Intuition for Me, Deliberation for Thee: Even Intuitive Thinkers Prefer Deliberation in Others on Complex Reasoning Tasks

Matthieu Raoelison[1], Wim De Neys[1]

November 10, 2026

[1]Université de Paris, LaPsyDÉ, CNRS, F-75005 Paris, France

Abstract

Although people often celebrate the power of intuitive thinking, recent work suggests that, when evaluating others solving reasoning problems, we prefer deliberation over intuition even when accuracy is held constant. The present study tested whether this "deliberation preference" varies with individual differences in participants' own thinking tendencies. Participants rated vignette-described reasoners who relied on intuition or deliberation and who were highly accurate, poorly accurate, or of unspecified accuracy. Replicating prior findings, deliberative reasoners were evaluated more positively than intuitive reasoners across all accuracy levels. Individual differences reliably predicted evaluations: compared to participants with more intuitive tendencies (composite of cognitive capacity and thinking-style measures), participants with more deliberative dispositions (composite of cognitive capacity and thinking-style measures) rated intuitive profiles less favorably across accuracy conditions, whereas evaluations of deliberative profiles depended on performance (higher for accurate deliberators and lower for inaccurate deliberators). Crucially, even participants with the strongest intuitive tendencies still showed an absolute preference for deliberation in others. These results indicate that folk evaluations of reasoning exhibit a robust pro-deliberation bias on complex reasoning tasks and suggest an asymmetry whereby one's own intuitive style primarily shifts attitudes toward intuition rather than toward deliberation.

Keywords: Dual-Process Theory, Individual Differences, Folk Theory

1 Introduction

In his prime, Roger Federer was lauded for his seemingly effortless play on the court, the intuitive grace he displayed. He would later recount how frustrating this was, as such comments overlooked the years of rigorous, deliberate preparation required to reach this level. While the illusion of effortlessness wooed the media and the public, his peers (with more at stake when competing) recognized the hard work that supported his success.

Oftentimes, hard work is shunned while (supposedly) innate talent gets all the attention in popular sports. Likewise, intuition often gets praised in popular culture for its effortlessness, whereas academic findings usually highlight effortful deliberation as necessary to overcome human shortcomings. Indeed, although we pride ourselves on our ability to reason, decades of research have shown that people often fall prey to a number of biases (Evans & Stanovich, 2013; Frederick, 2005; Kahneman, 2011), which can have real-life consequences (Eberhardt et al., 2006; Oster & Koesterich, 2013; Schmidt et al., 2017; W. C. Thompson & Schumann, 1987). These humbling findings have made their way into popular culture as well, with attention brought to intuition and deliberation (also known as System 1 and System 2, or Type 1 and Type 2 processes, e.g., Kahneman, 2011; Stanovich & Toplak, 2012) and their interaction in fast-and-slow thinking. Given that opposite views favoring either intuition (e.g., de Becker, 1997; Gigerenzer, 2007) or deliberation (e.g., Brafman & Brafman, 2008; Chabris & Simons, 2010; Legault, 2006; World Bank, 2015) coexist in the public sphere, it remains unclear how people truly feel about these reasoning processes, and to what extent their own reasoning behavior influences their perception. To better understand how people perceive and value intuition and deliberation, we need to build a "folk theory" of fast-and-slow thinking (Bonnefon & Rahwan, 2020).

Early inquiries suggest that several factors might play a role in determining which reasoning process is preferred (Inbar et al., 2010). For instance, in contexts where authenticity matters, like when looking for a romantic partner, we value intuition more (Oktar & Lombrozo, 2022). Conversely, we often shift our preference to deliberation for cognitively challenging tasks (Kupor et al., 2014), even from a young age (Richardson & Keil, 2022).

It may be tempting to attribute this preference for deliberation to a preference for accuracy. Indeed, earlier reasoning research made deliberation a requirement for accuracy. However, recent investigations suggest that accurate reasoners can rely on their intuition too (Bago & De Neys, 2017; De Neys, 2023; Reyna, 2004, 2012). That is, they don't always need deliberation to correct their intuitive responses because those responses are already correct. Furthermore, it appears that higher cognitive capacity is linked to higher-quality intuitions (Peters, 2012; Raoelison et al., 2020; V. A. Thompson et al., 2018). This could explain why so many experts are lauded for their intuition. Still, if intuitive thinkers can be as good as deliberative ones, the opposite can be true as well. Indeed, when intuitions are misleading, deliberation often fails to correct them (Bago & De Neys, 2017; Raoelison et al., 2020; V. A. Thompson et al., 2018). Together, these findings suggest that there are grounds to disentangle the reasoning process that leads to a response from its accuracy.

A recent investigation totaling 13 studies found an overall preference for deliberation in the context of solving reasoning problems, regardless of accuracy (De Neys & Raoelison, 2025). That is, when

presented with several decision makers with varied reasoning modes and performance levels, participants on average rated decision makers who relied on deliberation higher than those who relied on intuition instead, across all levels of performance. However, although the study identified a strong modal preference for deliberation, it did not examine possible individual differences.

In the current study we used the same vignettes as De Neys and Raoelison (2025) and measured individual differences to investigate whether various indicators of people's own reasoning style predict their perception of intuition and deliberation. Bluntly put, do people who themselves engage more in intuitive thinking do tend to show preference for intuition over deliberation in others?

2 Methods

2.1 Open Science and Data

The research question and study design were preregistered on AsPredicted (https://aspredicted.org/bk2ah2.pdf). No specific analyses were preregistered. All data, material and analysis scripts can be retrieved from our OSF page: https://osf.io/5pqvy/.

2.2 Participants

We recruited 241 participants (120 male, 119 female, 2 other/undisclosed; mean age = 40.4, SD = 13.9, range = 19 to 81) via Prolific (https://www.prolific.com). This sample size allowed us to detect medium effects (.23) between intuitive and deliberate vignettes with 95% power. Participants were compensated at a rate of £6.00 per hour. They had English as their first language and held UK, USA, Ireland, Australia, Canada, or New Zealand nationality.

2.3 Material

2.3.1 Vignettes

Participants were presented with seven profiles, each describing a hypothetical decision maker's reasoning mode (intuition or deliberation) and accuracy (high, low, or unspecified) when engaged in complex reasoning. The following examples illustrate an intuitive-high accuracy profile and a deliberative-low accuracy one:

> *"Person C follows their intuition when reasoning about a problem. They do not spend much time or effort to arrive at a conclusion. The accuracy of their answers is very high."*
>
> *"Person A reflects deeply when reasoning about a problem. They spend a lot of time and effort to arrive at a conclusion. The accuracy of their answers is very low."*

In the unspecified accuracy condition, the last sentence with the accuracy information was omitted. Each of the vignettes was labelled with a unique letter and specified a unique combination of the reasoning mode and accuracy level. In addition to the six experimental vignettes we also presented an extra filler

vignette with the text "There is no information about this person's profile". Table 1 presents the content of all vignettes.

Table 1
Vignette texts.

Reasoning mode	Accuracy	Description
Intuition	High	Person A follows their intuition when reasoning about a problem. They do not spend much time or effort to arrive at a conclusion. The accuracy of their answers is very high.
Deliberation	High	Person B reflects deeply when reasoning about a problem. They spend a lot of time and effort to arrive at a conclusion. The accuracy of their answers is very high.
Intuition	Low	Person C follows their intuition when reasoning about a problem. They do not spend much time or effort to arrive at a conclusion. The accuracy of their answers is very low.
Deliberation	Low	Person D reflects deeply when reasoning about a problem. They spend a lot of time and effort to arrive at a conclusion. The accuracy of their answers is very low.
Intuition	Unspecified	Person E follows their intuition when reasoning about a problem. They do not spend much time or effort to arrive at a conclusion.
Deliberation	Unspecified	Person F reflects deeply when reasoning about a problem. They spend a lot of time and effort to arrive at a conclusion.
Control		Person O. There is no information about Person O's profile.
Practice		Person Z likes puzzles. They also spend time solving crosswords.

The letters for each Person were randomized for each participant.

Rating scales For each vignette, participants had to rate the perceived reasoning quality of the profile described by clicking on the corresponding value on an 11-point rating scale (from 0 = "Not very good" to 10 = "Very good"). Participants were shown examples of the rating scale in the instructions and were given one practice trial to familiarize themselves with it (see Supplementary Material for full instructions). The presentation order of the vignettes was randomized for each participant with both unspecified vignettes being presented together first or last.

Ranking question After participants had rated the four profiles where accuracy information was provided (i.e., the "core" vignettes), they were asked to directly rank-order them by perceived intelligence. We opted for the label "intelligence" to avoid repeating the rating scale label. Participants saw the four vignettes all at once and clicked on a rank order for each one of them with the following instructions:

"From the descriptions below, please now rank the people from the most intelligent (number 1) to the least intelligent (number 4). You can only place one person at each rank (i.e., no ties)".

The rankings allowed us to classify participants as preferring either intuition or deliberation.

Betting question Once all profiles had been rated and the ranking question had been completed, participants were given the following instructions:

"After this study is completed we will have four individuals with profiles matching the ones you've seen solve a reasoning problem. You will get to see the specific profiles in question again below and can bet on one of them. If the person you select solves the problem correctly, you will get a bonus payment of 1 pound. Click on the profile you'd like to bet on, then click on Next to validate."

Participants then saw the four profiles they previously rank-ordered and had to select one of them. This allowed us to see which profile participants would deem as most likely to solve a future reasoning problem, thereby maximizing their chance to earn a bonus. Note that our design used deception. All participants received the bonus payment regardless of their choice which was not announced to participants. Hence, the bonus was not contingent on their choice but was only used to frame this task as more consequential than the rating task.

2.3.2 Individual differences measures

To measure participants' tendency to rely on intuition or deliberation, we used adapted versions of commonly used scales from the literature. The different scales we used can be broadly categorized as assessing cognitive capacity (Raven and CRT), prescriptive cognitive style (AOT) and descriptive thinking dispositions (NFC, USDI and self-report items). There was no a priori hypothesis regarding how each category would contribute to evaluations.

Raven We used a 12-item version of the Raven Advanced Progressive Matrices (Bors & Stokes, 1998) as a measure of cognitive capacity (i.e., fluid intelligence). As deliberation is computationally expensive, this serves as a proxy for the capacity to engage in deliberation. Participants had to select the correct pattern out of eight that would complete a 3-by-3 matrix. Matrices were presented separately with increased difficulty, one per page.

CRT Cognitive reflection refers to the ability to override an intuitive but erroneous response and has usually been measured by the Cognitive Reflection Test (CRT, Frederick, 2005)[1]. Following Pennycook et al. (2023), we used the original 3 items of the CRT and 3 items from a non-numerical version (CRT-2, Thomson & Oppenheimer, 2016). CRT items used a free response format and were presented separately, one per page.

[1]Following one reviewer's suggestion, we added analyses using these erroneous intuitive responses (aka lured responses) for Raven and CRT in the Appendix.

AOT Actively Open-minded Thinking (AOT) refers to a standard of thinking to avoid relying on intuitive responses and look for alternatives instead (Baron, 1991; Stanovich & West, 1997). We used a 13-item AOT developped by Stanovich and Toplak (2023). Participants had to use a 6-point scale (from 1 = "Disagree strongly" to 6 = "Agree strongly") to indicate their agreement with statements such as "People should always take into consideration evidence that goes against their opinions." All thirteen items were presented on a single page[2].

NFC The need for cognition (NFC, Cacioppo & Petty, 1982) refers to the tendency to engage in and enjoy thinking for its own sake. We used a 6-item version developed by Coelho et al. (2020). Participants had to use a 5-point scale (from 1 = "Extremely uncharacteristic of me" to 5 = "Extremely characteristic of me") to indicate their agreement with statements such as "I would prefer complex to simple problems." All six items were presented on a single page.

USID To assess the propensity to rely on intuition and deliberation, we used the 32-item Unified Scale to Assess Individual Differences in Intuition and Deliberation (USID, Pachur & Spaar, 2015). Participants had to use a 5-point scale (from 1 = "I don't agree" to 5 = "I agree completely") to indicate their agreement with statements such as "I quickly do the right thing when deciding because I've often faced almost the same thing before." All thirty-two items were presented on a single page, in a predetermined pseudo-random order which was the same for every participant.

Self-report scales In addition, we used a short 3-item questionnaire to assess the relative and absolute extent to which participants in general relied on intuition and deliberation. The first item was a simple binary choice asking which reasoning mode they usually engage in, followed by two 11-point rating scales (from 0 = "Not at all" to 10 = "All the time") to assess to what extent participants relied on intuition and deliberation when solving problems. Participants were reminded above the first rating scale that the two ratings were independent and didn't need to add up to 10. All three items of this self-report questionnaire were presented on a single page.

Reflection-Impulsivity Reflection-Impulsivity (RI, Baron et al., 2015) refers to a cognitive style concerned with the relative preference for accuracy versus speed, that is, the extensiveness of thinking. In addition to cognitive capacity, CRT, as well as Raven matrices, can be used as a measure of RI (Baron et al., 2015). RI was computed as z(accuracy) + z(log(RT)) for each participant. Higher values indicate thoroughness and deliberation as individuals spend more time to prioritize accuracy over speed, while lower values indicate impulsivity where individuals prioritize speed over accuracy. Note that our planned analyses used accuracy for CRT and Raven. As suggested during the review, we replaced it with RI measures. Results pointed to similar conclusions. For completeness, tables with accuracy as well as log(RT) can be found in the appendix.

[2]Note that, representing a standard of thinking that people should strive toward, AOT is a prescriptive scale, while NFC, USID and our self-report questionnaire were simply descriptive scales.

2.3.3 Procedure

After providing their consent, participants were presented with the rating task. After a practice example, they proceeded to rate the seven vignettes which were presented in a counterbalanced order such that each combination of the reasoning mode and accuracy level would be presented first an equal number of times. Participants completed the ranking task after the four core vignettes were rated. The betting question was then presented. The individual differences measures were presented after the betting question in the following order: the self-report questionnaire, Raven matrices, the AOT, the NFC, the CRT and the USID. Apart from the self-report questionnaire, the introduction of each measure indicated the number of items to complete. Last, they were presented with demographics questions before completing the experiment.

2.4 Statistical analysis

All analyses and visualizations were done in R using the following packages (in alphabetical order): *afex* (Singmann et al., 2024), *broom* (Robinson et al., 2024), *emmeans* (Lenth, 2024), *ggpubr* (Kassambara, 2023a), *ggrepel* (Slowikowski, 2024), *gt* (Iannone et al., 2024), *ggsignif* (Ahlmann-Eltze & Patil, 2021), *here* (Müller, 2020), *janitor* (Firke, 2023), *knitr* (Xie, 2024), *patchwork* (Pedersen, 2024), *plotly* (Sievert, 2020), *rstatix* (Kassambara, 2023b), *tidyverse* (Wickham et al., 2019). Cohens' d for contrasts was computed using JASP (JASP Team, 2025).

3 Results and discussion

3.1 Average ratings

Figure 1 illustrates the ratings for the different profiles. We ran a 3 (Accuracy Information: High, Low, Unspecified) x 2 (Reasoning Mode: Intuition, Deliberation) within-subject ANOVA on the ratings using the *afex* package in R, treating System and Accuracy as factors, using orthogonal sum-to-zero contrasts. In practice, System (Intuition vs Deliberation) was centered using orthogonal contrast weights (-0.5,0.5), and Accuracy was modeled as a 3-level ordered factor centered at zero (Low=-1, Unspecified=0, High=1).

The ANOVA revealed a main effect of Accuracy information ($F(1.70, 408.91) = 1025.17$, $p < .001$, $\eta_p^2 = .81$). In line with previous findings, participants preferred reasoners with high accuracy over those with low accuracy, with the unspecified condition in between, which indicates that the accuracy manipulation worked as intended.

More critically, there was a main effect of Reasoning mode ($F(1, 240) = 294.90$, $p < .001$, $\eta_p^2 = .55$) [3]. Holm-corrected t tests confirmed the visual impression from Figure 1, indicating a general preference for the deliberative over the intuitive reasoner that was strongest when accuracy information was absent and slightly attenuated in the high- and low-accuracy conditions (High: diff = 2.6, $t(240) = 14.59$, $p < .001$,

[3] We also found a significant Reasoning mode x Accuracy interaction ($F(1.89, 452.74) = 41.02$, $p < .001$, $\eta_p^2 = .15$). However, we couldn't interpret it since this interaction is removable (Wagenmakers et al., 2012) and because of possible ceiling or floor effects, which was raised during the review.

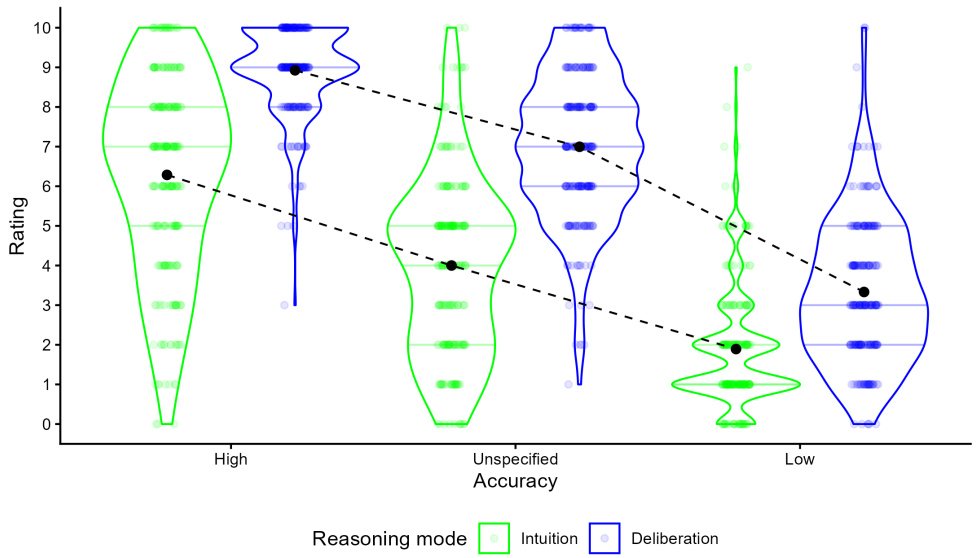


Figure 1. Perceived Reasoning Quality as a Function of Reasoning mode and Accuracy. Violin plots showing the perceived reasoning quality for all profiles. Black dots and dotted lines indicate the mean scores.

$d = 1.40$; Unspecified: diff = 3, $t(240) = 14.61$, $p < .001$, $d = 1.59$; Low: diff = 1.4, $t(240) = 12.08$, $p < .001$, $d = 0.76$). Importantly, the preference for deliberation never disappeared or reversed and remained significant across all accuracy conditions, as indicated by these comparisons. These findings thus indicate a general preference for deliberation over intuition, regardless of accuracy.

To check for potential bias from the within-subject design, we also ran a between-subjects analysis on the first item each participant evaluated, yielding similar results (see Appendix A for details). For further validation, after participants had rated all vignettes, they were also asked to directly rank-order four core vignettes (i.e., high/low accuracy intuitor and deliberator) by perceived reasoning quality. In addition, average ranking scores followed the ratings, with deliberative reasoners ranked higher than intuitive ones and accurate reasoners ranked higher than inaccurate ones (see Appendix B for details), giving further credence to a robust preference for deliberation.

Following De Neys and Raoelison (2025), we tested whether the observed deliberation preference would still apply in a more consequential framing. Our betting paradigm asked participants to bet on the performance of either a highly accurate intuitor or deliberator (as well as their lowly accurate counterparts for control purposes). Results were in line with the rating and ranking findings. Participants were far more likely to bet on the accurate deliberator (81.7%) than on the accurate intuitor (17.4%; $\chi^2(1) = 100.52$, $p < .001$). Thus, when choices had real monetary consequences, the preference for deliberation over intuition (here interpreted as which process is most likely to lead to a bonus payment) became even more pronounced (see Appendix C for further details).

3.2 Individual differences

Taken together, at the aggregate level, the present results replicated the results of De Neys and Raoelison (2025). We now turn to the critical analysis of our individual difference predictors.

3.2.1 Predictors

The predictors showed moderate-to-good reliability (Cronbach's alpha for Raven: .68; CRT: .77; AOT: .84; NFC: .85; USID-Deliberation: .90; USID-Intuition: .92). Table 2 presents the correlations between the different predictors. Results suggest that the self-report scales tracked their intended reasoning mode as we found overall moderate positive correlations across indicators reflecting a preference for intuition (i.e., the intuition subscale of USID and our self-report scale for intuition) or deliberation (i.e., NFC, the deliberation subscale of USID, Raven, CRT, AOT and our self-report scale for deliberation), and moderate negative correlations between contrasting indicators.

Table 2
Correlations between Individual Differences Predictors.

Predictor	1	2	3	4	5	6	7	8	9
1. USID-Intuition	-								
2. Self-Intuition	**.50*** **	-							
3. NFC	.01	-.01	-						
4. USID-Deliberation	.08	-.01	**.60*** **	-					
5. RI_Raven	**-.28*** **	**-.16***	**.21****	**.19****	-				
6. RI_CRT	.02	-.08	**.17*** **	**.21****	**.45*** **	-			
7. AOT	**-.58*** **	**-.38*** **	**.15***	.02	**.18****	.03	-		
8. Self-Binary[a]	**-.48*** **	**-.63*** **	.08	.09	**.24*** **	**.13***	**.34*** **	-	
9. Self-Deliberation	**-.41*** **	**-.4*** **	**.25*** **	**.33*** **	**.30*** **	.12	**.39*** **	**.63*** **	-

[a] 0 = intuition, 1 = deliberation
*p <.05. **p <.01. ***p <.001

3.2.2 Rating correlations

Composite score First, given the overall coherent pattern, to summarize the effect of predictors, we took the Z scores and took their average (after reversing those for the intuitive subscale of USID and our self-report scale on intuition) to create a composite score combining cognitive capacity, cognitive style and thinking dispositions and thus reflecting participants' overall thinking tendencies. Higher scores indicate a higher likelihood an individual prefers to engage in deliberation when reasoning themselves.

Figure 2 shows a scatterplot of the ratings faceted by accuracy level and reasoning mode, in addition to the correlations with the composite score. The key insight is that scores for intuitive profiles were negatively correlated with the composite score at all accuracy levels: the more one personally relies on deliberation, the lower one evaluates others' intuition. A different pattern emerges for deliberative profiles: correlations were positive for high accuracy, not significant when accuracy was unspecified, and

negative for low accuracy. In other words, the more one relies on deliberation, the more one values accurate deliberators but the more critically one evaluates inaccurate ones.

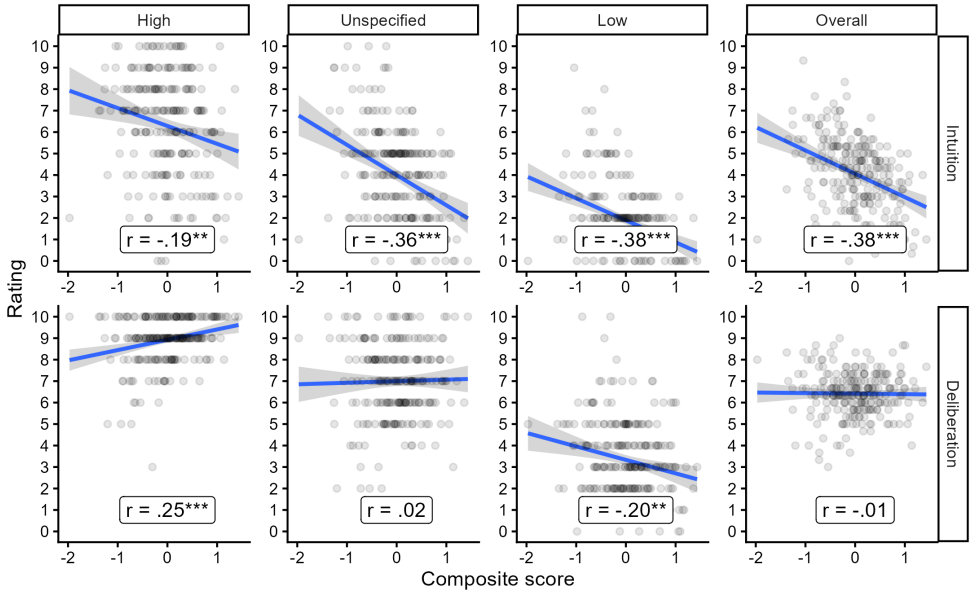


Figure 2. Correlations Between Participants' own Thinking Tendencies (Composite Score) and Profile Ratings. Scatterplots faceted by Accuracy and Reasoning Mode with Pearson correlations. Blue lines represent linear regressions.

Hence, overall the Figure 2 results indicate that the higher one's tendency to rely on intuition (vs deliberation) oneself, the higher one tends to rate intuition in others. However, this leaves open the question of whether more intuitive reasoners still show an absolute preference for deliberation over intuition. To gauge this issue, we also calculated a difference score between one's rating of the deliberate and intuitive profiles (i.e., a score above 0 indicates a preference for deliberation) and correlated this with the composite predictor. Results are shown in Figure 3, they replicate the association between one's own tendency to rely on intuition and the preference for intuition in others. However, as the figure shows, the difference score was predominantly above 0 (average = 2.4, $t(240) = 17.2$, $p < .001$), even among those with negative composite scores (i.e., the more intuitive thinkers, average = 1.9, $t(120) = 9.7$, $p < .001$). This indicates that even the most avid intuitive reasoners typically maintain a preference for deliberation over intuition in others. Figure 4 further illustrate this by plotting the overall preference for deliberation and intuition in the top and bottom deciles of our composite predictor score. As the figure indicates, people with a stronger preference for intuition when they reason themselves value intuition in others more than those showing a stronger preference for deliberation themselves (and vice-versa for deliberation scores), but both groups show the same overall pattern: a general preference for deliberation over intuition across all accuracy levels. Further details can be found in Appendix D.

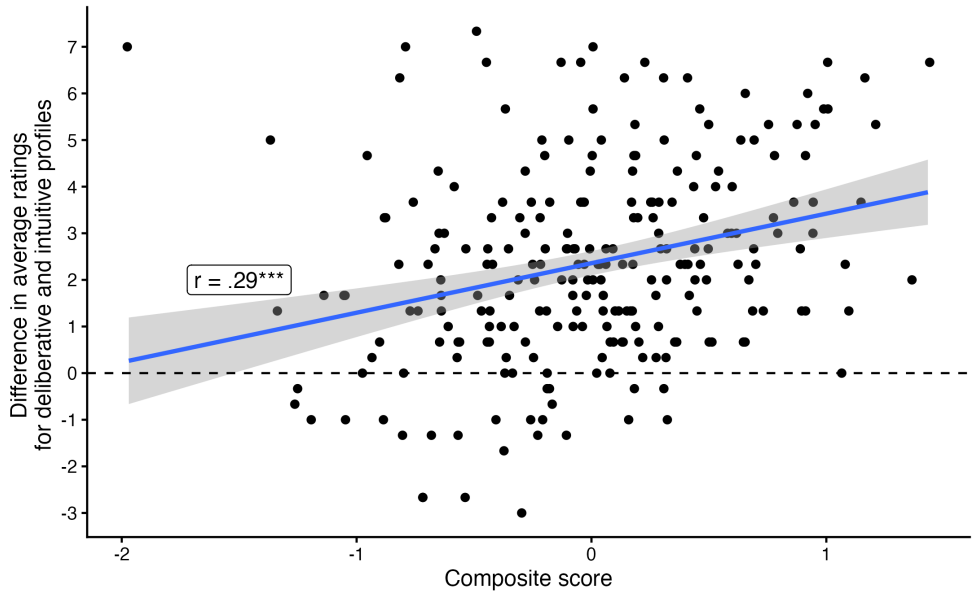


Figure 3. Difference in Perceived Quality Between Deliberative and Intuitive Thinkers, Correlated with Composite Score. Scatterplot showing the difference score (average deliberation rating minus average intuitive rating) for all participants relative to their composite scores with Pearson correlation.

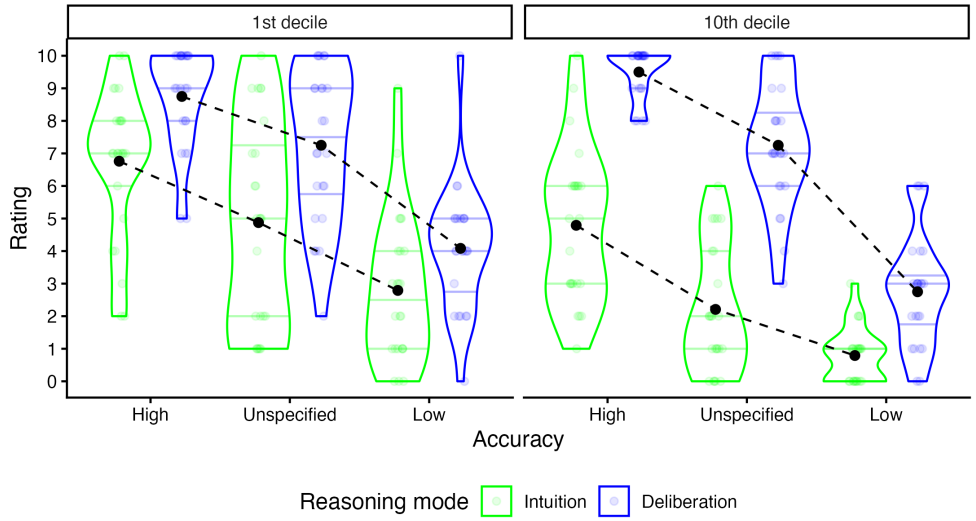


Figure 4. Comparative Ratings for the Most Intuitive vs Most Deliberative Participants. Violin plots showing the perceived reasoning quality, faceted by the bottom (1^{st}: most intuitive) and top (10^{th}: most deliberate) deciles of the composite score.

Table 3
Correlations between Predictors and Perceived Quality separated by Accuracy and Reasoning mode.

Predictor	Reasoning mode	Accuracy			
		High	No	Low	Overall
RI_Raven	Intuition	.01	-.11	**-.24*** **	-.12
	Deliberation	.11	-.01	**-.21** **	-.08
AOT	Intuition	**-.30*** **	**-.45*** **	**-.48*** **	**-.50*** **
	Deliberation	**.25*** **	**.17** **	**-.21** **	.06
NFC	Intuition	.04	-.02	-.02	.01
	Deliberation	.06	-.12	-.04	-.07
RI_CRT	Intuition	.12	-.01	-.07	.03
	Deliberation	.08	-.08	-.08	-.06
USID-Intuition	Intuition	**.33*** **	**.42*** **	**.41*** **	**.48*** **
	Deliberation	**-.20** **	-.09	**.16***	-.03
USID-Deliberation	Intuition	**.16***	.07	.07	**.13***
	Deliberation	.12	.02	-.01	.05
Self-Intuition	Intuition	**.21** **	**.33*** **	**.26*** **	**.33*** **
	Deliberation	-.08	-.01	.10	.02
Self-Deliberation	Intuition	-.12	**-.25*** **	**-.21** **	**-.24*** **
	Deliberation	**.14***	.06	**-.13***	.01
Self-Binary[a]	Intuition	**-.19** **	**-.29*** **	**-.25*** **	**-.30*** **
	Deliberation	**.17** **	.06	-.11	.03
RI_Composite	Intuition	-.11	**-.31*** **	**-.35*** **	**-.30*** **
	Deliberation	**.22*** **	.02	**-.22** **	-.04

[a] 0 = intuition, 1 = deliberation
*p <.05. **p <.01. ***p <.001

Separate predictors Table 3 reports the correlations with the different profiles separated by reasoning mode and accuracy for all predictors. By and large, the different measures showed the same pattern: participants' own reasoning mode tendency predicted their ratings of others' reasoning mode. The more participants reasoned intuitively themselves, the higher their rating for intuitive reasoning in others. The more participant reasoned deliberately themselves, the higher their rating for accurate deliberate reasoning and the lower their rating for inaccurate deliberate reasoning. The AOT, the intuition subscale of USID and the self-reports showed the strongest effects on ratings for intuitive profiles. The Raven matrices, the CRT and the deliberation subscale of USID had somewhat weaker effects, while the NFC had no effect. The AOT and the intuition subscale of USID had the strongest effects on ratings for deliberative profiles. Raven matrices, our self-report for deliberation and the CRT had weaker effects as well, with the NFC still showing no significant effect.

3.2.3 Rankings and bets

For completeness, note that exploratory analyses showed that our composite score also predicted the preference for intuition or deliberation as seen in the rankings and bets. T tests showed that the composite score was lower, but not significantly so, for those who ranked intuition higher than deliberation ($t(118.4)$ = -1.9, p = .05, d = -0.3) as well as significantly lower for those who bet on intuition instead of deliberation ($t(75.77)$ = -3.6, p = .001, d = -0.6). Again, the separate predictors broadly followed the same trends as the composite score regarding the rankings and the bets: the preference for a given reasoning mode in others was predicted by the tendency to engage in that reasoning mode oneself. See Appendix E for full results.

4 General discussion

We investigated whether a person's own reasoning style shapes how they evaluate the thinking of others on complex reasoning tasks. As in previous work, participants showed a general preference for deliberation over intuition but the magnitude of this preference was smaller for participants who personally exhibited a more intuitive reasoning style. Nevertheless, even individuals with the strongest tendency to rely on intuition when making decisions themselves typically still preferred deliberation over intuition in others. Bluntly put, no matter how strongly we tend to rely on our own intuition, we remain skeptical of others' intuitions (i.e., prefer them to use deliberation over intuition). Taken together, this indicates that the previously observed general "deliberation" preference is remarkably robust.

Interestingly, it seems that our own reasoning tendencies don't influence our perceptions of intuition and deliberation in others equally. More precisely, it appears that the degree to which one leans toward intuition or deliberation primarily impacts one's attitude toward intuition, as the predictors did not significantly shift the perception of deliberation overall (see Figure 2 and Table 3). Our predictors were negatively correlated with the perception of intuition at all accuracy levels, showing a consistent effect. However, the predictors were positively correlated with the perception of deliberation for highly accurate profiles but negatively correlated for inaccurate profiles. In other words, while both intuitive and deliberative thinkers value deliberation equally overall, intuitive thinkers show a higher valuation for intuition in others than deliberative thinkers do. As one reviewer noted, this asymmetry could also be due to different baselines. The reasoning literature routinely praises deliberation over intuition (e.g., Frederick, 2005; Kahneman, 2011). This would lead participants to value deliberation more, absent any accuracy information. This information, in turn, would influence the relationship between individual differences and evaluations. Indeed, as accuracy decreased, so did the correlations (see Figure 2). Had baseline evaluations been closer (i.e., when no accuracy information is provided), we might have observed increasingly negative correlations for both intuition and deliberation, or conversely, positive correlations for highly accurate intuitors similarly to deliberators. One additional interpretation compatible with the asymmetry lies in more deliberate rea-soners who may see (idealized versions of) themselves in other highly accurate deliberators but not the inaccurate ones, thus giving higher evaluations to the former but lower evaluations to the latter. Future research efforts may investigate which is more likely.

While our results suggest that the overall preference for deliberation is robust, one may wonder why intuitors would not trust others' intuition. One tentative answer for this seemingly counterintuitive finding lies in deliberation's role in providing justifications (De Neys, 2025; Evans, 2019). This requires cognitive transparency (i.e., its output comes with an awareness of how it was derived, Bonnefon, 2018), a feature that intuition lacks. Consequently, we are often unable to justify our intuitions, which is one of the reasons they are often referred to as "gut feelings" (Bago & De Neys, 2019b). Indeed, as has been observed across several tasks and domains, people are frequently unable to provide a sound justification for responses generated intuitively but they are able to do so more consistently when their responses are generated through deliberation (e.g., Bago & De Neys, 2019a, 2019b; Beauvais et al., 2025). Why does justification give an edge to deliberation over intuition? As others' intuitions are not transparent to us (i.e., we can't "feel" their intuitions), this may explain why even a highly accurate intuitor, however good outcomes his intuition may lead to, will still prefer others rely on deliberation. While one's own intuition will typically feel "right" (i.e., is accompanied by an affective feeling of rightness, V. A. Thompson et al., 2011), this feeling may be enough for themselves to rely on intuition without fully grasping how it operates (Bonnefon, 2018). However, given we can't feel others' feelings, we may still prefer others to deliberate so they are able to explain and justify their reasoning.

In addition, one extra factor may be the (opportunity) cost of deliberation. Given that deliberation is slow and burdens our cognitive resources, our own deliberation presents a high processing cost. However, when we ask others for advice, we do not have to (directly) invest time and cognitive resources ourselves. Although speculative, this "outsourcing" of the cost of deliberation may be an extra factor that explains the greater preference for deliberation in others.

More specific insights can also be gleaned from a more in-depth look at the correlations of our different predictors. For instance, according to Baron (1991), "good thinking involves optimal search for possibilities, evidence, and goals, and fairness in the search for evidence and in inference". Well-justified arguments (from someone else) may thus be more likely to be evaluated positively by someone scoring high on the AOT, which is what we observed in our data. Furthermore, AOT had the strongest correlations with ratings among all our predictors, as opposed to NFC, which had no significant correlation with the ratings. Since NFC relates to the enjoyment of deep thinking per se, it may be considered to be focused on one's own reasoning, while AOT also relates to engaging with other's thinking (Baron, 2024), which may explain this discrepancy. Relatedly, one could argue that both scales are concerned with different objects: as a prescriptive scale, the AOT represent a standard of thinking that people should strive toward. That is, participants might answer it while envisioning an idealized version of themselves (i.e., more distant), while the other thinking disposition scales might only remind them of past or current version of themselves (i.e., closer). Envisioning a more distant version of oneself may thus relate more closely to evaluating other reasoners than more self-centered scales such as NFC. Alternatively, as was suggested during the review, an individual scoring high on NFC might spend time bolstering his initial, erroneous conclusions instead of looking for an alternative. Such bolstering or rationalization is a form of reflection, but this type of high NFC individuals may therefore not believe that other people's reflection is advantageous (i.e. as in leading to a different, "better" judgment). This variability could

contribute to the NFC's decreased predictive power in the current context. Additionally, note that we used a shortened, 6-item version of NFC, which may have further weakened its relation to behavioral indicators. Finally, uncertainty aversion may also play a role in devaluing intuition. As suggested earlier, the lack of justifications from intuition may make it less appealing to those scoring higher on the AOT.

Perhaps more surprisingly, CRT scores didn't correlate (negatively) with measures of intuitive thinking dispositions, although they usually track the two key features of reflective reasoning (Byrd, 2025) — i.e., deliberate reconsideration of one's initial response and conscious reasoning (Byrd et al., 2023). An analysis using lured responses (e.g., Byrd & Białek, 2021; Byrd & Conway, 2019; Byrd et al., 2023, see Appendix) also didn't find any association with intuitive thinking dispositions (see also Pennycook et al., 2015). Previous works also suggest that thinking dispositions and (assumedly) related behavioral performance don't always correlate (e.g., Coutinho et al., 2021 for intuition and CRT; Gärtner et al., 2021 for NFC and executive functions). Solving CRT items is often considered to require overriding the intuitively-cued response with a correct reflective response (Frederick, 2005), in case a subject didn't intuitively produce a correct response that doesn't require correction (e.g., Bago & De Neys, 2017, 2019b). Both disposition and cognitive resources are necessary to engage in fruitful deliberation while it could be argued that an intuitive thinking disposition is sufficient to engage in intuition. This asymmetry might explain why CRT scores didn't correlate with USID-Intuition or Self-Intuition. Another potential explanation is that participants lacked self-knowledge, in that they would describe themselves as more intuitive than they really are. However, both correct and lured responses on the Raven task did correlate with intuitive thinking dispositions (see in Appendix), which makes it less likely and may hint at other features that CRT doesn't share with Raven matrices (e.g., it has a math component, Meyer et al., 2024).

Relatedly, the Raven test we used is a progressive test, where items are supposedly ordered by difficulty level, which wasn't the case for the AOT or the NFC for instance. This could explain why its reliability was only moderate (i.e., $\alpha = .68$). While a low reliability based on raw scores might attenuate correlations, we used RI measures instead following the review, which should alleviate this issue. Note that this task produced similar reliability in a lab setting with 260 participants (Study 2, Raoelison et al., 2020).

Note that our self-report scales were able to track their respective relative reasoning mode. Correlations with the AOT, the intuitive subscale of the USID and the ratings for intuitive reasoners suggest that our self-report scale for intuition was an appropriate measure of actual intuition use. Similarly, correlations with the AOT, Raven, the reflexive subscale of USID and the ratings for deliberative reasoners suggest that our self-report scale for deliberation was an appropriate measure of actual deliberation use. Together with the binary question, our self-report scales could thus allow a quick, if rather crude, assessment of thinking tendencies toward intuition and deliberation.

Regarding the betting task, one may wonder why, contrary to the modal pattern, a few participants (18%) didn't consider the deliberative profile the most likely to solve a future reasoning problem. While correlations suggest that those were mostly participants less inclined toward deliberation, even some of the participants who ranked deliberation higher than intuition bet on intuition. One potential explanation lies in good deliberators who dislike inaccurate deliberators. They may not believe others have the necessary capacities to deliberate well and therefore want them to intuit. Asking participants to justify their choice

may provide better understanding in future investigations.

The present study and findings also faced some limitations. First, our study presented participants with a constrained, hypothetical situation: reasoners involved in solving challenging problems. As noted earlier, the literature in reasoning has routinely shunned intuition and instead emphasized the role of deliberation in sound reasoning (e.g., Frederick, 2005; Kahneman, 2011), which suggests that a problem-solving context would heavily favor deliberation. Past research has shown that, in contexts such as looking for a romantic partner or a restaurant, intuition and deliberation are valued (Oktar & Lombrozo, 2022) and relied on differently (Inbar et al., 2010; Pachur & Spaar, 2015). Engaging in deliberation to answer trivial questions is also likely to be shunned (Kupor et al., 2014). If our participants were to evaluate decision makers in such contexts, rather than in problem solving, it is thus possible that our primary pattern (i.e., that deliberation is favored over intuition) would not hold. Whether we would then observe similar correlations with thinking tendencies to the current study remains to be tested. Note that Baron (in Study 3, Baron, 2024) previously looked at the generality of AOT across different topics or domains. He found positive correlations across topics, although there was considerable topic-related variability. We therefore remain cautious about making too broad a generalization regarding the deliberation preference and call for more research into how the decision domain or context may alter it.

Second, one may consider that our participants were not evaluating intuition and deliberation as isolated labels but were evaluating bundled reasoning-and-effort cues instead. Deliberation has many properties and functions (De Neys, 2023, 2025). Among them, time and effort are frequently tied to deliberation in the literature (Evans & Stanovich, 2013). While they are not defining features, evidence suggests they are cues that people rely on to infer the use of thoughtful deliberation (e.g., Kupor et al., 2014). It thus seems reasonable that profiles that evoke either reasoning process could act as their proxies. Nevertheless, stronger precise claims about a folk preference for deliberation per se would ideally require a design that disentangles all the various possible deliberation features and functions (e.g., isolate process from effort, speed, thoroughness, and the capacity to explain one's answer, for example). Future work may examine the role of each individual feature in more detail.

Third, our approach combined several measures (i.e., rating, ranking and betting tasks) to ensure the robustness of our findings, but note that we didn't imply any construct equivalence. For instance, we had previously used 3 separate scales for evaluating the different profiles (De Neys & Raoelison, 2025) instead of a single scale. The different components evaluated (i.e., reasoning quality, smartness and trust in each profile) were not equivalent but nonetheless provided consistent and correlated ratings, which led us to use a single evaluation scale. Similarly, the different measures we used in this study give credence to a preference toward deliberation but don't necessarily assess a unique, shared dimension of this preference. Future research may focus on each measure in more detail.

Fourth, while it allows us to look at accuracy and reasoning mode as separate factors, our vignette-based design doesn't capture more ecological factors present in daily interactions. For instance, while a CEO relying on data scientists to steer their company's strategy would likely be more confident in deliberation-led data crunching, they may also become more comfortable when dealing with intuitive creatives from the advertisement department through repeated meetings. Our preferences may thus shift

with experience and context, which our current vignettes don't consider. Future work may use more refined stimuli to capture this phenomenon.

In sum, our findings reveal a robust preference for deliberation in the context of problem solving that remains significant regardless of one's thinking disposition, yet strengthens alongside one's own tendency to engage in deliberation. While individuals who rely on intuition value that same reasoning mode in others more than their deliberative peers do, they nonetheless maintain a preference for deliberation in others. Returning to our opening illustration: journalists and casual observers may be wooed by the seemingly effortless grace Federer displayed, but a fellow competitor or an aspiring player seeking his advice on how to replicate his backhand would surely prefer a thorough explanation of the moves involved rather than hear him say "Just feel it".

Acknowledgements

Preparation of this manuscript benefitted from support from the European Union under Horizon Europe Programme Grant Agreement no. 101120763 – TANGO. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency (HaDEA). Neither the European Union nor the granting authority can be held responsible for them.

Competing Interests Statement

The authors have no competing interests to declare.

Financial Support Statement

This work was supported by the Horizon Europe Programme Grant Agreement no. 101120763 – TANGO.

Inclusion and ethics statement

Our experimental procedures were approved by the Comité d'Ethique de la Recherche Université Paris Cité protocol 00012023-151.

Declaration of Generative AI Use

The authors used Gemini 3 Flash (Free tier), a large language model developed by Google, between February 18, 2026 and February 26, 2026 to assist in the preparation of this manuscript. The tool was used to refine the clarity and flow of the text and perform final proofreading for grammar and style. The tool is a web-based service accessible at gemini.google.com. The authors remained responsible for the conceptual integrity of the work, and all AI-suggested edits were reviewed and verified.

Appendix

A Between-subject analyses

When looking only at the first profile rated by participants, results are similar to analyses with all profiles.

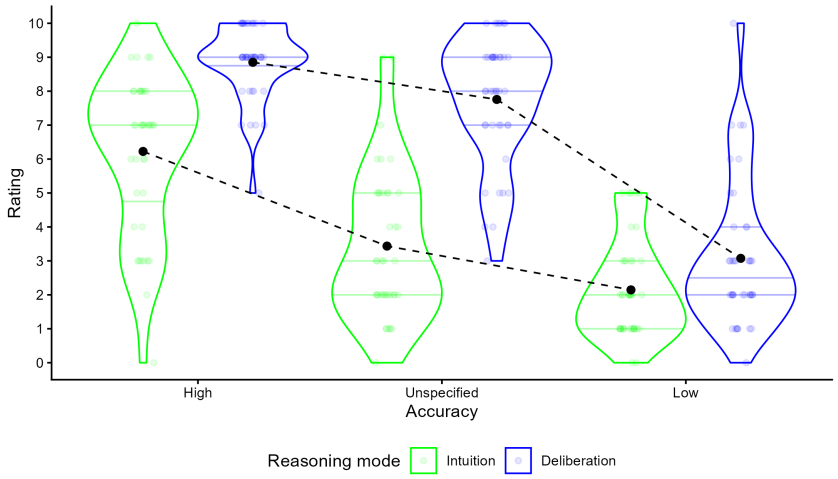


Figure A1. Combined rating scores for the first item. Violin plots indicate the overall perceived reasoning quality for each profile when it was presented first. Black dots indicate the average scores. When looking at the first item only, the same pattern as with all items emerges: a general preference for deliberation over intuition, at every level of accuracy.

Table A2
Between-subject ANOVA.

Effect	*df*	*F*	*p*	η_p^2
Mode	1, 235	117.60 ***	<.001	.34
Accuracy	2, 235	140.63 ***	<.001	.55
Mode:Accuracy	2, 235	16.40 ***	<.001	.12

Table A3
Contrasts for the between-subject ANOVA.

Accuracy	*estimate*	*SE*	*df*	*t*	*p*	*d*
High	2.63	0.42	235	6.25 ***	<.001	1.40
Unspecified	4.32	0.42	235	10.28 ***	<.001	2.30
Low	0.93	0.42	235	2.23 *	.027	0.49

A Holm correction was applied.
Deliberation was always rated significantly higher than intuition, at all accuracy levels.

B Ranking profiles

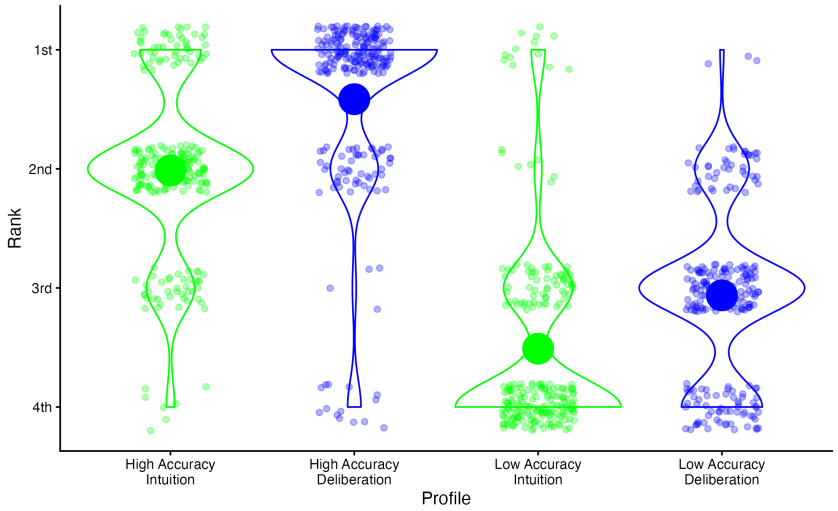


Figure B1. Ranking scores. Violin plots indicate individual rankings. The bigger dots indicate the average rank for each profile. Results indicate that deliberation was ranked higher than intuition, both when past accuracy was high or low.

Table B2
Distribution of rankings.

Ranking	Preferred reasoning mode	Amount
ABCD	Intuition	10.8% (26)
ABDC	Intuition	9.1% (22)
ACBD	Intuition	0.4% (1)
ACDB	Intuition	0.8% (2)
ADBC	Intuition	0.4% (1)
CADB	Intuition	0.8% (2)
CDAB	Intuition	3.3% (8)
DACB	Intuition	0.4% (1)
DCAB	Intuition	0.4% (1)
BACD	Deliberation	14.9% (36)
BADC	Deliberation	42.3% (102)
BCAD	Deliberation	0.4% (1)
BCDA	Deliberation	0.8% (2)
BDAC	Deliberation	12.4% (30)
BDCA	Deliberation	0.8% (2)
CBDA	Deliberation	0.4% (1)
CDBA	Deliberation	0.8% (2)
DBAC	Deliberation	0.4% (1)

Profile labels: A = High accuracy, intuition, B = high accuracy, deliberation, C = low accuracy, intuition, D = low accuracy, deliberation.
73.4% of participants showed a preference for deliberation over intuition in others through their rankings.

C Betting results

Table C1
Overall bets.

Bet	*n*	percent
A	42	17.4%
B	197	81.7%
C	2	0.8%
D	0	0.0%

Profile labels: A = High accuracy, intuition, B = high accuracy, deliberation, C = low accuracy, intuition, D = low accuracy, deliberation.
The majority of participants bet on the highly accurate deliberative profile, rather than the highly accurate intuitive profile.

Table C2
Bets according to the preference for deliberation or intuition with high accuracy.

	Bet			
Preference	A	B	C	D
Deliberation first	7.9% (14)	92.1% (163)	0.0% (0)	0.0% (0)
Intuition first	43.8% (28)	53.1% (34)	3.1% (2)	0.0% (0)
Total	17.4% (42)	81.7% (197)	0.8% (2)	0.0% (0)

Profile labels: A = High accuracy, intuition, B = high accuracy, deliberation, C = low accuracy, intuition, D = low accuracy, deliberation. Deliberation first: all rankings in which the highly accurate deliberative profile was ranked higher than the highly accurate intuitive one. Intuition first: all rankings in which the highly accurate intuitive profile was ranked higher than the highly accurate deliberative one.
Critically, of the participants who ranked the deliberation-accurate profile the highest, 92.1% also consistently bet on it, while only 43.8% of those who ranked the intuition-accurate profile above the deliberation-accurate one bet on it when money was at stake, suggesting that when choices have real monetary consequences, the preference for deliberation over intuition is even more pronounced.

D ANOVAs for bottom and top deciles of the composite

When looking at the first and last deciles of the composite score, i.e., focusing on participants most likely to either rely on intuition or engage in deliberation, analyses point to a general preference for deliberation over intuition across all accuracy levels, as revealed by the main analyses.

Table D1
Separate ANOVAs for the 1st and 10th deciles of the composite score.

Decile	Effect	*df*	*F*	*p*	η_p^2
1st	Mode	1, 23	15.1 **	<.001	.40
	Accuracy	1.3, 29.9	52.8 ***	<.001	.70
	Mode:Accuracy	1.6, 36.3	1.99	.159	.08
10th	Mode	1, 23	53.9 ***	<.001	.70
	Accuracy	1.8, 40.8	174.1 ***	<.001	.88
	Mode:Accuracy	1.98, 45.7	14.8 ***	<.001	.39

The Reasoning mode x Accuracy interaction was not significant for the 1st decile. Apart from that, the results were similar to the general ANOVA in the main analysis.

Table D2
Contrasts for the ANOVAs for the 1st and 10th deciles of the composite score.

Decile	Accuracy	*estimate*	*SE*	*df*	*t*	*p*	*d*
1st	High	2.2	0.5	23	4.5 **	<.001	0.95
	Unspecified	2.4	0.8	23	2.95 *	.007	1.1
	Low	1.2	0.5	23	2.2 *	.035	0.5
10th	High	3.8	0.6	23	6.6 ***	<.001	2.1
	Unspecified	4.3	0.6	23	6.9 ***	<.001	2.4
	Low	1.8	0.4	23	5.2 ***	<.001	1.0

Holm corrections were applied separately for each decile.
Similar to the main analysis, deliberation was always rated significantly higher than intuition, at all accuracy levels.

E Predictors, Rankings and Bets

Table E1
T tests for Individual Differences Predictors.

Predictor	DV	*df*	*t*	*p*	*d*
Raven	Preference	115.40	1.73	.086	0.25
	Bet	65.06	1.38	.171	0.23
logRT_Raven	Preference	103.3	-0.9	.388	-0.1
	Bet	61.3	0.2	.821	0.0
RI_Raven	Preference	112.1	0.5	.646	0.1
	Bet	62.8	0.9	.363	0.2
AOT	Preference	98.73	2.39 *	.019	0.36
	Bet	59.93	2.17 *	.034	0.37
NFC	Preference	126.59	-0.39	.694	-0.06
	Bet	62.90	0.32	.749	0.05
CRT	Preference	103.26	0.41	.680	0.06
	Bet	58.55	0.67	.507	0.11
logRT_CRT	Preference	93.9	-2.4 *	.017	-0.4
	Bet	61.8	-0.9	.364	-0.2
RI_CRT	Preference	96.8	-1.4	.158	0.06
	Bet	58.6	-0.2	.885	-0.2
USID-Intuition	Preference	106.41	-3.38 **	.001	-0.0
	Bet	66.62	-4.18 ***	<.001	-0.68
USID-Deliberation	Preference	122.31	-0.73	.466	-0.10
	Bet	59.00	0.78	.440	0.13
Self-Intuition	Preference	106.64	-3.20	.002**	-0.47
	Bet	61.81	-4.42 ***	<.001	-0.74
Self-Deliberation	Preference	94.13	2.84	.006**	0.43
	Bet	60.83	3.03	.004**	0.51
Self-Binary[a]	Preference	96.38	3.67 ***	<.001	0.56
	Bet	54.24	4.46 ***	<.001	0.80

Table E1
(Continued)

Predictor	DV	*df*	*t*	*p*	*d*
Composite	Preference	118.43	3.02 **	.003	0.43
	Bet	70.51	4.10 ***	<.001	0.65
RI_Composite	Preference	118.43	1.9	.056	0.4
	Bet	75.8	3.6 **	.001	0.7

[a] 0 = intuition, 1 = deliberation

Overall, the preference (and bet) for a given reasoning mode followed the tendency to engage in this reasoning mode.

F Extra analyses

Following a reviewer's suggestion, we included lured responses for the CRT and the Raven task in our analysis, in addition to regular scores (e.g., Byrd & Białek, 2021; Byrd & Conway, 2019; Byrd et al., 2023). Note that the role of lure items in CRT performance is not always clear-cut in the literature (e.g., Baron et al., 2015; Meyer et al., 2024) but the use of RI measures is not dependent on lure items (Baron et al., 2015).

Responses on those tests mostly followed bimodal distributions, where participants virtually considered only the correct response and an intuitive but erroneous (hence lured) alternative response. Table F1 indicates correlations between predictors when using accuracy scores for CRT and Raven., while Table F2 replaced accuracy scores with lured responses. Lured responses for Raven were correlated with the intuition subscale of USID ($r = .37$) but this wasn't the case for lured responses for CRT, which suggests that intuitive responses on the CRT didn't track intuitive thinking dispositions (e.g., Pennycook et al., 2015; Coutinho et al., 2021). Still, lured responses almost mirrored regular scores in their relationships to other predictors (Table F2) and the ratings as well (Table F3), which remains consistent with our main analyses.

Table F1

Correlations between Individual Differences Predictors using Accuracy for CRT and Raven.

Predictor	1	2	3	4	5	6	7	8	9
1. USID-Intuition	-								
2. Self-Intuition	**.50*** **	-							
3. NFC	.01	-.01	-						
4. USID-Deliberation	.08	-.01	**.60*** **	-					
5. Raven	**-.37*** **	-.11	**.21** **	.12	-				
6. CRT	-.06	-.09	**.21*** **	**.17** **	**.37*** **	-			
7. AOT	**-.58*** **	**-.38*** **	**.15***	.02	**.24*** **	**.21*** **	-		
8. Self-Binary[a]	**-.48*** **	**-.63*** **	.08	.09	**.22*** **	**.17** **	**.34*** **	-	
9. Self-Deliberation	**-.41*** **	**-.4*** **	**.25*** **	**.33*** **	**.31*** **	**.17** **	**.39*** **	**.63*** **	-

[a] 0 = intuition, 1 = deliberation

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

Table F2

Correlations between Individual Differences Predictors using Lured Responses for CRT and Raven.

Predictor	1	2	3	4	5	6	7	8	9
1. USID-Intuition	-								
2. Self-Intuition	**.50*** **	-							
3. NFC	.01	-.01	-						
4. USID-Deliberation	.08	-.01	**.60*** **	-					
5. Lure-Raven	**.37*** **	.11	**-.21** **	-.13	-				
6. Lure-CRT	.05	.1	**-.22*** **	**-.17** **	**.36*** **	-			
7. AOT	**-.58*** **	**-.38*** **	**.15***	.02	**-.25*** **	**-.2*** **	-		
8. Self-Binary[a]	**-.48*** **	**-.63*** **	.08	.09	**-.22*** **	**-.16** **	**.34*** **	-	
9. Self-Deliberation	**-.41*** **	**-.4*** **	**.25*** **	**.33*** **	**-.31*** **	**-.18** **	**.39*** **	**.63*** **	-

[a] 0 = intuition, 1 = deliberation

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

Table F3

Correlations between Predictors and Perceived Quality separated by Accuracy and Reasoning mode, including Accuracy, Lured responses and RI measures for CRT and Raven.

Predictor	Reasoning mode	Accuracy			
		High	No	Low	Overall
Raven	Intuition	-.06	**-.15***	**-.27*****	**-.19****
	Deliberation	**.13***	-.04	**-.17****	-.08
Lure-Raven	Intuition	-.06	**-.15***	**.28*****	**.18****
	Deliberation	**-.13***	.04	**.17****	.08
logrT_Raven	Intuition	.09	-.04	**-.15***	-.02
	Deliberation	.06	.03	**-.19****	-.07
RI_Raven	Intuition	.01	-.11	**-.24*****	-.12
	Deliberation	.11	-.01	**-.21****	-.08
CRT	Intuition	-.07	-.11	**-.18****	**-.14***
	Deliberation	**.16***	-.07	-.09	-.04
Lure-CRT	Intuition	.06	.10	**.16***	-.12
	Deliberation	**-.16***	.07	.07	.02
logRT_CRT	Intuition	**.24*****	.09	.08	**.19****
	Deliberation	-.05	-.04	-.02	-.05
RI_CRT	Intuition	.12	-.01	-.07	.03
	Deliberation	.08	-.08	-.08	-.06
Composite	Intuition	**-.19****	**-.36*****	**-.38*****	**-.38*****
	Deliberation	**.25*****	.02	**-.20****	-.01
Composite with lures	Intuition	**-.17****	**-.33*****	**-.37*****	**-.35*****
	Deliberation	**.25*****	-.00	**-.20****	-.03
RI_Composite	Intuition	-.11	**-.31*****	**-.35*****	**-.30*****
	Deliberation	**.22*****	.02	**-.22*****	-.04

[a] 0 = intuition, 1 = deliberation

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

Supplementary Material

Rating instructions

After indicating their Prolific ID, participants were introduced to the rating task with the following instructions:

> *"Please read these instructions carefully.*
>
> *We will give you information about the reasoning style of different individuals. We want to know how you evaluate their reasoning skills. We want you to indicate how good you find this person at reasoning in general (i.e., how smart is this person? To what extent would you trust this person's advice about a reasoning problem?). Each time you will see the following rating scale:"*

A screenshot of the rating scales was displayed on the same page.
Instructions continued below the screenshot:

> *"You will need to click on the rating of your choice for each question and when you're finished you can click on the next button.*
>
> *We will show you an example next."*

This was followed with a practice item. After completing practice, participants proceeded to the main experiment.

Individual differences scales

Table S1
Self-report questions

Order	Question
1	Do you yourself rather use intuition or deliberation?
2	On a 1-to-10 scale, how much do you rely on intuition?
3	On a 1-to-10 scale, how much do you rely on deliberation?

The following instructions were placed above the last 2 questions: *The following 2 questions are independent (i.e., you don't need to have one scale at 4/10 and the other at 6/10 for example).*

Table S2
AOT-13

Order	Statement	Reversed
1	People should always take into consideration evidence that goes against their opinions	No
2	Changing your mind is a sign of weakness	Yes
3	I like to think that my actions are motivated by sound reasons	No
4	It is important to stick to your opinions even when evidence is brought to bear against them	Yes
5	Intuition is the best guide in making decisions	Yes
6	Considering too many different opinions often leads to muddled thinking	Yes
7	One should disregard evidence that conflicts with your current opinions	Yes
8	Coming to decisions quickly is a sign of wisdom	Yes
9	Allowing oneself to be convinced by a solid opposing argument is a sign of good character	No
10	If something I think feels right then I am comfortable, whether or not it is true	Yes
11	A person should always consider new information	No
12	People should revise their conclusions in response to relevant new information	No
13	Certain opinions are just too important to abandon no matter how good a case can be made against them	Yes

The Likert scales went from 1 (*Disagree strongly*) to 6 (*Agree strongly*).

Table S3
NFC-6

Order	Statement	Reversed
1	I would prefer complex to simple problems.	No
2	I like to have the responsibility of handling a situation that requires a lot of thinking.	No
3	Thinking is not my idea of fun.	Yes
4	I would rather do something that requires little thought than something that is sure to challenge my thinking abilities.	Yes
5	I really enjoy a task that involves coming up with new solutions to problems.	No
6	I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought.	No

The Likert scales went from 1 (*Extremely uncharacteristic of me*) to 5 (*Extremely characteristic of me*).

Table S4
CRT questions

Order	Question	Heuristic	Correct
1	If you are running a race and you pass the person in second place, what place are you in?	1st	2nd
2	A farmer had 15 sheep and all but 8 died. How many are left?	7	8
3	Emily's father has three daughters. The first two are named April and May. What is the third daughter's name?	June	Emily
4	A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?	10	5
5	If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?	100	5
6	In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?	24	47

Table S5
USID

Order	System	Subscale	Item	Statement
24	Intuition	Affective	1	When I make a decision, it is more important for me to feel the decision is right than to have a rational reason for it.
32	Intuition	Affective	2	When I make a decision, I trust my inner feeling and reactions.
23	Intuition	Affective	3	With most decisions it makes sense to completely rely on your feelings.
29	Intuition	Affective	4	I prefer drawing conclusions based on my feelings, my knowledge of human nature, and my experience of life.
19	Intuition	Affective	5	Using my gut feelings usually works well for me in figuring out problems in my life.
14	Intuition	Affective	6	I believe in trusting my hunches.
16	Intuition	Affective	7	I hardly ever go wrong when I listen to my deepest gut feelings to find an answer.
12	Intuition	Affective	8	I tend to use my heart as a guide for my actions.
18	Intuition	Spontaneous	1	I generally make snap decisions.
25	Intuition	Spontaneous	2	I make quick decisions.
4	Intuition	Spontaneous	3	I am often aware of how to decide even before I review all aspects.
6	Intuition	Spontaneous	4	I've had enough experience to just know what I need to do most of the time without trying to figure it out every time.
31	Intuition	Spontaneous	5	The right way to decide usually comes to mind almost immediately.
20	Intuition	Spontaneous	6	I typically figure out the way to decide swiftly.
1	Intuition	Spontaneous	7	I quickly do the right thing when deciding because I've often faced almost the same thing before.
10	Intuition	Spontaneous	8	I rarely need to mull things over; how to decide usually becomes quickly apparent.
17	Deliberation	Planning	1	Developing a clear plan is very important to me.
8	Deliberation	Planning	2	I like detailed action plans.

Table S5
(Continued)

Order	System	Subscale	Item	Statement
15	Deliberation	Planning	3	I prefer well-prepared meetings with a clear agenda and strict time management.
27	Deliberation	Planning	4	I make definite engagements, and I follow up meticulously.
3	Deliberation	Planning	5	When I make decisions, I proceed step-by-step.
5	Deliberation	Planning	6	Before making decisions I usually think about the goals I want to achieve.
26	Deliberation	Planning	7	I prefer making detailed plans rather than leaving things to chance.
9	Deliberation	Planning	8	I usually have clear, explainable reasons for my decisions.
11	Deliberation	Knowing	1	I want to have a full understanding of all problems.
2	Deliberation	Knowing	2	I like to analyze problems.
30	Deliberation	Knowing	3	I study every problem until I understand the underlying logic.
7	Deliberation	Knowing	4	I have no problem thinking things through carefully.
22	Deliberation	Knowing	5	I enjoy intellectual challenges.
13	Deliberation	Knowing	6	I enjoy solving problems that require hard thinking.
21	Deliberation	Knowing	7	I prefer complex problems to simple problems.
28	Deliberation	Knowing	8	I enjoy thinking in abstract terms.

The Likert scales went from 1 (*I don't agree*) to 5 (*I agree completely*).

References

Ahlmann-Eltze, C., & Patil, I. (2021). ggsignif: R package for displaying significance brackets for 'ggplot2'. *PsyArxiv*. https://doi.org/10.31234/osf.io/7awm6.

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. (2019a). The intuitive greater good: Testing the corrective dual process model of moral cognition. *Journal of experimental psychology: General*, *148*(10), 1782–1801. https://doi.org/10.1037/xge0000533.

Bago, B., & De Neys, W. (2019b). 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.

Baron, J. (1991). Informal reasoning and education. In J. F. Voss, D. N. Perkins, & J. W. Segal (Eds.). Lawrence Erlbaum Associates, Inc.

Baron, J. (2024). Two components of individual differences in actively open-minded thinking standards: Myside bias and uncertainty aversion. *Thinking & Reasoning*, *30*(4), 648–673. https://doi.org/10.1080/13546783.2024.2360491.

Baron, J., Scott, S., Fincher, K., & Metz, S. E. (2015). Why does the cognitive reflection test (sometimes) predict utilitarian moral judgment (and other things)? [Modeling and Aiding Intuition in Organizational Decision Making]. *Journal of Applied Research in Memory and Cognition*, *4*(3), 265–284. https://doi.org/10.1016/j.jarmac.2014.09.003.

Beauvais, N., Voudouri, A., Boissin, E., & De Neys, W. (2025). System 2 and cognitive transparency: Deliberation helps to justify sound intuitions during reasoning. *Thinking & Reasoning*, *0*(0), 1–26. https://doi.org/10.1080/13546783.2025.2532653.

Bonnefon, J.-F. (2018). The pros and cons of identifying critical thinking with system 2 processing. *Topoi*, *37*(1), 113–119. https://doi.org/10.1007/s11245-016-9375-2.

Bonnefon, J.-F., & Rahwan, I. (2020). Machine thinking, fast and slow. *Trends in Cognitive Sciences*, *24*(12), 1019–1027. https://doi.org/10.1016/j.tics.2020.09.007.

Bors, D. A., & Stokes, T. L. (1998). Raven's advanced progressive matrices: Norms for first-year university students and the development of a short form. *Educational and Psychological Measurement*, *58*(3), 382–398. https://doi.org/10.1177/0013164498058003002.

Brafman, O., & Brafman, R. (2008). *Sway: The irresistible pull of irrational behavior*. Crown Currency.

Byrd, N. (2025). A two-factor explication of 'reflection': Unifying, making sense of, and guiding the philosophy and science of reflective reasoning. https://doi.org/10.31234/osf.io/d628j_v3.

Byrd, N., & Białek, M. (2021). Your health vs. my liberty: Philosophical beliefs dominated reflection and identifiable victim effects when predicting public health recommendation compliance during the covid-19 pandemic. *Cognition*, *212*, 104649. https://doi.org/https://doi.org/10.1016/j.cognition.2021.104649.

Byrd, N., & Conway, P. (2019). Not all who ponder count costs: Arithmetic reflection predicts utilitarian tendencies, but logical reflection predicts both deontological and utilitarian tendencies. *Cognition*, *192*, 103995. https://doi.org/10.1016/j.cognition.2019.06.007.

Byrd, N., Joseph, B., Gongora, G., & Sirota, M. (2023). Tell us what you really think: A think aloud protocol analysis of the verbal cognitive reflection test. *Journal of Intelligence*, *11*(4). https://doi.org/10.3390/jintelligence11040076.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*, 116–131. https://doi.org/10.1037/0022-3514.42.1.116.

Chabris, C., & Simons, D. (2010). *The invisible gorilla: And other ways our intuition deceives us*. Crown Publishers/Random House.

Coelho, G. L. d. H., Hanel, P. H. P., & Wolf, L. J. (2020). The very efficient assessment of need for cognition: Developing a six-item version* [PMID: 30095000]. *Assessment*, *27*(8), 1870–1885. https://doi.org/10.1177/1073191118793208.

Coutinho, M. V. C., Thomas, J., Alsuwaidi, A. S. M., & Couchman, J. J. (2021). Dunning-kruger effect: Intuitive errors predict overconfidence on the cognitive reflection test. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.603225.

De Neys, W. (2023). Advancing theorizing about fast-and-slow thinking. *Behavioral and Brain Sciences*, *46*, e111. https://doi.org/10.1017/S0140525X2200142X.

De Neys, W. (2025). Defining deliberation for dual-process models of reasoning. *Nature Reviews Psychology*, *4*(8), 544–552. https://doi.org/10.1038/s44159-025-00466-6.

De Neys, W., & Raoelison, M. (2025). Humans and llms rate deliberation as superior to intuition on complex reasoning tasks. *Communications Psychology*, *3*(1), 141. https://doi.org/10.1038/s44271-025-00320-8.

de Becker, G. (1997). *The gift of fear: Survival signals that protect us from violence*. Little, Brown.

Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking deathworthy: Perceived stereotypicality of black defendants predicts capital-sentencing outcomes. *Psychological Science*, *17*(5), 383–386.

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

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

Firke, S. (2023). *janitor: Simple Tools for Examining and Cleaning Dirty Data* [R package version 2.2.0]. Available at: https://CRAN.R-project.org/package=janitor.

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

Gärtner, A., Grass, J., Wolff, M., Goschke, T., Strobel, A., & Strobel, A. (2021). No relation of need for cognition to basic executive functions. *Journal of Personality*, *89*(6), 1113–1125. https://doi.org/10.1111/jopy.12639.

Gigerenzer, G. (2007). *Gut feelings: The intelligence of the unconscious*. Viking.

Iannone, R., Cheng, J., Schloerke, B., Hughes, E., Lauer, A., Seo, J., Brevoort, K., & Roy, O. (2024). *Gt: Easily create presentation-ready display tables* [R package version 0.11.1]. Available at: https://CRAN.R-project.org/package=gt.

Inbar, Y., Cone, J., & Gilovich, T. (2010). People's intuitions about intuitive insight and intuitive choice. *Journal of Personality and Social Psychology*, *99*(2), 232–247.

JASP Team. (2025). JASP (Version 0.19.3)[Computer software]. Available at: https://jasp-stats.org/.

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

Kassambara, A. (2023a). *Ggpubr: 'ggplot2' based publication ready plots* [R package version 0.6.0]. Available at: https://CRAN.R-project.org/package=ggpubr.

Kassambara, A. (2023b). *Rstatix: Pipe-friendly framework for basic statistical tests* [R package version 0.7.2]. Available at: https://CRAN.R-project.org/package=rstatix.

Kupor, D. M., Tormala, Z. L., Norton, M. I., & Rucker, D. D. (2014). Thought calibration: How thinking just the right amount increases one's influence and appeal. *Social Psychological and Personality Science*, *5*(3), 263–270. https://doi.org/10.1177/1948550613499940.

Legault, M. R. (2006). *Think!: Why crucial decisions can't be made in the blink of an eye*. Simon & Shuster.

Lenth, R. V. (2024). *Emmeans: Estimated marginal means, aka least-squares means* [R package version 1.10.5]. Available at: https://CRAN.R-project.org/package=emmeans.

Meyer, A., Attali, Y., Bar-Hillel, M., Frederick, S., & Kahneman, D. (2024). Cognitive reflection is a distinct and measurable trait. *Proceedings of the National Academy of Sciences*, *121*(49), e2409191121. https://doi.org/10.1073/pnas.2409191121.

Müller, K. (2020). *here: A Simpler Way to Find Your Files* [R package version 1.0.1]. Available at: https://CRAN.R-project.org/package=here.

Oktar, K., & Lombrozo, T. (2022). Deciding to be authentic: Intuition is favored over deliberation when authenticity matters. *Cognition*, *223*, 105021. https://doi.org/10.1016/j.cognition.2022.105021.

Oster, N., & Koesterich, R. (2013). Breaking bad behaviors: Understanding investing biases and how to overcome them. *iShares Market Perspectives*, 1–9.

Pachur, T., & Spaar, M. (2015). Domain-specific preferences for intuition and deliberation in decision making. *Journal of Applied Research in Memory and Cognition*, *4*(3), 303–311. https://doi.org/10.1016/j.jarmac.2015.07.006.

Pedersen, T. L. (2024). *Patchwork: The composer of plots* [R package version 1.3.0]. Available at: https://CRAN.R-project.org/package=patchwork.

Pennycook, G., Bago, B., & McPhetres, J. (2023). Science beliefs, political ideology, and cognitive sophistication. *Journal of Experimental Psychology: General*, *152*(1), 80–97. https://doi.org/10.1037/xge0001267.

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.

Peters, E. (2012). Beyond comprehension: The role of numeracy in judgments and decisions. *Current Directions in Psychological Science*, *21*(1), 31–35. https://doi.org/10.1177/0963721411429960.

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.

Reyna, V. F. (2004). How people make decisions that involve risk: A dual-processes approach. *Current Directions in Psychological Science*, *13*(2), 60–66. https://doi.org/10.1111/j.0963-7214.2004.00275.x.

Reyna, V. F. (2012). A new intuitionism: Meaning, memory, and development in fuzzy-trace theory. [PMID: 25530822]. *Judgment & Decision Making*, *7*(3), 332–359.

Richardson, E., & Keil, F. C. (2022). Thinking takes time: Children use agents' response times to infer the source, quality, and complexity of their knowledge. *Cognition*, *224*, 105073. https://doi.org/10.1016/j.cognition.2022.105073.

Robinson, D., Hayes, A., & Couch, S. (2024). *Broom: Convert statistical objects into tidy tibbles* [R package version 1.0.7]. Available at: https://CRAN.R-project.org/package=broom.

Schmidt, H. G., Van Gog, T., Schuit, S. C. E., Van den Berge, K., Van Daele, P. L., Bueving, H., Van der Zee, T., Van den Broek, W. W., Van Saase, J. L. C. M., & Mamede, S. (2017). Do patients' disruptive behaviours influence the accuracy of a doctor's diagnosis? a randomised experiment. *BMJ Quality & Safety*, *26*(1), 19–23. https://doi.org/10.1136/bmjqs-2015-004109.

Sievert, C. (2020). *Interactive web-based data visualization with r, plotly, and shiny*. Chapman; Hall/CRC.

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2024). *Afex: Analysis of factorial experiments* [R package version 1.4-1]. Available at: https://CRAN.R-project.org/package=afex.

Slowikowski, K. (2024). *Ggrepel: Automatically position non-overlapping text labels with 'ggplot2'* [R package version 0.9.6]. Available at: https://CRAN.R-project.org/package=ggrepel.

Stanovich, K. E., & Toplak, M. E. (2012). Defining features versus incidental correlates of type 1 and type 2 processing. *Mind & Society*, *11*(1), 3–13. https://doi.org/10.1007/s11299-011-0093-6.

Stanovich, K. E., & Toplak, M. E. (2023). Actively open-minded thinking and its measurement. *Journal of Intelligence*, *11*(2). https://doi.org/10.3390/jintelligence11020027.

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of educational psychology*, *89*(2), 342–357. https://doi.org/10.1037/0022-0663.89.2.342.

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.

Thompson, W. C., & Schumann, E. L. (1987). Interpretation of statistical evidence in criminal trials: The prosecutor's fallacy and the defense attorney's fallacy. *Law and Human Behavior*, *11*(3), 167–187.

Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment & Decision Making*, *11*(1), 99–113.

Wagenmakers, E.-J., Krypotos, A.-M., Criss, A. H., & Iverson, G. (2012). On the interpretation of removable interactions: A survey of the field 33 years after loftus. *Memory & Cognition*, *40*(2), 145–160. https://doi.org/10.3758/s13421-011-0158-0.

Wickham, H., Averick, M., Bryan, J., Chang, W., D'Agostino McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686.

World Bank. (2015). *World Development Report 2015: Mind, Society, and Behavior*. https://doi.org/10.1596/978-1-4648-0342-0.

Xie, Y. (2024). *Knitr: A General-Purpose Package for Dynamic Report Generation in R* [R package version 1.49]. Available at: https://yihui.org/knitr/.