Methodology:** EB, WDN. **Software**: EB. **Formal analysis:** EB**. Investigation:** EB, CB, JAH. **Resources:** SC, WDN. **Data curation:** EB. **Writing – Original Draft:** EB. **Writing – Review & Editing:** EB, CB, JAH, WDN, SC**. Visualization**: EB. **Supervision:** WDN, SC. **Funding acquisition:** EB, WDN.

References

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 & Reasoning*, *25*(3), 257–299. https://doi.org/10.1080/13546783.2018.1507949

Bago, B., & De Neys, W. (2020). Advancing the specification of dual process models of higher cognition: A critical test of the hybrid model view. *Thinking & Reasoning*, *26*(1), 1–30. https://doi.org/10.1080/13546783.2018.1552194

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., & De Neys, W. (2023a). Examining the role of deliberation in de-bias training. *Thinking & Reasoning*, *0*(0), 1–29. https://doi.org/10.1080/13546783.2023.2259542

Boissin, E., Caparos, S., & De Neys, W. (2023b). No easy fix for belief bias during syllogistic reasoning? *Journal of Cognitive Psychology*, *35*(4), 401–421.

https://doi.org/10.1080/20445911.2023.2181734

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. https://doi.org/10.1017/S1930297500008895

Boissin, E., Josserand, M., De Neys, W., & Caparos, S. (2024). Debiasing thinking among non-WEIRD reasoners. *Cognition*, *243*, 105681. https://doi.org/10.1016/j.cognition.2023.105681

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.

Claidière, N., Trouche, E., & Mercier, H. (2017). Argumentation and the diffusion of counter-intuitive beliefs. *Journal of Experimental Psychology: General*, *146*(7), 1052–1066. https://doi.org/10.1037/xge0000323

De Neys, W. (2006). Automatic–Heuristic and Executive–Analytic Processing during Reasoning: Chronometric and Dual-Task Considerations. *Quarterly Journal of Experimental Psychology*, *59*(6), 1070–1100. https://doi.org/10.1080/02724980543000123

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. (2022). Advancing theorizing about fast-and-slow thinking. *Behavioral and Brain Sciences*, 1–68. https://doi.org/10.1017/S0140525X2200142X

De Neys, W., Cromheeke, S., & Osman, M. (2011). Biased but in Doubt: Conflict and Decision Confidence. *PLOS ONE*, *6*(1), e15954. https://doi.org/10.1371/journal.pone.0015954

Evans, J. St. B. T. (2010). Intuition and Reasoning: A Dual-Process Perspective. *Psychological Inquiry*, *21*(4), 313–326. https://doi.org/10.1080/1047840X.2010.521057

Evans, J. St. 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

Franssens, S., & De Neys, W. (2009). The effortless nature of conflict detection during thinking. *Thinking & Reasoning*, *15*(2), 105–128. https://doi.org/10.1080/13546780802711185

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

Gigerenzer, G. (2009). Making sense of health statistics. Bulletin of the World Health Organization, 87(8), 567 . https://doi.org/10.2471/BLT.09.069872

Gigerenzer, G., Hell, W., & Blank, H. (1988). *Presentation and Content: The Use of Base Rates as a Continuous Variable*. 13.

Goel, V. (2022). *Reason and Less: Pursuing Food, Sex, and Politics*. MIT Press.

Hoover, J. D., & Healy, A. F. (2017). Algebraic reasoning and bat-and-ball problem variants: Solving isomorphic algebra first facilitates problem solving later. *Psychonomic Bulletin & Review*, *24*(6), 1922–1928. https://doi.org/10.3758/s13423-017-1241-8

Hoover, J.D., & Healy, A. F. (2021). The bat-and-ball problem: A word-problem debiasing approach. *Thinking & Reasoning*, *27*(4), 567–598. https://doi.org/10.1080/13546783.2021.1878473

Isler, O., & Yilmaz, O. (2022). How to activate intuitive and reflective thinking in behavior research? A comprehensive examination of experimental techniques. *Behavior Research Methods*, 1–20. https://doi.org/10.3758/s13428-022-01984-4

Isler, O., Yilmaz, O., & Dogruyol, B. (2020). Activating reflective thinking with decision justification and debiasing training. *Judgment and Decision Making*, *15*(6), 926–938.

Jeffreys, H. (1961). *Theory of probability, Clarendon*. Oxford.

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

Kaplan, J. T., Gimbel, S. I., & Harris, S. (2016). Neural correlates of maintaining one's political beliefs in the face of counterevidence. *Scientific Reports*, *6*(1), Article 1.

https://doi.org/10.1038/srep39589

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian Cognitive Modeling: A Practical Course*. Cambridge University Press. https://doi.org/10.1017/CBO9781139087759

Lin, H., Pennycook, G., & Rand, D. G. (2023). Thinking more or thinking differently? Using drift-diffusion modeling to illuminate why accuracy prompts decrease misinformation sharing. Cognition, 230, 105312. https://doi.org/10.1016/j.cognition.2022.105312

Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. *Journal of Open-Source Software*, 4(40), 1541. https://doi.org/10.21105/joss.01541

Mata, A. (2020). An easy fix for reasoning errors: Attention capturers improve reasoning performance. *Quarterly Journal of Experimental Psychology*, *73*(10), 1695–1702. https://doi.org/10.1177/1747021820931499

Mata, A., Ferreira, M., & Sherman, S. (2013). The metacognitive advantage of deliberative thinkers: A dual-process perspective on overconfidence. *Journal of Personality and Social Psychology*, *105*(3), 353–373. https://doi.org/10.1037/a0033640

Mata, A., Schubert, A.-L., & B. Ferreira, M. (2014). The role of language comprehension in reasoning: How "good-enough" representations induce biases. *Cognition*, *133*(2), 457–463. https://doi.org/10.1016/j.cognition.2014.07.011

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, *130*(4), 621–640. https://doi.org/10.1037/0096-3445.130.4.621

Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C. W., Korris, J. H., & Kassam, K. S. (2015). Debiasing Decisions: Improved Decision Making With a Single Training Intervention. *Policy Insights from the Behavioral and Brain Sciences*, *2*(1), 129–140. https://doi.org/10.1177/2372732215600886

Morey, R. D., Rouder, J. N., & Jamil, T. (2015). BayesFactor: Computation of Bayes factors for common designs. *R Package Version 0.9*, 9, 2014.

Moutier, S., Plagne-Cayeux, S., Melot, A.-M., & Houdé, O. (2006). Syllogistic reasoning and belief-bias inhibition in school children: Evidence from a negative priming paradigm.

Developmental Science, *9*(2), 166–172. https://doi.org/10.1111/j.1467-7687.2006.00476.x

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. Nature, 592(7855),. https://doi.org/10.1038/s41586-021-03344-2

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

Pennycook, G., Trippas, D., Handley, S. J., & Thompson, V. A. (2014). Base rates: Both neglected and intuitive. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 544–554. https://doi.org/10.1037/a0034887

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

R Core Team. (2023). *R: A Language and Environment for Statistical Computing* (4.3.2). R Foundation for Statistical Computing. https://www.R-project.org/

Sloman, S. A. (1996). The Empirical Case for Two Systems of Reasoning. *Psychological Bulletin*, *119*(1), 3. https://doi.org/10.1037/0033-2909.119.1.3

Stanovich, K. E., & West, R. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*, 645–726.

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

Trouche, E., Sander, E., & Mercier, H. (2014). Arguments, more than confidence, explain the good performance of reasoning groups. *Journal of Experimental Psychology: General*, *143*(5), 1958–1971. https://doi.org/10.1037/a0037099

Voudouri, A., Białek, M., Domurat, A., Kowal, M., & De Neys, W. (2023). Conflict detection predicts the temporal stability of intuitive and deliberate reasoning. *Thinking & Reasoning*, *29*(4), 427–455. https://doi.org/10.1080/13546783.2022.2077439

Supplementary material

A. Example of Boissin et al. (2022)'s explicit intervention.

Participants had to respond first to this item: "This study contains businessmen and firemen. Person 'K' is brave. There are 996 businessmen and 4 firemen."
After providing a response, they were introduced to this short explanation:

"The correct answer to the previous problem is that person 'K' is most likely a "businessman". Many people think it is "fireman", but this answer is wrong. Most people base their answer solely on the description ("Person K is brave"). If this were all information you got, this answer would be correct, as it is likely that there are more brave firemen in the world than brave businessmen. However, in the problem you also got information about the specific number of businessmen and firemen in the group that person K got drawn from. You were informed that person K was drawn randomly from a group with 996 businessmen and only 4 firemen. Since there are so much more businessmen in the group than firemen (200 times more!), it becomes more likely that person K is a businessman. After all, although firemen might in general be braver than businessmen, there are also some businessmen who are brave. If you combine this with the vastly larger number of businessmen in the group, it will be more plausible that you're dealing with a brave businessman."

B. Material

Conflict items	**No-conflict items**
This study contains high school students and librarians. Person 'M' is loud. There are 5 high school students and 995 librarians.	This study contains high school students and librarians. Person 'M' is loud. There are 995 high school students and 5 librarians.
This study contains politicians and librarians. Person 'P' is dishonest. There are 4 politicians and 996 librarians.	This study contains politicians and librarians. Person 'P' is dishonest. There are 996 politicians and 4 librarians.
This study contains women and drummers. Person 'I' is careful. There are 997 women and 3 drummers.	This study contains women and drummers. Person 'I' is careful. There are 3 women and 997 drummers.
This study contains flight attendants and surgeons. Person 'E' is kind. There are 5 flight attendants and 995 surgeons.	This study contains flight attendants and surgeons. Person 'E' is kind. There are 995 flight attendants and 5 surgeons.
This study contains rich people and gardeners. Person 'G' is arrogant. There are 4 rich people and 996 gardeners.	This study contains rich people and gardeners. Person 'G' is arrogant. There are 996 rich people and 4 gardeners.

This study contains lawyers and gardeners. Person 'W' is argumentative. There are 3 lawyers and 997 gardeners.	This study contains lawyers and gardeners. Person 'W' is argumentative. There are 997 lawyers and 3 gardeners.
This study contains high school coaches and dentists. Person 'O' is loud. There are 3 high school coaches and 997 dentists.	This study contains high school coaches and dentists. Person 'O' is loud. There are 997 high school coaches and 3 dentists.
This study contains secretaries and telemarketers. Person 'J' is persuasive. There are 995 secretaries and 5 telemarketers.	This study contains secretaries and telemarketers. Person 'J' is persuasive. There are 5 secretaries and 995 telemarketers.
This study contains clowns and accountants. Person 'L' is funny. There are 5 clowns and 995 accountants.	This study contains clowns and accountants. Person 'L' is funny. There are 995 clowns and 5 accountants.
This study contains scientists and assistants. Person 'C' is intelligent. There are 4 scientists and 996 assistants.	This study contains scientists and assistants. Person 'C' is intelligent. There are 996 scientists and 4 assistants.
This study contains lab technicians and aerobics instructors. Person 'D' is active. There are 996 lab technicians and 4 aerobics instructors.	This study contains lab technicians and aerobics instructors. Person 'D' is active. There are 4 lab technicians and 996 aerobics instructors.
This study contains real estate agents and poor people. Person 'K' is persuasive. There are 5 real estate agents and 995 poor people.	This study contains real estate agents and poor people. Person 'K' is persuasive. There are 995 real estate agents and 5 poor people.
This study contains I.T. technicians and boxers. Person 'F' is strong. There are 995 I.T. technicians and 5 boxers.	This study contains I.T. technicians and boxers. Person 'F' is strong. There are 5 I.T. technicians and 995 boxers.
This study contains accountants and boys. Person 'H' is immature. There are 997 accountants and 3 boys.	This study contains accountants and boys. Person 'H' is immature. There are 3 accountants and 997 boys.
This study contains businessmen and firemen. Person 'K' is brave. There are 996 businessmen and 4 firemen.	This study contains businessmen and firemen. Person 'K' is brave. There are 4 businessmen and 996 firemen.
This study contains nurses and artists. Person 'S' is creative. There are 997 nurses and 3 artists.	This study contains nurses and artists. Person 'S' is creative. There are 3 nurses and 997 artists.

Heuristic intervention items	**Logical intervention items**
This study contains nannies and businessmen. Person 'Q' is ambitious.	This study contains nannies and businessmen. Person 'W' has two ears.

There are 500 nannies and 500 businessmen.	There are 997 nannies and 3 businessmen.
This study contains writers and constructions workers. Person 'W' is strong. There are 500 writers and 500 construction workers.	This study contains writers and constructions workers. Person 'Z' has a mouth. There are 997 writers and 3 constructions workers.
This study contains real estate agent and poor people. Person 'F' is persuasive. There are 500 real estate agents and 500 poor people.	This study contains poor people and real estate agents. Person 'L' has a forehead. There are 995 poor people and 5 real estate agents.
This study contains consultants and construction workers. Person 'A' is helpful. There are 500 consultants and 500 construction workers.	This study contains consultants and construction workers. Person 'C' has two eyes. There are 4 consultants and 996 construction workers.
This study contains executive managers and sixteen-year-olds. Person 'J' is immature. There are 500 executive managers and 500 sixteen-year-olds.	This study contains executive managers and sixteen-year-olds. Person 'D' has ten toes. There are 995 executive managers and 5 sixteen-year-olds.
This study contains kindergarten teachers and boxers. Person 'E' is kind. There are 500 kindergarten teachers and 500 boxers.	This study contains kindergarten teachers and boxers. Person 'F' has two hands. There are 3 kindergarten teachers and 997 boxers.
This study contains nurses and artists. Person 'S' is creative. There are 500 nurses and 500 artists.	This study contains nurses and artists. Person 'S' has ten fingers. There are 997 nurses and 3 artists.
This study contains businessmen and firemen. Person 'K' is brave. There are 500 businessmen and 500 firemen.	This study contains businessmen and firemen. Person 'K' has two knees. There are 4 firemen and 996 businessmen.
This study contains doctors and farmers. Person 'A' is intelligent. There are 500 doctors and 500 farmers.	This study contains doctors and farmers. Person 'A' has two feet. There are 5 doctors and 995 farmers.
This study contains computer programmers and hippies. Person 'Z' is nerdy. There are 500 computer programmers and 500 hippies.	This study contains computer programmers and hippies. Person 'M' has a nose. There are 5 computer programmers and 995 hippies.

C. Bayesian factor when not controlling for number of conflict trials in the post-intervention block.

a. Conflict accuracy

Table S1. Bayes Factor for the contrast in accuracy before and after the intervention for both intervention groups and response stages.

Group	**Initial response** *Pre- vs. Post-intervention*	**Final response** *Pre- vs. Post-intervention*
Logical intervention	BF10 = 50.27	BF10 > 100
Heuristic intervention	BF01 = 3.20	BF01 = 1.98

b. Direction of change

Table S2. Bayes Factor for the contrast in direction of change before and after the intervention for both intervention groups and response stages.

Group	**"00" patterns** *Pre- vs. Post-intervention*	**"01" patterns** *Pre- vs. Post-intervention*	**"11" patterns** *Pre- vs. Post-intervention*
Logical intervention	BF10 > 100	BF10 = 1.28	BF10 > 100
Heuristic intervention	BF10 = 26.77	BF01 = 6.94	BF10 = 2.31

D. Inferential analyses

a. With controlling for number of conflict trials in the post-intervention block.

i. Conflict accuracy

We used Generalized Linear Mixed Models for contrasting conflict performance before and after the intervention for both groups (Logical and Heuristic intervention groups) and both response stages (Initial and Final response stage). In these models, Block (Pre vs. Post-intervention) was included as a predictive factor as well as the Number of Conflict Trial factor (2 vs. 4) for controlling the number of conflict trials during the post-intervention block.

Participants were entered as random intercepts.

Table S3. Generalized Linear Mixed Model outputs for both initial and final response accuracies before and after the intervention for both intervention groups. The model includes number of conflict trial during post-intervention for controlling the manipulation error.

	Logical intervention		**Heuristic intervention**	
	Initial response	**Final response**	**Initial response**	**Final response**
Intercept	-3.76	-0.91	-4.14 *	-2.01
	(2.04)	(1.59)	(1.63)	(2.23)
Number of conflict trial during post-intervention	0.53	0.37	0.87	0.75
	(0.60)	(0.50)	(0.49)	(0.71)
Block	1.45 **	1.46 ***	0.48	0.69
	(0.45)	(0.41)	(0.39)	(0.44)

$***p < 0.001$; $**p < 0.01$; $*p < 0.05$

ii. Direction of change

We used Linear Mixed Models for contrasting proportion of each direction of change (00, 01 and 11) before and after the intervention for both groups (Logical and Heuristic intervention groups). In these models, Block (Pre vs. Post-intervention) was included as a predictive factor as well as the Number of Conflict Trial factor (2 vs. 4) for controlling the number of conflict trials during the post-intervention block. Participants were entered as random intercepts.

Table S4. Generalized Linear Mixed Model outputs for each direction of change before and after the intervention for both intervention groups. The model includes number of conflict trial during post-intervention for controlling the manipulation error. We report the estimate (and the standard error).

	Logical intervention			**Heuristic intervention**		
	"00" patterns	**"01" patterns**	**"11" patterns**	**"00" patterns**	**"01" patterns**	**"11" patterns**
Intercept	48.91***	21.72***	26.58**	47.63***	26.18***	23.02**
	(7.65)	(5.92)	(7.77)	(8.46)	(6.04)	(7.21)
Number of conflict trial during post-intervention	-5.06	2.82	2.71	-9.91	-5.64	14.63
	(10.92)	(8.34)	(11.14)	(12.27)	(8.68)	(10.14)
Block	-18.57***	5.89*	10.40***	-6.40 ***	1.02	4.94*
	(2.26)	(2.68)	(1.88)	(1.91)	(2.06)	(2.04)

$***p < 0.001$; $**p < 0.01$; $*p < 0.05$

b. Without controlling for number of conflict trials in the post-intervention block.

i. Conflict accuracy

We used Generalized Linear Mixed Models for contrasting conflict performance before and after the intervention for both groups (Logical and Heuristic intervention groups) and both response stages (Initial and Final response stage). In these models, Block (Pre vs. Post-intervention) was included as a predictive factor. Participants were entered as random intercepts.

Table S5. Generalized Linear Mixed Model outputs for both initial and final response accuracies before and after the intervention for both intervention groups. The model does not include number of conflict trial during post-intervention for controlling the manipulation error. We report the estimate (and the standard error).

	Logical intervention		**Heuristic intervention**	
	Initial response	**Final response**	**Initial response**	**Final response**
Intercept	-2.12**	0.22	-1.53**	0.24***
	(0.67)	(0.55)	(0.54)	(0.00)
Block	1.47**	1.49***	0.52	0.71***
	(0.45)	(0.41)	(0.39)	(0.00)

∗∗∗p < 0.001; ∗∗p < 0.01; ∗p < 0.05

c. Direction of change

We used Linear Mixed Models for contrasting proportion of each direction of change (00, 01 and 11) before and after the intervention for both groups (Logical and Heuristic intervention groups). In these models, Block (Pre vs. Post-intervention) was included as a predictive factor Participants were entered as random intercepts.

Table S6. Generalized Linear Mixed Model outputs for each direction of change before and after the intervention for both intervention groups. The model does not include number of conflict trial during post-intervention for controlling the manipulation error. We report the estimate (and the standard error).

	Logical intervention			**Heuristic intervention**		
	"00" patterns	**"01" patterns**	**"11" patterns**	**"00" patterns**	**"01" patterns**	**"11" patterns**
Intercept	46.47***	23.08***	27.89**	42.97***	23.52***	29.90***
	(5.52)	(4.32)	(5.50)	(6.17)	(4.41)	(5.34)
Block	-18.60***	5.94*	10.42***	-6.44***	0.97	5.04*
	(2.26)	(2.68)	(1.88)	(1.90)	(2.06)	(2.03)

$***p < 0.001$; $**p < 0.01$; $*p < 0.05$

E. No-conflict item accuracies

Table S7. Mean no-conflict initial and final accuracies (and Standard Error of the Mean, SEM) before and after the intervention for both intervention groups.

Group	**Initial response**		**Final response**	
	Pre-intervention	**Post-intervention**	**Pre-intervention**	**Post-intervention**
Logical Intervention	97.3 (1.5)	93.8 (2.0)	99.0 (1.0)	96.5 (1.5)
Heuristic intervention	100.0 (0.0)	96.4 (1.7)	97.8 (1.4)	99.5 (0.5)