Thinking & Reasoning

ISSN: 1354-6783 (Print) 1464-0708 (Online) Journal homepage: www.tandfonline.com/journals/ptar20

Reasoning does hurt: deliberation is associated with heightened levels of doubt

Cédric Cortial, Jérôme Prado & Serge Caparos

To cite this article: Cédric Cortial, Jérôme Prado & Serge Caparos (2026) Reasoning does hurt: deliberation is associated with heightened levels of doubt, Thinking & Reasoning, 32:1, 92-124, DOI: 10.1080/13546783.2025.2531067

To link to this article: https://doi.org/10.1080/13546783.2025.2531067

THINKING & REASONING
2026, VOL. 32, NO. 1, 92–124
https://doi.org/10.1080/13546783.2025.2531067

Reasoning does hurt: deliberation is associated with heightened levels of doubt

Cédric Cortial[a,b], Jérôme Prado[c] and Serge Caparos[a,d]

[a]Université Paris 8, DysCo lab, Saint-Denis, France; [b]Université Clermont-Auvergne, LAPSCO, Clermont-Ferrand, France; [c]Centre de Recherche en Neurosciences de Lyon (CRNL), INSERM U1028 - CNRS UMR5292, Université de Lyon, Lyon, France; [d]Institut Universitaire de France, Paris, France

ABSTRACT
Human cognition involves two distinct processes: intuition and deliberation. While deliberation requires significant cognitive effort and is often resource-intensive, intuition occurs effortlessly. A key challenge for reasoning theories is to identify the triggers for deliberation. Some compelling theories suggest that deliberative processes are activated by a metacognitive experience of uncertainty; however, they struggle to incorporate findings that connect emotion with cognition. To assess the role of emotions in the transition from intuition to deliberation, we utilised problems designed to elicit such metacognitive experiences and measured emotional responses using a doubt scale, where doubt is defined as a negative valence emotional response resulting from discontinuous or incoherent situations. Study 1 aimed to examine the relationship between cognitive conflict and the emotional response of doubt. Study 2 explored the connection between doubt and deliberation through a two-response paradigm, where participants provided an initial response under time pressure and then gave a final response after deliberation. Our findings demonstrated that cognitive conflict is appraised as negative and triggers an emotional response of doubt, and that doubt is associated with increased engagement in deliberation.

ARTICLE HISTORY Received December 3, 2024; Accepted July 3, 2025

KEYWORDS Reasoning; emotion; doubt; cognitive conflict; intuition; appraisal

Introduction

A prominent theoretical framework in psychology suggests that human reasoning operates through two distinct systems: intuition and deliberation (Epstein, 1994; Evans, 2006; James, 1890; Kahneman, 2003; Sloman, 1996;

CONTACT Cédric Cortial cedric.cortialpro@gmail.com Université Clermont-Auvergne, 17 Rue Paul Collomp, 63000 Clermont-Ferrand, France.
Supplemental data for this article can be accessed online at https://doi.org/10.1080/13546783.2025.2531067.

Stanovich & West, 2000). Intuition represents a rapid and instinctive form of reasoning, whereas deliberation embodies a deep and analytical mode, which requires substantial cognitive resources. While deliberation has traditionally been posited as a mechanism used to produce responses that are consistent with logico-mathematical principles, through the rectification of illogical heuristic-based intuitions, recent observations challenge this notion and suggest that intuitions can be logically sound (Bago & De Neys, 2019; Raoelison et al., 2020; Thompson et al., 2018). Consequently, the prevailing model that often portrays intuitions as flawed, biased, or irrational may need reconsideration. Intuitions may serve as a proficient problem-solving tool (Gigerenzer & Todd, 1999) and, more importantly, they may follow logico-mathematical rules and be qualified as 'fully rational'.

Despite this recent re-evaluation of the role of intuitions in human thinking (De Neys, 2022), it is clear that intuitive reasoning has limitations. For example, it may not be systematic (i.e., intuitions may not always come to mind), it may not be fast enough (and therefore seem uncertain), or it may not be unique (such that several intuitions associated with different responses compete with each other). In all these cases, intuitive reasoning appears to have failed, necessitating the intervention of deliberation (Ackerman & Thompson, 2017; De Neys & Pennycook, 2019; De Neys, 2022).

However, in order to explain why people would use deliberation at all, models must explain why individuals frequently seek to avoid it (David et al., 2024; Stanovich, 2011). Although it may appear reasonable to conclude that a failure in problem-solving should prompt a shift in strategy, from intuition to deliberation, one may nevertheless wonder why a failure of intuition should bother us at all. We might be more efficient by simply taking the first intuition that comes to mind or choosing an intuition at random when several compete. Knowing that the logical validity of intuitions is not a reliable criterion for predicting which one an individual will choose (Frederick, 2005; Kahneman et al., 1982; Wason, 1968; Wason & Evans, 1974), we would save precious cognitive resources and react much more quickly to our environment. Arguably, a distal explanation for why we may call upon deliberation even if it is costly lies in evolution. Not caring at all about our mistakes would make us very poor candidates for survival. However, the proximal explanation, namely, how this evolutionary advantage is instantiated in the organism, remains unexplained. While several models attempt to describe the mechanism underlying the switch from intuition to deliberation (De Neys, 2022; De Neys & Glumicic, 2008; Pennycook et al., 2015; Thompson et al., 2011), these models often fail to fully address the emotional component of the switch, an element that we believe is a fruitful and relevant way of addressing the problem of motivation to use an unpleasant and aversive process such as deliberation (David et al., 2024).

In fact, in order to explain why the organism would make use of deliberation despite its striving to avoid it as much as possible, we recently proposed that the trigger for deliberative reasoning originates in an emotional and embodied state of negative valence, reminiscent of the feelings of anxiety, surprise, or confusion (Cortial et al., 2025). We called this emotional manifestation "doubt", following the footsteps of pragmatist philosophers John Dewey and Charles S. Peirce (Dewey, 1910; Peirce, 1877). Doubt is a negative emotion caused by something appraised as a disruption of an individual's automatic interaction with the environment. In the present paper, we use doubt as an umbrella term that encompasses a subset of related emotions which all share the same negative valence, originate from a discrepant or disturbed situation, and have been shown to promote deliberation (i.e., surprise has been shown to drive people to use more costly cognitive processes: Meyer et al., 1991; Muis et al., 2015; Pezzo, 2003; Topolinski & Strack, 2015; confusion promotes learning: Craig et al., 2004; D'Mello & Graesser, 2014; VanLehn et al., 2003; Vogl et al., 2019, 2020; and anxiety motivates deeper thinking: Vazard, 2021, 2022).

Our proposal is generally rooted in the Appraisal Theory of Emotion (Moors et al., 2013; Scherer & Moors, 2019). According to this theoretical framework, emotions arise from evaluating whether information or stimuli align with an individual's goals. The more a new element contradicts one's goals, the stronger and more negative the emotional response would be, and vice versa (Frijda, 1986; Lazarus et al., 1970; Moors & Fischer, 2019; Parkinson & Manstead, 1992; Reisenzein, 2001; Roseman, 1984; Smith, 1989; Smith & Lazarus, 1990). When intuition fails, it disrupts the fluidity of our interaction with the environment and this failure would trigger an emotional response, motivating us to take action in order to restore coherence through deliberation. This view differs from other recent models that are rooted in a metacognitive framework, whereby poor intuitions would generate a low Feeling of Rightness (FOR; Thompson et al., 2011), a strong Feeling of Error (FOE; Fernandez Cruz et al., 2016; Gangemi et al., 2015), a strong Feeling of Wrongness (FOW; Ackerman & Thompson, 2017), or a strong Response Uncertainty (RU; De Neys, 2022). Indeed, those models typically do not conceptualise the trigger of deliberation as being emotional in nature[1].

A key question, in our view, concerns the role of emotion in metacognitive experiences, such as the Feeling of Rightness (FOR). When situated within a metacognitive framework, the FOR is not itself an appraisal in the sense defined by the Appraisal Theory of Emotion, and cannot be

[1]While the FOR model does consider an affective component (Thompson et al., 2011; Thompson & Morsanyi, 2012), it does not go beyond the idea that strong intuitions are associated with pleasantness (Ackerman & Thompson, 2017; Morsanyi & Handley, 2012; Thompson et al., 2011; Trippas et al., 2016) and it does not integrate the observation that physiological arousal is observed in situations of cognitive conflict during reasoning tasks (De Neys et al., 2010; Purcell et al., 2023).

equated to an emotional response. According to the Appraisal theory, when an organism appraises a situation—for example, as fluent or disfluent—this appraisal can elicit an emotional reaction. Metacognitive responses such as the FOR would emerge at a later stage, serving to interpret or regulate the motivational signal. In the case of reasoning, the negative affect triggered by weak intuitions may act as a cue to reconsider the situation and engage further cognitive resources.

In sum, our proposal is that a more comprehensive model of reasoning should position FOR or RU as a cognitive evaluation of intuitive failure, one that draws upon the emotional cue of doubt in order to initiate deliberative thinking. The present study aims to test the relevance of the Appraisal framework to the field of reasoning. It may allow us to explain why individuals engage in costly deliberative reasoning when intuitive reasoning has failed. In order to emulate situations where intuitive reasoning is suboptimal, it is possible to use reasoning problems that can elicit conflicting intuitions (De Neys et al., 2010; Sá et al., 1999). Consider, for example, the following problem: "All primates have legs. Humans have legs. Therefore, humans are primates". Although at first glance the argument may seem logically sound, the conclusion "humans are primates" does not follow logically from the premises (i.e., primates are not the only species with legs, birds for example also have legs). It is a fallacious form of logical reasoning known as affirmation of the consequent. Yet, reasoners may be inclined to endorse the conclusion because it is believable. In other words, there is conflict between the believability of the conclusion and the logical validity of the argument (Evans et al., 1983). Interestingly, reasoners who endorse the conclusion tend to report less confidence (or more uncertainty) in their answer in such problems than in problems where believability and logicality do not conflict. This suggests that the conflict has been detected (Bago & De Neys, 2019; Botvinick, 2007; De Neys, 2012; 2022; De Neys et al., 2008, 2010; De Neys & Glumicic, 2008; De Neys & Pennycook, 2019; Pennycook et al., 2015). In this study, we test the hypothesis that the evaluation of the product of intuitive reasoning is not solely cognitive in nature and also relies on an emotional response. The negative nature of such response would explain our motivation to trigger deliberation, in order to eliminate the negative emotional state by resolving the failure of intuitive reasoning. In sum, the two hypotheses tested in the present work are that (1) intuition failures provoke not only a second-order cognitive judgement about confidence in the response, but also a broader appraisal leading to an emotional response, named doubt, and that (2) the presence of such negative emotion prompts reasoners to engage in deliberative reasoning.

In Experiment 1, we asked participants to evaluate the logical validity of syllogisms in which conclusions could be consistent or inconsistent with prior beliefs. Participants were required to respond under time pressure in order to promote the use of intuitive reasoning. We introduced a subjective measure of doubt and explored its association with participants' performance, as well as the association between doubt and state anxiety. In Experiment 2, we assessed whether doubt triggers deliberation by employing a two-response paradigm (Thompson et al., 2011). This allowed us to discriminate intuitive from deliberative responses. Participants first responded to a reasoning problem under time constraint (first "fast" response) to limit the involvement of deliberation. They then responded once again to the same problem with all the time necessary (second "slow" response), allowing them to use deliberation. We used three indices of deliberation: the number of times participants changed their response between the first and second presentation of the problem (Evans & Stanovich, 2013; Kahneman, 2011), the response time associated with the second response (longer RTs indicating a more likely involvement of deliberation; Thompson et al., 2011), and participants' self-report of whether or not they had engaged in deliberation before giving their second response. Our hypotheses were that 1) conflict problems would generate more doubt than non-conflict problems, and that 2) there would be a positive relationship between doubt and deliberation.

Experiment 1

Method

Pre-registration

The study design, hypotheses, research question and analyses were pre-registered on the Open Science Framework (https://osf.io/v3pq5)

Participants

Participants were students and employees from the University Clermont-Auvergne recruited by e-mail, using a mailing list covering the entire pool of students and employees of the university. A total of 620 French-speaking participants took part in the experiment. Education level was collected for all participants. A total of 46 people were excluded from the analyses because they failed the attention test that was embedded in the experiment (see below). Therefore, a sample of 574 participants were included in the analyses (390 women, mean age: $M = 20.8$ years, $SD = 5.09$). Most of them reported having completed secondary school ($n = 448$), while others reported having completed undergraduate education ($n = 99$), post-graduate education ($n = 25$), or only primary education ($n = 2$).

Given that we had never previously tested our measure of doubt before, we anticipated that the difference in doubt between conflict and no-conflict problems might be small (i.e., between $d = 0.10$ and $d = 0.20$). Power calculations performed with G*Power (version 3.1.9.7) indicated that a minimum of 620 participants would be needed to detect an effect size of $d = 0.10$ with a power of 80% in a paired t-test ($p = .05$). Therefore, we stopped collecting data when we reached that sample size.

Materials

Syllogisms. Participants were asked to solve syllogisms that were built using categorical versions of two possible conditional inferences. The first was equivalent to the valid *Modus Ponens* (If P then Q, P is true, therefore Q is true), and the second was equivalent to the invalid *Affirmation of the Consequent* (If P then Q, Q is true, therefore P is true).

Each problem comprised a major premise (e.g., “All vehicles have wheels”), a minor premise (e.g., “Boats are vehicles”) and a conclusion (e.g., “Therefore, Boats have wheels”). Participants were asked to assess whether the conclusion to each problem followed logically from the two premises (that they had to consider true regardless of their degree of belief in them). The following format was used for (1) valid problems of which the conclusion followed logically and (2) invalid problems of which the conclusion did not follow logically:

1. All vehicles have wheels.
   Boats are vehicles.
   Therefore, boats have wheels.
   Does the final statement follow logically from the first two?
2. All felines can walk.
   Cats can walk.
   Therefore, cats are felines.
   Does the final statement follow logically from the first two?

Twelve syllogisms were presented. They were taken from the work of Boissin et al. (2023) and translated into French (see full list of problems in supplemental materials). Half of the problems involved conflicts, where logical validity and believability did not align, while the other half were non-conflict problems, where logical validity and believability aligned (see Table 1).

Doubt scale. After responding to each syllogism, the participants reported whether they had experienced doubt when solving the problem, by rating the following five statements: “I felt confused”, “I felt blocked”, “I felt the problem was pushing me to think harder”,

Table 1. Examples of problems.

	Believable	Unbelievable
Logical	All flowers need water. Roses are flowers. Therefore, roses need water.	**All vehicles have wheels.** **Boats are vehicles.** **Therefore, boats have wheels.**
Illogical	**All mammals can walk.** **Cows can walk.** **Therefore, cows are mammals.**	All birds can fly. Planes can fly. Therefore, planes are birds.

Note: Conflict problems in bold.

"I felt doubt", "I found the reflection unpleasant", using a four-option Likert scale (i.e., "not at all", "not really", "quite so" and "completely"). Statements were chosen on the basis of the theoretical dimensions of the emotion of doubt. Accordingly, we aimed to assess an emotional appraisal that evaluates whether the experienced situation aligns with the individual's goals. To achieve this, we asked participants about feeling blocked (i.e., "I can't do what I want to do") and the need for further thought (i.e., "I must change the situation"). Additionally, we incorporated an assessment of the perceived valence of the situation (i.e., "unpleasantness") and of the subjective experience of the emotion, namely, "doubt" and "confusion". The doubt scale showed good internal consistency (Cronbach α = .88). For each participant and each problem, we calculated a doubt score by averaging the scores to the five items and transforming it into a percentage (i.e., min = 0, max = 100). We then computed the mean doubt score for conflict problems and the mean score for non-conflict problems.

Questionnaires

At the end of the experiment, the participants completed an anxiety questionnaire, using a reduced 5-item version of the State-Trait Anxiety Inventory (STAI) measuring state anxiety (Zsido et al., 2020; Cronbach α = .91). Participants rated their agreement with five statements about how they felt ("upset", "scared", "nervous", "jittery", and "confused") again using a 4-option Likert scale for each item ("strongly disagree", "somewhat disagree", "somewhat agree", and "strongly agree"). For each participant, we computed an average state-anxiety score and transformed it into a percentage (Cronbach α = .71).

Participants also completed a reduced version of the Need For Cognition scale (NFC; Chiesi et al., 2018) which assesses participants' personal tendency to engage in and enjoy effortful cognitive activities, using a 10-item Likert scale with 5 response choices for each item ("Not at all", "A little", "Moderately", "Quite so", and "Completly"). Examples of items include: "I prefer complex to simple problems", "I like to have the responsibility of handling a situation that requires a lot of thinking" or "Thinking is not my idea of fun". Although Chiesi et al. (2018) reported a McDonald ω ranging from .76 to .90, suggesting a very good consistency between items, in our study this scale appeared to show low consistency (Cronbach α = .46), which is why we decided not to report the findings linked to this scale here (analyses of the findings in relation with this scale can be found in the Experiment 1 analyses of the supplemental material.)

Procedure

The experiment was run online using the OpenSesame software (Mathôt et al., 2012) supported online by Jatos and hosted on the MindProbe server created by the European Society for Cognitive Psychology. Participants were informed that the study would last around 10 min. They were instructed to isolate themselves in a quiet, distraction-free environment. The instructions explained that the participants would have to assess whether reasoning problems were logically valid or not, under time pressure and using their intuition (see full description of instructions in supplemental material).

After having read the set of instructions, the participants were presented with a practice problem. They were then presented with the 12 test problems. The conflict and non-conflict problems were presented in a pseudo-random order (see supplemental material). For each trial, the full problem content (consisting in two premises and one conclusion) and the two response choices ("logically valid" and "logically invalid") were presented all at once. Participants had 10 s to respond after problem onset. Note that problem presentation time was selected following an informal pre-test, showing that 10 s was enough for participants to be able to read and respond to the problem. After 10 s had elapsed, the problem disappeared and was replaced by the 5-item doubt scale, whether an answer had been given or not. The participants could take all the time they needed to complete the doubt scale, after which another problem was presented. After the completion of 12 trials (each trial consisting of a problem followed by the doubt scale), an unexpected attention test was presented. Participants were asked to tick the 7th box in a 10-box panel. Participants were excluded from further analysis if they failed the attention test. Finally, after the attention test, the participants filled the NFC and State-Anxiety scales and answered a short socio-demographic questionnaire.

Ethics

The study was approved by the Ethics Committee of the University of Clermont-Auvergne (IRB00011540-2023-17).

Data analysis

The data were processed using Python and the Pandas library (code available on https://osf.io/v3pq5). For each participant, we computed the total number of logical responses (min = 0, max = 6) to both non-conflict and conflict problems. Each missing response was considered as a wrong answer. Given the low number of missing data (6.96%), removing trials in which there was a missing response from the calculation of error rates did not affect the results and did not change the conclusions of the study.

The distributions of all relevant scores (Number of logical responses; doubt score; State-Anxiety score) did not follow normal distributions. Therefore, we used non-parametric statistical tests wherever possible.

Results

In this section we begin by describing the results of the pre-registered analyses before moving to exploratory analyses.

Pre-registered analyses

As expected, participants were overall very accurate when responding to non-conflict problems, with an average of 5.11 logical responses out of 6 ($SD = 1.06$; see Table 2). This confirms that participants understood the task well. In contrast, participants were much less accurate on conflict problems, with an average of 1.38 logical responses out of 6 ($SD = 1.45$). A Wilcoxon signed rank test comparing the number of logical responses in conflict and non-conflict problems showed a significant difference between the two conditions, $W = 86$, $p < .001$, $r = 0.99$. The effect size r in the Wilcoxon test measures the magnitude of the difference between two groups. It is used for nonparametric tests and is interpreted in the same way as the Pearson coefficient that is used for parametric tests (where $r = .10$ is considered small, $r = .30$ is considered medium and $r > .50$ is considered large). One important hypothesis of this study was that participants would experience more doubt overall with conflict than with non-conflict problems. This was the case, as the mean level of doubt was significantly higher for conflict problems ($M_{conflict} = 21.41$, $SD_{conflict} = 16.44$) than for non-conflict problems ($M_{non\text{-}conflict} = 15.44$, $SD_{non\text{-}conflict} = 14.67$), $W = 26795$, $p < .001$, $r = .63$.

Additionally, a Spearman correlation test showed that the average level of doubt for conflict problem was positively related to the number of logical responses for these problems ($\rho = .13$, $p = .002$; see Table 3). To further examine this relationship, we conducted a series of generalised linear models (GLMs), the results of which are reported in the Experiment 1 analyses of the supplemental material. Finally, the mean level of doubt promoted by conflict problems was related to the level of state anxiety reported at the end of the experiment, $\rho = .31$, $p < .001$.

Exploratory analyses

First, we focused solely on belief-driven responses (i.e., responses that were correct in non-conflict trials and incorrect in conflict trials). We tested whether people doubted more in the conflict than the non-conflict condition, despite having given the expected intuitive answer in both

Table 2. Means of doubt and Number of logical responses across both conflict conditions.

		Mean *(SD)*
Number of logical responses	Conflict	1.38 *(1.45)*
	Non-Conflict	5.11 *(1.06)*
Doubt (%)	Conflict	21.41 *(16.44)*
	Non-Conflict	15.44 *(14.67)*

Table 3. Spearman's correlation table between doubt, logical responses and state-anxiety scores in Experiment 1.

		Doubt		State-anxiety
		Non-conflict	Conflict	
State-anxiety		.35***	.31***	–
Number of logical responses	Non-conflict	−0.29**	–	−0.11*
	Conflict	–	.13*	−0.07

Note.
*p <.05. **p < .01. ***p < .001.

Table 4. Number of participants in each accuracy groups.

Accuracy groups (number of logical responses in the conflict condition)	0	1	2	3	4	5	6
N (%)	191 (33%)	174 (30%)	103 (18%)	57 (10%)	17 (3%)	21 (4%)	11 (2%)

conditions. After excluding some participants from the analysis because they had no belief-driven response in one or both conditions, the analysis was conducted on a subset of $N = 539$ participants. As expected, doubt remained higher in conflict problems ($M = 21.2$, $SD = 20$) than in non-conflict problems ($M = 10.2$, $SD = 13.7$) in this subset of participants. A Wilcoxon rank-signed test showed that the difference was significant, $W = 87336$, $p < .001$, $r = .74$.

Although we did predict that there would be a relationship between the level of doubt and the number of logical responses, we did not make any specific prediction regarding the shape of this relationship. Here, we examined whether the observed relationship was linear or non-linear. We separated the participants into 7 groups (see Table 4), according to their number of logical responses in the conflict condition (min = 0, max = 6). A Kruskal-Wallis's ANOVA using group as an independent variable showed that the average level of doubt in the conflict condition varied according to the participants' number of logical responses, $\chi^2(6) = 14.8$, $p = .021$, $\varepsilon^2 = .026$. A contrast analysis further showed that the latter relationship was best explained by a quadratic function, $t(567) = -2.32$, $p = .021$ (see Figure 1).

Discussion

The goal of Experiment 1 was to test the hypothesis that participants experience a negative emotion, doubt, when a reasoning problem generates a cognitive conflict, for example when prior beliefs conflict with logical validity (De Neys, 2012, 2017; De Neys & Pennycook, 2019). To this aim, we used syllogisms in which the conclusion was either consistent or inconsistent with prior beliefs, and was either logical or illogical. We measured the level of doubt experienced after the presentation of each syllogism. In line with the literature, participants gave fewer logical responses

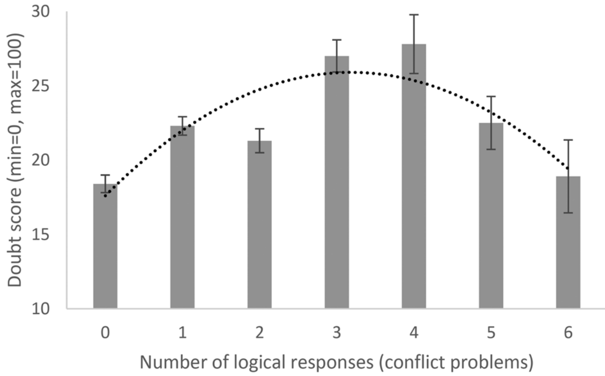

Figure 1. Level of doubt according to accuracy groups.

to conflict problems than to non-conflict ones, showing a belief bias effect when assessing the logical validity of a problem conclusion (Evans et al., 1983; Evans & Pollard, 1990). In addition, the level of doubt was higher for conflict problems than for non-conflict problems, and this was true even when focusing only on belief-driven responses. The latter finding is consistent with the idea that participants can detect (to some extent at least) when two conflicting intuitions emerge (one logical and one belief-based; De Neys, 2012; De Neys & Bonnefon, 2013), and this results in an increase in doubt. Interestingly, participants who gave either few or many logical responses experienced less doubt than participants in the medium range. It is possible that participants at both extremes felt a lower level of conflict between intuitions and could just rely on their strong (dominant) intuition (belief-based intuition for participants at the low end, and logical intuition for participants at the high end; Bago & De Neys, 2019; Raoelison et al., 2020; Thompson et al., 2018). Conversely, participants who gave an intermediate number of logical responses (i.e., with a medium level of accuracy) possibly experienced a stronger conflict between belief-based and logical intuitions, and this may have resulted in a stronger feeling of doubt.

Experiment 1 also showed that participants who doubted more also reported higher levels of state anxiety at the end of the experiment, which confirms the negative valence of conflict detection (Cortial et al., 2025). While we cannot conclude whether anxiety-prone people tend to doubt more or whether the experience of doubt leads to an increase in state anxiety, the present findings suggest that the two types of emotions overlap at least partially.

Finally, there was a negative relationship between (i) accuracy in the non-conflict condition and (ii) state anxiety and doubt. Although this relationship was not predicted, it suggests that the measure of doubt captures not only conflict detection but also other factors. For example, the measure of doubt may also reflect general stress that is unrelated to the problems themselves, but is linked to the reasoning situation. This in turn may reduce cognitive resources available for problem solving and generate some errors in the (easy) non-conflict condition (Moran, 2016).

In sum, Experiment 1 showed a link between cognitive conflict and doubt. Critically, our findings hinge on the assumption that intuitions were central to our task. This is why we limited response time to ten seconds, in order to promote intuitive reasoning. While this allotted time is relatively short, it remains longer than in some previous studies (e.g., Boissin et al., 2023), making it possible that deliberation was involved to some extent. In Experiment 2, we decreased the allotted time even more, in order to further limit the involvement of deliberation during the production of the problem response. In addition, we gave participants the opportunity to give a second response to each problem, with no time constraint, which allowed us to test the hypothesis that deliberation is triggered in situations where an appraisal of cognitive conflict generates doubt. Specifically, this two-response procedure allowed us to examine the involvement of deliberation using three indices: the time it took to give the second response, the number of response changes (i.e., when participants gave a second response which was different from the first one), and participants' self-report of deliberation (SRD). We tested whether an increase in doubt is associated with an increase in deliberation.

Experiment 2 also allowed us to explore the idea that the strength of doubt affects the type of deliberation that is triggered (Cortial et al., 2025). While weak doubt may simply trigger a rationalisation strategy, which involves finding a justification to validate and reinforce a given intuition (also known as motivated reasoning; Evans, 1996; Kunda, 1990; Wason & Evans, 1974), stronger doubt may push reasoners to engage in a more costly strategy known as decoupling (Pennycook et al., 2015; Stanovich, 1999, 2004, 2009). Decoupling consists in finding a justification to confirm or disconfirm one's intuitions, for instance when two intuitions have equal strength. Specifically, we focused on the subset of trials where participants had reported deliberating, and compared response-change with no-response-change trials within this subset. We postulate that decoupling is more likely to lead to a response change, and should thus be associated with longer RTs (more effort) and stronger doubt, compared to rationalisation (absence of response change). Finally, in Experiment 2, we further explored the emotional nature of doubt by adding a subjective measure of arousal.

Experiment 2

Method

Pre-registration

The study design, hypotheses, research question and analyses were preregistered on the Open Science Framework (https://osf.io/hdwpx)

Participants

Participants were recruited through e-mails sent first to students from the University of Clermont-Auvergne (using a different mailing list than that used in Experiment 1) and then to students from the University of Paris 8 to complete the sample. Participants were also recruited through social networks. A total of 831 people participated, of whom 793 passed the attention test (528 women, $M = 23.4$ years, $SD = 6.9$) and were included in the analyses. Five reported having completed only primary school, 411 had completed secondary school, 192 had completed an undergraduate degree, and 185 had completed a postgraduate degree or higher. Due to changes in the display of the experiment compared with Experiment 1, we instructed the participants to complete the task on a computer and not on a mobile device.

Our preregistered goal was to reach at least 674 participants based on the effect sizes of the ANOVA between accuracy groups and doubt level observed in Experiment 1. Power calculations performed with G*Power (version 3.1.9.7) indicated that a minimum of 674 participants would be needed to detect an effect size of eta^2= 0.02. with a power of 80% in a ANOVA F-test ($p = 0.05$). Given that enrolment rate was slower than in Experiment 1, we had to revive participation by disseminating the experiment through another mailing list, leading the participant number to exceed that which was pre-registered.

Materials

Syllogisms with two responses. Experiment 2 was conducted online using the same tools as in Experiment 1 (i.e., online presentation using the OpenSesame software; Mathôt et al., 2012). We also used the same type of syllogisms as in Experiment 1 but with different contents, also coming from Boissin et al. (2023). In Experiment 1, argument content was not randomly assigned to conflict and non-conflict conditions, potentially confounding the conflict effect. To address this confound in Experiment 2, participants were randomly allocated to the first or second list, and problems, within the lists, were presented in a random order for each participant. Here is an example of the same syllogism in its conflict and non-conflict forms:

Non-Conflict (believable and logically valid):

All planets revolve around the sun.
The Earth is a planet
Therefore, the Earth revolves around the sun.

Conflict (believable and logically invalid):

All planets revolve around the sun.
The Earth revolves around the sun.
Therefore, the Earth is a planet

In Experiment 2, participants responded to each problem using a two-response procedure (Thompson et al., 2011). In each trial, they first provided a "fast" response (within 3 s after the presentation of the problem conclusion) followed by a second "slow" response (with no time limit). The time limit of 3 s was chosen based on studies indicating that it corresponded to the time it took to read the conclusion, move the mouse, and select an answer (Bago & De Neys, 2017; Raoelison et al., 2020). This method allowed us to capture both an initial "intuitive" response and then a final "deliberate" one (Boissin et al., 2023). For each participant, we computed the number of correct responses (min = 0, max = 6) both for conflict and non-conflict problems, separately for time 1 (t1) and time 2 (t2) responses. We also measured response time for t2 responses, and number of response changes between t1 and t2 responses. The latter two indices were used to assess the involvement of deliberation (Bago & De Neys, 2017; Boissin et al., 2023; Thompson et al., 2011). Finally, we introduced a third self-report measure of deliberation, not previously employed in studies using the two-response paradigm. For each reasoning problem, after giving their second slow response, participants were asked whether (1) they had "given some thought before giving the second response", or whether (2) they had "kept their initial answer without giving any more thought to the problem". For each participant, we computed a self-report deliberation index (min = 0, max = 6) both for conflict and for non-conflict problems.

Doubt. Given the strong internal consistency of the doubt scale in Experiment 1, we reduced the scale to three items, keeping one item for each dimension of doubt (appraisal, subjective experience and negative valence), using the same Likert scale with four options ("not at all", "not really", "quite so" or "completely"). The three items that we kept were "I felt blocked", "I felt doubt", and "I found the reflection unpleasant". The doubt scale showed a good internal consistency ($\alpha = .82$). For each participant and each problem, we calculated a doubt score by averaging the scores to the three items. For each participant, we computed the mean doubt score for conflict problems, and the mean doubt score for non-conflict problems, and transformed the two scores into percentages.

Arousal. Participants reported how "they felt when solving the problem" using a Manikin arousal scale, from calm/serene to restless/awake/aroused (Bradley & Lang, 1994). For each participant, we computed the mean arousal score for conflict problems, and that for non-conflict problems, and transformed the two scores into percentages.

State-trait anxiety inventory (STAI). At the end of the experiment, the participants completed a state-anxiety questionnaire, using the same reduced 5-item version of the State-Trait Anxiety Inventory (STAI; Zsido et al., 2020) as in Experiment 1.

Procedure

Participants were instructed that the experiment would take between 10 and 15 min and would require their full attention. A general description of the task was presented at the beginning, informing participants that they would have to solve reasoning problems and provide two consecutive answers for each. They were informed that we were interested in the very first response that came to their mind. They were also instructed that after providing their initial response for each problem, they could reflect once more on the problem and take as much time as necessary to provide a second (final) response. In order to familiarise themselves with the two-response procedure, the participants solved a training problem under the conditions of the two-response procedure.

Figure 2 shows a typical trial. In Experiment 2, we adopted the presentation format of Bago and De Neys (2017). All trials began with the presentation of a fixation cross for 2000 ms, followed by the first sentence (the major premise, e.g., "All mammals can walk"), for 2000 ms, and then the second sentence (the minor premise, e.g., "Whales are mammals"), which was presented directly underneath the first premise for another 2000 ms. This was followed by the presentation of the conclusion, directly underneath the second premise, and by the question "Does the last statement logically follow from the first two?" with two response options (yes/no). At this stage, participants had 3000 ms to choose an answer (i.e., they had 7000 ms in total to read the entire problem and choose an answer; see Figure 2). The response buttons turned red 2000 ms after the conclusion appeared. This warned the participants that they only had a short time left to respond. Whether the participants had time to provide an answer or not, the problem disappeared and a doubt scale appeared. Finally, the same reasoning problem was presented again in full (all at once) and participants were asked to provide a final response, with no time limit. Following the second response, participants were asked to report whether they had used reflection to give their second response (see Figure 2).

As in Experiment 1, an attention test was randomly presented during the course of the experiment. Participants were asked to check the 6th box in a 10-box panel and to choose the "no" box from two choices (yes/no).

At the end of the study, the participants completed a 5-item scale measuring state anxiety and they provided their demographic information.

Ethics

This experiment was approved by the Ethics Committee of the University of Clermont-Auvergne (IRB00011540-2023-17).

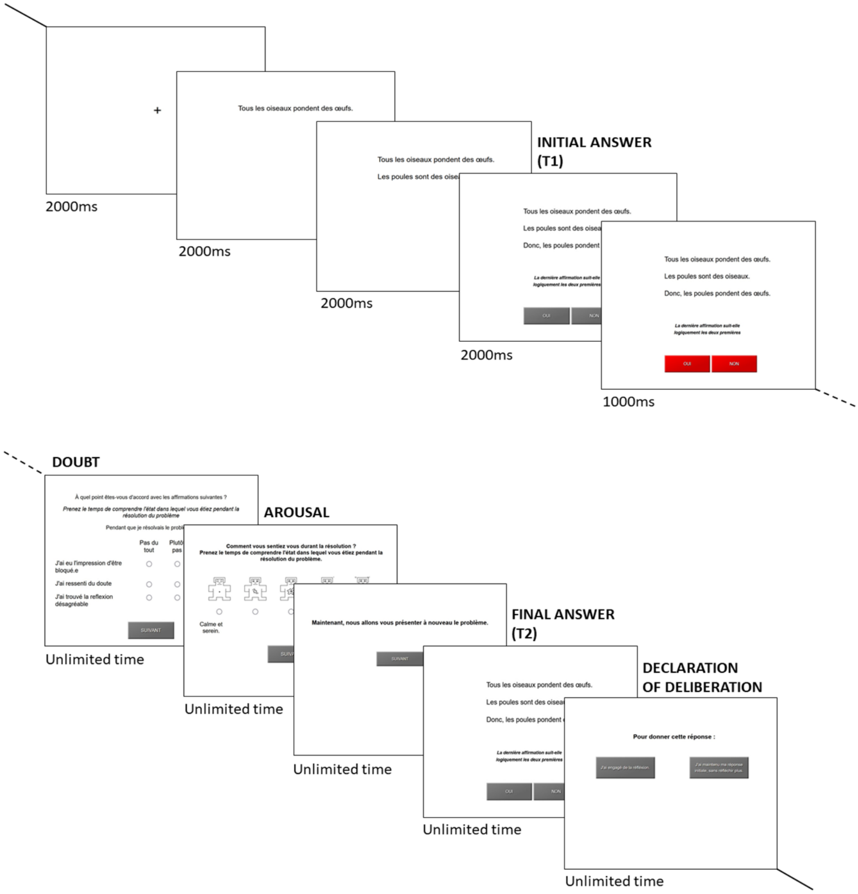

Figure 2. Illustration of one test trial in Experiment 2.

Data analyses

The data were processed using Python and the Pandas library. For each participant, we calculated the total number of logical t1 and t2 responses to non-conflict and to conflict problems (min = 0, max = 6, for each score). Given that the number of missing responses was low (4.97%), like in Experiment 1, both an absence of response and a non-logical response were considered an error (the conclusions were similar when using only non-logical responses as errors). Given that the distributions of relevant scores did not follow normal distributions, we used non-parametric statistical tests wherever possible.

Results

On average, participants took around 13 min to complete the experiment ($M = 773$ s). We excluded the participants who took less than half this time or more than twice this time, under the assumption that these participants did not engage with the experiment. We analysed the data of the remaining 695 participants.

Pre-registered analyses

Conflict and doubt. As anticipated, the mean number of correct t1 responses was high for non-conflict problems ($M_{non\text{-}conflict}$ = 4.43 out of 6; $SD_{non\text{-}conflict}$ = 1.62) confirming that the participants understood the task. In contrast, the mean number of correct t1 responses was low for conflict problems ($M_{conflict}$ = 1.13 out of 6; $SD_{conflict}$ = 1.39). The difference was significant, $W = 996$, $p < .001$, $r = .99$. Although participants gave more accurate final (than initial) responses, the difference in accuracy between non-conflict and conflict problems remained significant ($M_{non\text{-}conflict}$ = 5.38, $SD_{non\text{-}conflict}$ = 0.9; $M_{conflict}$ = 2.23, $SD_{conflict}$ = 1.78), $W = 550$, p = <.001, r = .99. One of the key hypotheses of this study was that, after giving their initial "fast" answer, participants would experience greater levels of doubt on average in the context of conflict than in the absence of conflict. This was indeed the case ($M_{conflict}$ = 15.1 $SD_{conflict}$ = 15.3; $M_{non\text{-}conflict}$ = 13.7, $SD_{non\text{-}conflict}$ = 13.8), $W = 93946$, p = .002, r = .15. Subjective arousal was also higher in the context of conflict than in its absence, $W = 85957$, p <.001, $r = 0.17$ (see Table 5).

Spearman correlation analyses revealed that the level of doubt was related to the number of logical responses for conflict problems both at t1 (ρ = .17) and at t2 (ρ = .23, see Table 6). We also found positive correlations between doubt, arousal, and level of state-anxiety (see Table 6). Furthermore, for the non-conflict problems, we found the same negative correlation as in Experiment 1 between the number of logical t1 responses and the level of doubt (ρ = −0.16, p < .001). The interaction between doubt and conflict on accuracy at both T1 and T2 was further supported by GLM analyses, with full results reported in the Experiment 2 analyses of the Supplementary Materials. We also found a negative correlation between the number of logical t1 responses and the level of arousal (ρ = −0.10, p = .01).

Conflict and deliberation. In Experiment 2, we introduced three indices in order to measure the involvement of deliberation: self-report of

Table 5. Mean and SD of accuracy (at t1 and t2), doubt, and arousal in response to conflict and non-conflict problems.

Number of logical responses				Doubt		Arousal	
Initial response (t1)		Final response (t2)					
Conflict	Non-Conflict	Conflict	Non-Conflict	Conflict	Non-conflict	Conflict	Non-conflict
1.13 (*1.39*)	4.43 (*1.62*)	2.23 (*1.78*)	5.38 (*0.90*)	15.1 (*15.3*)	13.7 (*13.8*)	20.2 (*19.3*)	19.0 (*18.3*)

Table 6. Correlation between accuracy measures and emotionals measures.

	Doubt		Arousal	
	Conflict	Non-Conflict	Conflict	Non-Conflict
Arousal	.65***	.64***	–	–
State-Anxiety	.50***	.48***	.48***	.48***
Logical responses (t1)	.17***	−0.16***	.15***	−0.10***
Logical responses (t2)	.23***	−0.05	.16***	−0.04

Note: The table presents Spearman's ρ correlation values.
*p <.05. **p < .01. ***p < .001.

Table 7. Relationship between self-report of deliberation, t2 RT, and response change.

	SRD		RespChange		RT	
	Conflict	Non-Conflict	Conflict	Non-Conflict	Conflict	Non-Conflict
RespChange	.36***	.17***	–	–	–	–
t2 RT	.45***	.42***	.34***	.09*	–	–
Logical responses (t1)	.26***	−0.09*	.12***	−0.15***	.25***	−0.11**
Logical responses (t2)	.41***	−0.07	.40***	−0.54***	.43***	−0.02

Note: The table presents Spearman's ρ correlation values.
*p <.05. **p < .01. ***p < .001.

deliberation (SRD), change in response from t1 to t2 (RespChange), and response time (RT) for the t2 (final) response.

We observed positive relationships across all three measures of deliberation in conflict problems, suggesting that they all measured the same construct. We also found positive correlations between deliberation measures and logical responses for conflict problem, but negative ones for non-conflict problems (see Table 7).

Participants reported having deliberated significantly more often for conflict than for non-conflict problems (see Table 8), $W = 69963$, p <.001, r = .31. They also changed their response more often from t1 to t2, $W = 53308$, p <.001, r = .51, and took longer to give their second response, $W = 154235$, p <.001, r = .27 (see Table 8), with conflict than with non-conflict problems.

Doubt and deliberation. One of our key predictions is that doubt triggers deliberation. In order to test this hypothesis, we examined the relationships between the indices of doubt (doubt score, arousal score, and state-anxiety score) and the indices of deliberation (self-report of deliberation, response change from t1 to t2, and t2 RT). Note that we consider every response change as a marker of deliberation (even if the final answer is wrong, as it shows a reassessment of the initial intuition; Bago & De Neys, 2017, 2019; Boissin et al., 2021, 2022, 2023). We found positive relationships between all these indices (see Table 9).

To analyse the different patterns of responses, logical responses were coded as "1" and non-logical ones were coded as "0", at each response

Table 8. SRD, ChangeResp and RT means difference between conflict and non-conflict problems.

	Conflict		Non-conflict	
	M	*(SD)*	*M*	*(SD)*
Self-report deliberation score (Min = 0, max = 6)	3.06	*(1.99)*	2.69	*(2.01)*
Response Change (frequency)	0.17	*(0.21)*	0.09	*(0.14)*
t2 RT (ms)	8220	*(6000)*	7230	*(5772)*

Table 9. Relationships across emotional and deliberation indices.

	Doubt		Arousal		State-anxiety	
	Conflict	Non-conflict	Conflict	Non-conflict	Conflict	Non-conflict
SRD	.34***	.34***	.24***	.25***	.13***	.12**
RespChange	.22***	.10***	.14***	.12**	.12**	.05
t2 RT	.35***	.34***	.32***	.31***	.20***	.17***

Note: The table presents Spearman's ρ correlation values.
*p <.05. **p < .01.
***p < .001. There is only one measure of state-anxiety per participant, we compare it to the conflict or non-conflict status of the deliberation measures (SRD, RespChange and t2 RT).

Table 10. Doubt and arousal mean across pattern groups.

	00		11		10		01	
	M	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*
Doubt (%)	9.5	*(17.5)*	11	*(18.8)*	20.8	*(26.6)*	21.9	*(25.5)*
Arousal (%)	16.3	*(22.5)*	17.25	*(22.67)*	26.0	*(27.5)*	24.75	*(25.75)*
RT (ms)	4918	*(4113)*	5494	*(4363)*	6691	*(5256)*	9026	*(8639)*

stage (i.e., t1 and t2). This gave us four possible response patterns. "11" patterns correspond to trials where both the first (fast) response and the slow (second) response were correct (35% of all responses). "00" patterns correspond to trials where both responses were incorrect (54% of all responses). "01" patterns correspond to trials where the first response was incorrect and the second was correct (7% of all trials). Finally, "10" patterns correspond to trials where the first response was correct and the second incorrect (4% of all trials).

Kruskal-Wallis ANOVAs revealed that the reported levels of doubt and arousal, as well as t2 RTs, varied as a function of response patterns (see Table 10), respectively, $\chi^2(3) = 167$, p <.001, $\varepsilon^2 = .025$, $\chi^2(3) = 85.2$, p <.001, $\varepsilon^2 = .013$, and $\chi^2(3) = 270$, p <.001, $\varepsilon^2 = .041$. Furthermore, Mann-Whitney tests suggest the levels of doubt and arousal were higher, respectively $U_{doubt} = 1.64^{e+6}$, p <.001, r=0.24, and $U_{arousal} = 1.77^{e+6}$, p <.001, r=0.18, and RTs were slower, $U_{RT} = 1.48^{e+6}$, p <.001, r=0.32, in patterns involving a response change (01 or 10 patterns) compared to patterns not involving a response change (11 or 00 patterns).

Finally, a χ^2 test revealed that the proportions of self-report deliberation varied as a function of response patterns (see Figure 3), $\chi^2(3) = 532$, p <.001.

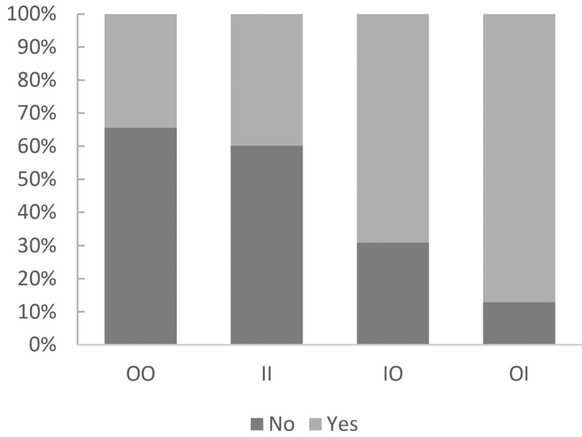

Figure 3. Distribution of self-report of deliberation answers according to response patterns.

Differences between reasoners. Like in Experiment 1, we grouped participants as a function of their number of logical responses, using the first (fast) response. Given that there were only three participants who gave 6 logical responses, we grouped them together with the participants who gave five logical responses, leading to 6 groups in total ("0", "1", "2", "3", "4", and "5+" logical responses). A Kruskal-Wallis ANOVA showed that the level of doubt varied as a function of group, $\chi^2(5) = 28$, $p < .001$, $\varepsilon^2 = .04$, and that the relationship was best explained by a quadratic trend, like in Experiment 1, $t(689) = -3.28$, $p = .001$ (see Figure 4).

Frequency of response change, and RTs for the second response, also varied as a function of group, respectively, $\chi^2(5) = 33.9$, $p < .001$, $\varepsilon^2 = .050$ and $\chi^2(5) = 49$, $p < .001$, $\varepsilon^2 = .071$, with significant quadratic trends, respectively, $t(689) = -2.42$, $p = .016$ and $t(689) = -2.97$, $p = .003$ (see Figure 5).

Finally, self-reports of deliberation and arousal also varied across groups, respectively, $\chi^2(5) = 52.3$, $p < .001$, $\varepsilon^2 = .075$ and $\chi^2(5) = 19.7$, $p = .001$, $\varepsilon^2 = .028$, but were qualified by linear trends, respectively, $t(689) = 5.59$, $p < .001$, and $t(689) = 3.15$, $p = .002$ (see Figure 5).

Exploratory analyses

Subset of belief-driven t1 responses. Like in Experiment 1, we tested if the difference in levels of doubt across conflict and non-conflict problems remained significant when focusing only on belief-driven t1 responses (i.e., responses that were correct in non-conflict trials but incorrect in conflict trials). We excluded participants for whom the mean could not be calculated (because of an absence of belief-driven

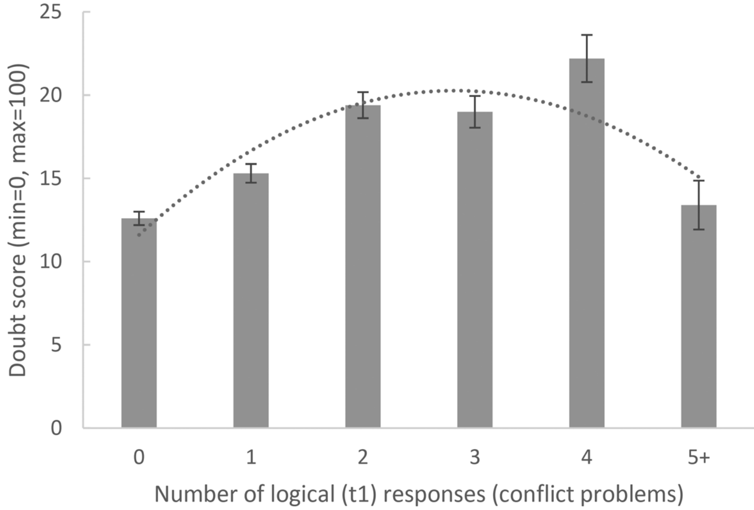

Figure 4. Level of doubt according to the number of logical t1 responses in conflict condition.

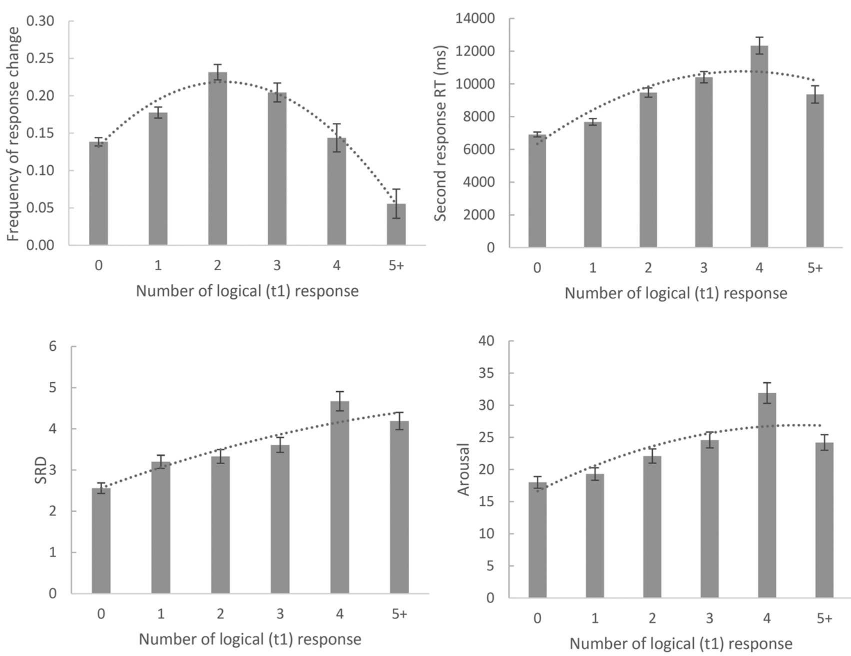

Figure 5. The change of response rates, RT, SRD and arousal according to the number of correct t1 responses in conflict condition.

response) and focused on the subset of remaining participants ($N = 677$). A Wilcoxon signed-rank test showed that even when focusing on belief-driven responses, participants reported more doubt in the conflict condition ($M = 13.3$, $SD = 13.6$) than in the non-conflict one ($M = 9.01$, $SD = 16.2$), $W = 31325$, $p < .001$, $r = .42$.

Subset of self-reported deliberative trials. Figure 5 shows that participants who reported having deliberated had not necessarily changed their response from t1 to t2. This might reflect the fact that deliberation can lead either to rationalisation or decoupling of ones' responses. Within the subset of self-reported deliberative trials, we compared the trials with no response change ("00" and "11" response patterns grouped together) to those with response change ("01" and "10" response patterns grouped together), using Mann-Whitney t-tests. We found that RTs and levels of doubt were lower in no-response-change trials (respectively, M_{RT} = 9301 ms, SD_{RT} = 10600, and M_{doubt} = 16.8, SD_{doubt} = 22) than in response-change trials (M_{RT} = 14770 ms, SD_{RT} = 16903; M_{doubt} = 24.7, SD_{doubt} = 26.2), U_{RT} = 722173, p <.001, r = .17; U_{doubt} = 626550, p <.001, r = .28. This finding is consistent with the idea that, when participants reported that they reflected but did not change their mind, they may have been rationalising (which involves lower levels of doubt and is faster). In contrast, when participants reported that they reflected and actually did change their mind, they may have been decoupling (which involves higher levels of doubt and is slower).

Discussion

Experiment 2 was designed with two objectives in mind. The first objective was to replicate the results of Experiment 1, which related cognitive conflict in deductive reasoning to doubt while suggesting a quadratic relationship between the level of doubt and the number of logical intuitive responses. The second goal was to examine the purported association between doubt and deliberation, which we measured using three indices (t2 response time, response change, and self-reported deliberation).

Doubt manifested itself in both the cognitive measures (i.e., feeling blocked, experiencing doubt) and the affective measures (i.e., unpleasantness, arousal). It was stronger in conflict than non-conflict trials, even when focusing solely on trials where participants gave belief-driven responses. Finally, doubt was overall stronger in participants with medium levels of t1 accuracy on conflict trials than in participants with low or high levels of t1 accuracy. These findings all converge towards the following interpretation: (1) conflict problems can simultaneously generate heuristic-based (flawed) and logically-sound intuitions, (2) the intuitive conflict can be detected and this generates a negative-valence emotion, and (3) the strength of the conflict varies across participants, such that in some participants it is stronger than in others (i.e., the latter are participants with either a strong heuristic-based intuition or a strong logically-sound intuition, whereas the former are participants with two intuitions of equivalent strength; De Neys, 2012, 2017; De Neys & Pennycook, 2019).

With respect to the second aim of the study, we found positive correlations between doubt and the three measures of deliberation. In addition, the findings are consistent with the idea that stronger doubt is associated with a deeper and more costly type of deliberation (decoupling), whereas weaker doubt is associated with a weaker and less costly type of deliberation (rationalising).

General discussion

The present study examined two main hypotheses. First, we tested whether the emotion of doubt emerges when intuitive reasoning fails to produce a straightforward answer to a problem, specifically, when two contradicting intuitions emerge at the same time and come into conflict. Second, we explored whether this emotional experience drives deliberation. To this end, we contrasted the level of doubt experienced when solving easy deductive-reasoning problems (non-conflict problems, unlikely to produce conflicting intuitions) to that experienced when solving difficult problems, in which several contradicting intuitions can emerge (conflict problems; Experiments 1 and 2). We further assessed whether the experience of doubt was associated with deliberation using a set of indices (Experiment 2). In both experiments, conflict problems generated a stronger emotional response (subjective doubt and subjective arousal) than non-conflict problems. In Experiment 2, all the measures of deliberation (reaction time, self-report of deliberation, and response changes in the two-response paradigm) were positively related with the emotional indices.

The association between conflict in deductive reasoning problems and emotion (i.e., subjective doubt and subjective arousal) is consistent with previous findings showing that discrepant, discontinuous or incoherent stimuli generate an emotional response (De Neys et al., 2010; Hoshikawa & Yamamoto, 1997; Kuzinas et al., 2016; Morris et al., 2008; Purcell et al., 2023; Van Steenbergen & Band, 2013; Zander et al., 2017; Zhang et al., 2010). Moreover, the correlation between emotional response and deliberation is consistent with previous findings indicating that certain emotional states can facilitate a deliberative mode of reasoning. A substantial body of literature demonstrates that emotions such as surprise, confusion, or anxiety indeed leads to reflection (see Cortial et al., 2025). All of these emotions share common traits of doubt, namely that they arise from a disrupted and non-fluid interaction between the organism and its environment and have a negative valence. The results of our study indeed indicate a relationship between doubt and anxiety. We argue that anxiety may represent a long-lasting form of doubt associated with a repeated exposure to ambiguous or problematic situations that cannot be satisfyingly solved through intuitive reasoning and that are susceptible to lead to a negative impact on the well-being of the organism (Miceli & Castelfranchi, 2005; Spielberger, 1966). Anxiety may thus emerge as a result of a prolonged state of doubt confirming the presence of a doubt-like affective element embedded into the reasoning process.

Although some models seem to have made reference to a "feeling" that would be associated with intuitions and that can prompt deliberation (Thompson et al., 2011; Thompson & Morsanyi, 2012), the remaining theoretical question is: What would be the exact role of emotion in FOR models? FOR, being a metacognitive experience, is built upon cues (Ackerman & Thompson, 2017; Efklides, 2001, 2006), one of which is the positive emotional valence arising from the fluency of strong intuition (Thompson & Morsanyi, 2012). Thus, if FOR is not an appraisal, there is no reason to view it as the motivation to deliberate, as emotional motivation already exists. An individual appraises non-fluent intuition as negative, experiences a physiological reaction, and feels action tendencies, which sufficiently explains the motivation to use deliberation. FOR would then acts as a metacognitive judgement to assess and direct the necessary time and effort allocated to problem-solving (De Neys, 2022; Efklides, 2006; Touroutoglou & Efklides, 2010).

We believe that an important finding from this study is that the experience of doubt during problem-solving is not entirely caused by cognitive conflict or disfluent intuitions. Reasoning contexts are likely to lead to nervousness, desire to perform well, or fear of failure, which would explain why doubt was also experienced to some extent in non-conflict problems, and was negatively linked to non-conflict-problem accuracy. According to the appraisal framework, doubt arises from an individual's overall evaluation of goal-situation congruence rather than from cognitive conflict alone. While conflict was used to induce doubt experimentally, some participants may have suspected a trick, leading them to overthink and give the wrong answer in non-conflict trials.

The capacity to elicit intuition is intrinsically dependent on the prior knowledge of individuals, also named Mindware (Raoelison et al., 2020; Stanovich, 2018; Thompson et al., 2018). The literature provides compelling evidence of substantial individual differences in this regard (de Chantal et al., 2020; Frey et al., 2018; Stanovich & West, 2000). It is also well established that a form of affective mindware plays a crucial role in reasoning. Individuals who have engaged in reasoning exercises related to traumatic events they have experienced—such as war or sexual violence—demonstrate enhanced reasoning abilities concerning these specific issues, as these contexts hold particular relevance for them (Blanchette et al., 2007; Blanchette & Caparos, 2013, 2016). The importance of individual differences is well accounted for by the Appraisal framework, explaining why some people may have found non-conflict problems puzzling and doubtful in the present studies (Moors & Fischer, 2019; Scherer & Moors, 2019; Silvia et al., 2009).

Our findings showed that poorer performance does not mean that people did not deliberate; in fact, the opposite was sometimes true. Doubt motivated people to mobilise cognitive resources to deliberate, but this was beneficial only for conflict problems. For non-conflict problems, doubt led individuals to question their intuition, but looking for a trap where there is not one will mechanically lower the level of accuracy.

In light of the present results, we propose that 1) a negative emotional response delineates the mechanistic processes involved in reasoning, 2) the intuition-to-deliberation switch mechanism is triggered by a general appraisal of a situation and 3) metacognitive experiences such as Feeling of Rightness or Response Uncertainty may not motivate to deliberate per se, but allow individuals to tailor the time and effort required by the task.

To be clear about the novelty of our proposition, we argue that distinguishing between uncertainty—conceptualised as a cognitive judgement of a lack of information—and doubt—conceptualised as a negative emotional response to a perceived incongruence in the organism–environment interaction—would be theoretically valuable but methodologically challenging. Doubt emerges from an appraisal process and thus incorporates elements of uncertainty, making a complete separation between uncertainty and doubt difficult. Although physiological measures such as pupillometry (Purcell et al., 2023) and electrodermal activity (De Neys et al., 2010) offer promising avenues, even with such tools, isolating the cognitive component (uncertainty) from the emotional component (doubt) is difficult, as the latter is predicated on the presence of the former. The present work captured the emotional dimension using a self-report tool comparable to previous measures (e.g., Feeling of Rightness, Thompson et al., 2011; or Response Uncertainty, De Neys, 2022), but with a specific emphasis on emotion. Our findings indicate that self-reported doubt and confusion—which can be taken as indices of uncertainty—are associated with unpleasantness and state anxiety. Importantly, these results support the relevance of using an appraisal-based framework over a purely metacognitive one, as they can account for both the cognitive and the affective dimensions of the monitoring of intuitions. Future research should aim to confirm that high RU or low FOR scores (employing the same measures as used in prior work) are associated with negative emotion.

A last point that we will discuss is the finding that the participants who experienced the lowest level of doubt overall were those who were either in the lower or in the higher range of accuracy. In contrast, participants in the mid-range experienced higher levels of doubt. These findings are consistent with current dual-process models, in that the most and least accurate reasoners are likely not to have experienced much conflict when solving conflicting problems because the strength of one intuition (either the belief-driven or the logical one) would have greatly exceeded the strength of the other. In contrast, mid-range accuracy may reflect stronger conflict between belief-driven and logical intuitions (Bago & De Neys, 2017, 2019; De Neys, 2012; Raoelison et al., 2020, 2021; Thompson et al., 2018). In addition, our results show that the 'number of logical responses' is not an optimal measure of the use of deliberation, and they add to the literature which challenges the notion that deliberation solely serves a corrective function when intuition "goes wrong" (De Neys, 2022). Deliberation appears to be relevant mostly in situations where intuition does not generate a straightforward response.

Our data are also consistent with the idea that the level of doubt is related to the depth of deliberation. Accordingly, weak doubt may be related to more superficial strategies of deliberation such as rationalisation, which aims at confirming why one's intuition is an appropriate response to a given problem. In contrast, strong doubt may be related to a deeper and more analytical form of deliberation, decoupling, which aims at challenging, and possibly disconfirming, an intuition. Although tentative, this interpretation of the present data is consistent with current models (Pennycook et al., 2015). It is also consistent with the idea that deliberation is not a unique process but may instead encompass several alternative mechanisms, such as rationalisation and decoupling.

In this vein, the present model of aversive doubt (Cortial et al., 2025) posits that, beyond cognitive abilities and mindware, an individual's capacity for emotional regulation—along with their distinctive way of experiencing emotions to varying degrees—may significantly influence their engagement in deliberation. Specifically, individuals who are prone to feeling overwhelmed by negative emotions, or who exhibit a marked intolerance towards them, are likely to experience doubt differently and/or to react to doubt differently. This variation in emotional processing is expected to have direct consequences on the reliance on deliberation.

We now raise a few limitations of the present work. First, we defined and measured doubt using only self-reports. Nevertheless, we note that our findings about participants' emotional state in situations of cognitive conflict are consistent with physiological findings observed elsewhere (De Neys et al., 2010; Purcell et al., 2023). Second, further studies may explore the source of doubt. While in the present study the impact of complex deductive-reasoning problems on doubt may be (in part) attributable to the conflict between logicality and believability, it may also be linked to a more general appraisal of the situation. Third, we emphasise our awareness of the limitations of our measure of doubt. This is one reason why the measurement changes from Experiment 1 to Experiment 2. While this measure allows us to establish that an emotional phenomenon is clearly involved in the transition from intuition to deliberation, we do not claim to have fully captured this phenomenon through our doubt scale. We therefore encourage future studies to refine and improve this measurement.

Another limitation of our study, and of previous studies using similar paradigms, is that we cannot ascertain that we have successfully isolated intuition from deliberation. However, even if deliberation was involved to some extent in the first "fast" answer, this would not invalidate the fact that doubt was found to be intimately linked to deliberation. A final limitation is that we limited the study to deductive-reasoning syllogisms The results of the present study would have to be generalised to other instances of complex/conflict reasoning problems. Examples of such problems include base-rate neglect problems (Tversky & Kahneman, 1974) and bat-and-ball problems (Frederick, 2005). Other studies have demonstrated that these types of problems also result in reduced confidence in responses. It is thus likely that the findings of this study would extant to other types of reasoning problems (Bago & De Neys, 2017; Białek & De Neys, 2016; De Neys et al., 2008; Thompson & Johnson, 2014).

In sum, the present findings substantiate the notion that an emotional state of doubt is associated with intuitive failure and the trigger of deliberation. Doubt is characterised by two distinct attributes: a negative valence and an origin embedded in the appraising of a discontinuous or incoherent situation, such as in cognitive conflict. We propose that conflict situations may elicit doubt, which itself prompts deliberation. The role of deliberation would be to lift the negative emotional state of doubt. The present findings place an emotional response at the centre of the process involved in the switch from intuition to deliberation, and thereby at the heart of dual-process models of reasoning. Our goal is not to reduce the entire question of the triggering of deliberation to a matter of emotional response, but rather to propose that the Appraisal framework should be fully integrated to the field of human reasoning.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

No funding.

ORCID

Cédric Cortial http://orcid.org/0000-0002-8866-5183
Jérôme Prado http://orcid.org/0000-0003-3548-5610
Serge Caparos http://orcid.org/0000-0001-6922-4449

Data availability statement

Raw data of both studies can be downloaded from our OSF page (https://osf.io/cyuvd)

References

Ackerman, R., & Thompson, V. A. (2017). Meta-reasoning: Monitoring and control of thinking and reasoning. *Trends in Cognitive Sciences*, *21*(8), 607–617. https://doi.org/10.1016/j.tics.2017.05.004

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

Białek, M., & De Neys, W. (2016). Conflict detection during moral decision-making: Evidence for deontic reasoners' utilitarian sensitivity. *Journal of Cognitive Psychology*, *28*(5), 631–639. https://doi.org/10.1080/20445911.2016.1156118

Blanchette, I., & Caparos, S. (2013). When emotions improve reasoning: The possible roles of relevance and utility. *Thinking & Reasoning*, *19*(3-4)-, 399–413. 4 https://doi.org/10.1080/13546783.2013.791642

Blanchette, I., & Caparos, S. (2016). Working memory function is linked to trauma exposure, independently of post-traumatic stress disorder symptoms. *Cognitive Neuropsychiatry*, *21*(6), 494–509. https://doi.org/10.1080/13546805.2016.1236015

Blanchette, I., Richards, A., Melnyk, L., & Lavda, A. (2007). Reasoning about emotional contents following shocking terrorist attacks: A tale of three cities. *Journal of Experimental Psychology. Applied*, *13*(1), 47–56. https://doi.org/10.1037/1076-898X.13.1.47

Boissin, E., Caparos, S., & De Neys, W. (2023). 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., & Neys, W. D. (2022). Debiasing System 1: Training favours logical over stereotypical intuiting. *Judgment and Decision Making*, *17*(4), 646–690. https://doi.org/10.1017/S1930297500008895

Botvinick, M. M. (2007). Conflict monitoring and decision making: Reconciling two perspectives on anterior cingulate function. *Cognitive, Affective & Behavioral Neuroscience*, *7*(4), 356–366. https://doi.org/10.3758/CABN.7.4.356

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

Chiesi, F., Morsanyi, K., Donati, M. A., & Primi, C. (2018). Applying item response theory to develop a shortened version of the need for cognition scale. *Advances in Cognitive Psychology*, *14*(3), 75–86. https://doi.org/10.5709/acp-0240-z

Cortial, C., Prado, J., & Caparos, S. (2025). Reconceptualizing Metacognitive Experience in Dual-Process Reasoning: The Role of Emotion in Triggering Deliberation. *Cognitive Science*, *49*(7), e70084. https://doi.org/10.1111/cogs.70084

Craig, S., Graesser, A., Sullins, J., & Gholson, B. (2004). Affect and learning: An exploratory look into the role of affect in learning with AutoTutor. *Journal of Educational Media*, *29*(3), 241–250. https://doi.org/10.1080/1358165042000283101

D'Mello, S., & Graesser, A. (2014). Confusion and its dynamics during device comprehension with breakdown scenarios. *Acta Psychologica*, *151*, 106–116. https://doi.org/10.1016/j.actpsy.2014.06.005

David, L., Vassena, E., & Bijleveld, E. (2024). The unpleasantness of thinking: A meta-analytic review of the association between mental effort and negative affect. *Psychological Bulletin*, *150*(9), 1070–1093. https://doi.org/10.1037/bul0000443

de Chantal, P.-L., Newman, I. R., Thompson, V. A., & Markovits, H. (2020). Who resists belief-biased inferences? The role of individual differences in reasoning strategies, working memory, and attentional focus. *Memory & Cognition*, *48*(4), 655–671. https://doi.org/10.3758/s13421-019-00998-2

De Neys, W. (2012). Bias and Conflict A Case for Logical Intuitions. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *7*(1), 28–38. https://doi.org/10.1177/1745691611429354

De Neys, W. (2017). Bias, conflict, and fast logic: Towards a hybrid dual process future?. In *Dual Process Theory 2.0*. Routledge.

De Neys, W. (2022). Advancing theorizing about fast-and-slow thinking. *The Behavioral and Brain Sciences*, *46*, e111. https://doi.org/10.1017/S0140525X2200142X

De Neys, W., & Bonnefon, J.-F. (2013). The 'whys' and 'whens' of individual differences in thinking biases. *Trends in Cognitive Sciences*, *17*(4), 172–178. https://doi.org/10.1016/j.tics.2013.02.001

De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition*, *106*(3), 1248–1299. https://doi.org/10.1016/j.cognition.2007.06.002

De Neys, W., Moyens, E., & Vansteenwegen, D. (2010). Feeling we're biased: Autonomic arousal and reasoning conflict. *Cognitive, Affective & Behavioral Neuroscience*, *10*(2), 208–216. https://doi.org/10.3758/CABN.10.2.208

De Neys, W., & Pennycook, G. (2019). Logic, fast and slow: Advances in dual-process theorizing. *Current Directions in Psychological Science*, *28*(5), 503–509. https://doi.org/10.1177/0963721419855658

De Neys, W., Vartanian, O., & Goel, V. (2008). Smarter than we think: When our brains detect that we are biased. *Psychological Science*, *19*(5), 483–489. https://doi.org/10.1111/j.1467-9280.2008.02113.x

Dewey, J. (1910). *How we think*. Heath and Co.

Efklides, A. (2001). Metacognitive experiences in problem solving: Metacognition, motivation, and self-regulation. In *Trends and prospects in motivation research* (pp. 297–323). Kluwer Academic Publishers.

Efklides, A. (2006). Metacognition and affect: What can metacognitive experiences tell us about the learning process? *Educational Research Review*, *1*(1), 3–14. https://doi.org/10.1016/j.edurev.2005.11.001

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *The American Psychologist*, *49*(8), 709–724. https://doi.org/10.1037/0003-066X.49.8.709

Evans, J. S. B. T. (1996). Deciding before you think: Relevance and reasoning in the selection task. *British Journal of Psychology*, *87*(2), 223–240. https://doi.org/10.1111/j.2044-8295.1996.tb02587.x

Evans, J. S. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review*, *13*(3), 378–395. https://doi.org/10.3758/BF03193858

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *8*(3), 223–241. https://doi.org/10.1177/1745691612460685

Evans, J. S. B. T., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, *11*(3), 295–306. https://doi.org/10.3758/BF03196976

Evans, J. S. B. T., & Pollard, P. (1990). Belief Bias and problem complexity in deductive reasoning. In J.-P. Caverni, J.-M. Fabre, & M. Gonzalez (Éds.), *Cognitive biases* (p. 131–154). North-Holland. https://doi.org/10.1016/S0166-4115(08)61320-5

Fernandez Cruz, A. L., Arango-Muñoz, S., & Volz, K. G. (2016). Oops, scratch that ! Monitoring one's own errors during mental calculation. *Cognition*, *146*, 110–120. https://doi.org/10.1016/j.cognition.2015.09.005

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

Frey, D., Johnson, E. D., & De Neys, W. (2018). Individual differences in conflict detection during reasoning. *Quarterly Journal of Experimental Psychology (2006)*, *71*(5), 1188–1208. https://doi.org/10.1080/17470218.2017.1313283

Frijda, N. H. (1986). The emotions (p. xii, 544). Editions De la Maison Des Sciences De L'Homme.

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

Gigerenzer, G., & Todd, P. M., ABC Research Group. (1999). *Simple heuristics that make us smart (1. issued as an Oxford Univ. Press paperback)*. Oxford University Press.

Hoshikawa, Y., & Yamamoto, Y. (1997). Effects of Stroop color-word conflict test on the autonomic nervous system responses. *The American Journal of Physiology*, *272*(3 Pt 2), H1113–1121. https://doi.org/10.1152/ajpheart.1997.272.3.H1113

James, W. (1890). *Principles of psychology*. Dover.

Kahneman, D, Slovic, P, & Tversky, A (Éds.) (1982). *Judgment under Uncertainty: Heuristics and Biases*. Cambridge University Press. https://doi.org/10.1017/CBO9780511809477

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *The American Psychologist*, *58*(9), 697–720. https://doi.org/10.1037/0003-066X.58.9.697

Kahneman, D. (2011). *Thinking, fast and slow* (1st ed). Straus and Giroux.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498. https://doi.org/10.1037/0033-2909.108.3.480

Kuzinas, A., Noiret, N., Bianchi, R., & Laurent, É. (2016). The effects of image hue and semantic content on viewer's emotional self-reports, pupil size, eye movements, and skin conductance response. *Psychology of Aesthetics, Creativity, and the Arts*, *10*(3), 360–371. https://doi.org/10.1037/a0040274

Lazarus, R. S., Averill, J. R., & Opton, E. M. (1970). Towards a cognitive theory of emotion. In M. B. Arnold (Éd.), *Feelings and emotions* (pp. 207–232). Academic Press. https://doi.org/10.1016/B978-0-12-063550-4.50023-1

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Meyer, W.-U., Niepel, M., Rudolph, U., & Schützwohl, A. (1991). An experimental analysis of surprise. *Cognition & Emotion*, *5*(4), 295–311. https://doi.org/10.1080/02699939108411042

Miceli, M., & Castelfranchi, C. (2005). Anxiety as an epistemic emotion: An uncertainty theory of anxiety. *Anxiety, Stress, and Coping*, *18*(4), 291–319. https://doi.org/10.1080/10615800500209324

Moors, A., Ellsworth, P. C., Scherer, K. R., & Frijda, N. H. (2013). Appraisal Theories of Emotion: State of the Art and Future Development. *Emotion Review*, *5*(2), 119–124. https://doi.org/10.1177/1754073912468165

Moors, A., & Fischer, M. (2019). Demystifying the role of emotion in behaviour: Toward a goal-directed account. *Cognition & Emotion*, *33*(1), 94–100. https://doi.org/10.1080/02699931.2018.1510381
Moran, T. P. (2016). Anxiety and working memory capacity: A meta-analysis and narrative review. *Psychological Bulletin*, *142*(8), 831–864. https://doi.org/10.1037/bul0000051
Morris, A. L., Cleary, A. M., & Still, M. L. (2008). The role of autonomic arousal in feelings of familiarity. *Consciousness and Cognition*, *17*(4), 1378–1385. https://doi.org/10.1016/j.concog.2008.04.005
Morsanyi, K., & Handley, S. J. (2012). Logic feels so good—I like it ! Evidence for intuitive detection of logicality in syllogistic reasoning. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *38*(3), 596–616. https://doi.org/10.1037/a0026099
Muis, K. R., Psaradellis, C., Lajoie, S. P., Di Leo, I., & Chevrier, M. (2015). The role of epistemic emotions in mathematics problem solving. *Contemporary Educational Psychology*, *42*, 172–185. https://doi.org/10.1016/j.cedpsych.2015.06.003
Parkinson, B., & Manstead, A. S. R. (1992). Appraisal as a cause of emotion. In *Emotion* (pp. 122–149). Sage Publications, Inc.
Peirce, C. S. (1877). The fixation of belief. *Popular Science Monthly*, *12*, 1–15.
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
Pezzo, M. (2003). Surprise, defence, or making sense: What removes hindsight bias? *Memory (Hove, England)*, *11*(4-5), 421–441. https://doi.org/10.1080/09658210244000603
Purcell, Z. A., Roberts, A. J., Handley, S. J., & Howarth, S. (2023). Eye Movements, Pupil Dilation, and Conflict Detection in Reasoning: Exploring the Evidence for Intuitive Logic. *Cognitive Science*, *47*(6), e13293. https://doi.org/10.1111/cogs.13293
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., Thompson, V. A., & De Neys, W. (2020). The smart intuitor: Cognitive capacity predicts intuitive rather than deliberate thinking. *Cognition*, *204*, 104381. https://doi.org/10.1016/j.cognition.2020.104381
Reisenzein, R. (2001). Appraisal processes conceptualized from a schema-theoretic perspective: Contributions to a process analysis of emotions. In *Appraisal processes in emotion: Theory, methods, research* (pp. 187–201). Oxford University Press.
Roseman, I. J. (1984). Cognitive determinants of emotion: A structural theory. *Review of Personality & Social Psychology*, *5*, 11–36.
Sá, W. C., West, R. F., & Stanovich, K. E. (1999). The domain specificity and generality of belief bias: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology*, *91*(3), 497–510. https://doi.org/10.1037/0022-0663.91.3.497
Scherer, K. R., & Moors, A. (2019). The emotion process: Event appraisal and component differentiation. *Annual Review of Psychology*, *70*(1), 719–745. https://doi.org/10.1146/annurev-psych-122216-011854
Silvia, P. J., Henson, R. A., & Templin, J. L. (2009). Are the sources of interest the same for everyone? Using multilevel mixture models to explore individual differences in appraisal structures. *Cognition and Emotion*, *23*(7), 1389–1406. https://doi.org/10.1080/02699930902850528

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
Smith, C. A. (1989). Dimensions of appraisal and physiological response in emotion. *Journal of Personality and Social Psychology*, *56*(3), 339–353. https://doi.org/10.1037/0022-3514.56.3.339
Smith, C. A., & Lazarus, R. S. (1990). Emotion and adaptation. In *Handbook of personality: Theory and research* (pp. 609–637). The Guilford Press.
Spielberger, C. D. (1966). *Anxiety and behavior*. Academic Press.
Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Lawrence Erlbaum Associates.
Stanovich, K. E. (2004). *The robot's rebellion: Finding meaning in the age of Darwin* (pp. xvi, 358). The University of Chicago Press. https://doi.org/10.7208/chicago/9780226771199.001.0001
Stanovich, K. E. (2009). *What Intelligence Tests Miss: The Psychology of Rational Thought* (1er édition). Yale University Press.
Stanovich, K. E. (2011). *Rationality and the reflective mind*. Oxford University Press.
Stanovich, K. E. (2018). Miserliness in human cognition: The interaction of detection, override and mindware. *Thinking & Reasoning*, *24*(4), 423–444. https://doi.org/10.1080/13546783.2018.1459314
Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *The Behavioral and Brain Sciences*, *23*(5), 645–665. https://doi.org/10.1017/S0140525X00003435
Thompson, V. A., & Johnson, S. (2014). Conflict, metacognition, and analytic thinking. *Thinking & Reasoning*, *20*(2), 215–244. https://doi.org/10.1080/13546783.2013.869763
Thompson, V. A., & Morsanyi, K. (2012). Analytic thinking: Do you feel like it? *Mind & Society*, *11*(1), 93–105. https://doi.org/10.1007/s11299-012-0100-6
Thompson, V. A., Pennycook, G., Trippas, D., & Evans, J. S. B. T. (2018). Do smart people have better intuitions? *Journal of Experimental Psychology. General*, *147*(7), 945–961. https://doi.org/10.1037/xge0000457
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
Topolinski, S., & Strack, F. (2015). Corrugator activity confirms immediate negative affect in surprise. *Frontiers in Psychology*, *6*, 134. https://www.frontiersin.org/article/10.3389/fpsyg.2015.00134 https://doi.org/10.3389/fpsyg.2015.00134
Touroutoglou, A., & Efklides, A. (2010). Cognitive Interruption as an Object of Metacognitive Monitoring: Feeling of Difficulty and Surprise. In A. Efklides & P. Misailidi (Éds.), *Trends and prospects in metacognition research* (pp. 171–208). Springer US. https://doi.org/10.1007/978-1-4419-6546-2_9
Trippas, D., Handley, S. J., Verde, M. F., & Morsanyi, K. (2016). Logic brightens my day: Evidence for implicit sensitivity to logical validity. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *42*(9), 1448–1457. https://doi.org/10.1037/xlm0000248
Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science (New York, N.Y.)*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124
Van Steenbergen, H., & Band, G. (2013). Pupil dilation in the Simon task as a marker of conflict processing. *Frontiers in Human Neuroscience*, *7*, 215. https://doi.org/10.3389/fnhum.2013.00215

VanLehn, K., Siler, S., Murray, C., Yamauchi, T., & Baggett, W. B. (2003). Why Do Only Some Events Cause Learning During Human Tutoring? *Cognition and Instruction*, *21*(3), 209–249. https://doi.org/10.1207/S1532690XCI2103_01

Vazard, J. (2021). (Un)reasonable doubt as affective experience: Obsessive–compulsive disorder, epistemic anxiety and the feeling of uncertainty. *Synthese*, *198*(7), 6917–6934. https://doi.org/10.1007/s11229-019-02497-y

Vazard, J. (2022). Everyday anxious doubt. *Synthese*, *200*(3), 224. https://doi.org/10.1007/s11229-022-03711-0

Vogl, E., Pekrun, R., Murayama, K., & Loderer, K. (2020). Surprised–curious–confused: Epistemic emotions and knowledge exploration. *Emotion (Washington, D.C.)*, *20*(4), 625–641. https://doi.org/10.1037/emo0000578

Vogl, E., Pekrun, R., Murayama, K., Loderer, K., & Schubert, S. (2019). Surprise, Curiosity, and Confusion Promote Knowledge Exploration: Evidence for Robust Effects of Epistemic Emotions. *Frontiers in Psychology*, *10*, 2474. https://www.frontiersin.org/article/10.3389/fpsyg.2019.02474 https://doi.org/10.3389/fpsyg.2019.02474

Wason, P. C. (1968). Reasoning about a Rule. *The Quarterly Journal of Experimental Psychology*, *20*(3), 273–281. https://doi.org/10.1080/14640746808400161

Wason, P. C., & Evans, J. S. T. B. T. (1974). Dual processes in reasoning? *Cognition*, *3*(2), 141–154. https://doi.org/10.1016/0010-0277(74)90017-1

Zander, T., Fernandez Cruz, A. L., Winkelmann, M. P., & Volz, K. G. (2017). Scrutinizing the emotional nature of intuitive coherence judgments. *Journal of Behavioral Decision Making*, *30*(3), 693–707. https://doi.org/10.1002/bdm.1982

Zhang, J., Kalinowski, J., Saltuklaroglu, T., & Hudock, D. (2010). Stuttered and fluent speakers' heart rate and skin conductance in response to fluent and stuttered speech. *International Journal of Language & Communication Disorders*, *45*(6), 670–680. https://doi.org/10.3109/13682820903391385

Zsido, A. N., Teleki, S. A., Csokasi, K., Rozsa, S., & Bandi, S. A. (2020). Development of the short version of the spielberger state—Trait anxiety inventory. *Psychiatry Research*, *291*, 113223. https://doi.org/10.1016/j.psychres.2020.113223