Defining System 2 deliberation for dual-process models

Wim De Neys

LaPsyDE (UMR CNRS 8240), Université Paris Cité, Paris, France

(In press). Nature Reviews Psychology.

Preprint - see the journal for the final published version

Corresponding author :

Wim De Neys
Sorbonne – LaPsyDE (UMR CNRS 8240)
Université Paris Cité
46 Rue Saint-Jacques
75005 Paris
France

Email: wim.de-neys@u-paris.fr

Abstract

Deliberation, commonly referred to as "System 2" thinking, is a core component of popular dual-process models of reasoning. However, its precise conceptualization has received limited attention. In this perspective, I present a basic framework and organizing principles to think of deliberation and avoid problematic misconceptions. I argue that deliberation should be understood as multifunctional, serving multiple, complementary purposes. I highlight four such functions (response control, response generation, response justification, and regulation), identify challenges in our current understanding, propose guiding principles, and discuss critical outstanding questions. The resulting framework should help advance the theoretical and empirical study of deliberation in the coming years.

Keywords: Thinking; Deliberation; Dual-Process Models

Defining System 2 deliberation for dual-process models

Human reasoning has long been conceptualized as an interplay between intuitive and deliberate thought processes—often referred to in popular dual-process models as System 1 and System 2 thinking (Kahneman, 2011). These models have sparked numerous debates, such as whether intuition and deliberation are qualitatively distinct (De Neys, 2021; Evans & Stanovich, 2013; Gawronski et al., 2014), whether they operate in parallel or serially (De Neys, 2012; Evans, 2007; Sloman, 1996), and how individuals switch between intuitive and deliberate thinking (Ackerman & Thompson, 2017; De Neys, 2023; Evans, 2019; Pennycook et al., 2015; Stanovich, 2018).

Amid these debates, much attention has been devoted to characterizing intuitive thinking (Betsch & Glöckner, 2010; De Neys, 2023; Ghasemi et al., 2022; Gigerenzer, 2007; Kruglanski & Gigerenzer, 2011; Meyer & Frederick, 2023; Morewedge & Kahneman, 2010; Reber & Allen, 2022; Thompson, 2014; Volz & Zander, 2014). In this paper, I focus on the other end of the spectrum—deliberation, or System 2—and aim to address common misconceptions in how we often conceptualize deliberation.

A central argument is that deliberation should be understood as multifunctional, serving multiple, non-mutually exclusive purposes. I will highlight four core functions of deliberation: response control, response generation, response justification, and regulation. While each of these functions is well-documented in its own right, they are often examined in isolation, which tends to distort theoretical understanding. Moreover, the conceptualization of these functions is beset by various potential challenges and misconceptions. This paper aims to address these issues and propose a more productive framework, offering guiding principles for the study of deliberation moving forward.

Context

To avoid confusion, the goal of this paper is not to develop a new dual process model but to present a fresh perspective on a critical element of dual process models—deliberation. I focus on deliberation in the context of contemporary dual-process frameworks (De Neys, 2023; Pennycook, 2017; Pennycook et al., 2015; Thompson & Newman, 2017) to illustrate

where deliberation typically fits within dual-process thinking and provide a concrete starting point for the general reader

In these frameworks, deliberation is typically triggered by an uncertainty monitoring process within System 1, which is commonly described as a collection of intuitively cued responses. The activation strength of these responses determines the level of certainty about System 1 operations. When uncertainty arises—such as when multiple conflicting responses are strongly activated—deliberation is initiated (see Figure 1).

Deliberation, associated with System 2, works to adjust or modulate the intuitions generated by System 1, typically aiming to resolve uncertainty. Once uncertainty is reduced, the deliberative process can stop. The key question I address here is how to conceptualize what happens inside the "black box" of System 2 deliberation—what functions it serves, how it operates, and how we can best think about it to advance theoretical and empirical understanding.

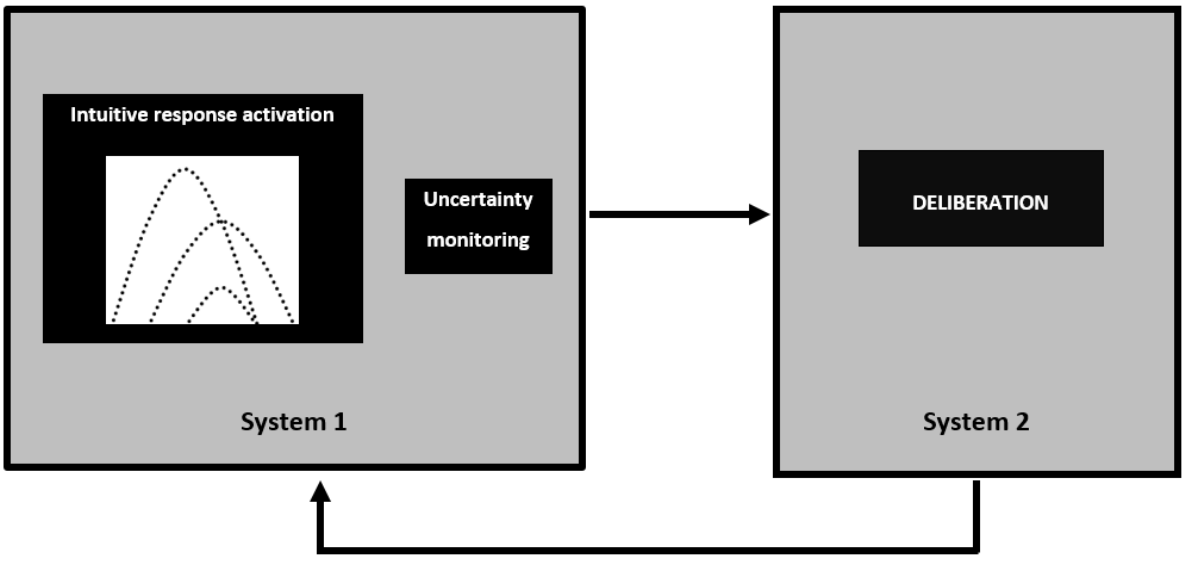

Figure 1. Simplified illustration of potential System 1 and System 2 interaction in contemporary dual-process models. In these models, System 1 processing is conceptualized as a collection of intuitively cued responses, each with varying activation strengths (depicted by the dotted curves). An uncertainty monitoring process evaluates the activation strengths to determine the level of uncertainty in the system. High uncertainty arises when multiple responses are strongly activated or when only a weakly activated—or no—response is cued. This uncertainty triggers System 2 deliberation, which can, in turn, modulate System 1 activation (illustrated by the arrow from System 2 to System 1).

The Multifunctional System 2

If System 2 serves multiple roles, what are these roles? Broadly, there are at least four main functions that scholars have considered, though often not in combination: (1) controlling or inhibiting intuitively generated responses, (2) generating new responses, (3) justifying intuitively generated responses, and (4) regulatory effort allocation. These functions have been referred to by a variety of terms, but I will use the generic labels of Control, Generation, Justification, and Regulation. The following section introduces each function in more detail and discusses misconceptions and challenges associated with them.

Response control. Sometimes the intuitive System 1 may generate a salient response that is not appropriate in the situation at hand. In these cases, we will need to refrain from acting on the intuitive impulse and inhibit or suppress it. For example, when on a diet you will need to control your craving to devour a bag of potato chips when you're hungry. When you are instructed to name the ink color in which a word is written in a Stroop task, you will need to refrain from reading the word. In reasoning tasks, you will often need to control the tendency to blast out the first response that comes to mind (e.g., "10 cents!" in the bat-and-ball problem, see below). In general, these are all cases in which effortful deliberation is required to control an intuitively generated response. This response control function is commonly referred to as "attentional control" (Draheim et al., 2022), "override" (Evans & Stanovich, 2013), "impulse control" (Baumeister et al., 2007; Hofmann et al., 2009), "inhibition" (De Neys & Bonnefon, 2013; Houdé & Borst, 2015), or "decoupling" (Pennycook et al., 2015; Stanovich, 2005) and it is widely considered as a core function of System 2 processing. One possible general instantiation is to think of it as a lowering of the activation strength of an intuitive response (Houdé, 2019).

There is little doubt that response control is an important aspect of deliberation. The point is that it is questionable that it is the only function of deliberation. This is especially clear in higher-order reasoning tasks. Consider for example the Cognitive Reflection Test (CRT, Frederick, 2005). The popular test features various problems in which intuitive reasoning cues a salient but incorrect hunch (e.g., "A bat and ball cost $1.10 together. The bat costs $1 more than the ball. How much does the ball cost?"—incorrect intuitive answer: 10 cents, correct answer: 5 cents) and it is widely used as an index of people's tendency to engage in

deliberation to correct their intuition. Clearly, if one does not refrain from shouting “10 cents” and give oneself the time to reason it through, one will often fail the test. Response control has its role to play here. But System 2 deliberation doesn’t stop there. Response control is not necessarily sufficient for successful reasoning. For example, in the cognitive reflection test, a reasoner may have controlled the tendency to shout “10 cents”, realized that the 10 cents answer is questionable but failed to find the correct procedure to arrive at the correct 5 cents response, and therefore nevertheless defaulted to their first “10 cents” hunch. Hence, what multifunctionality of deliberation implies in this case is that giving the 10-cents response (and a resulting low CRT score) cannot be equated with a lack of deliberation or a tendency to not go beyond one’s first hunch.

A failure to appreciate this simple but critical distinction is bound to distort our theories and applications. For example, based on the observation that participants perform worse on the CRT after testosterone administration, it has been hypothesized that testosterone reduces deliberation and increases impulsive, System 1 reasoning (Knight et al., 2020; Nave et al., 2017). But since reasoners who failed the test may nevertheless have engaged in deliberate response generation, this conclusion is invalid. Likewise, correlational findings such as that those who believe fake news and conspiracy theories show lower CRT scores, do not necessarily imply that these people are lazy thinkers and deliberate less (Martire et al., 2023; Pennycook, 2023; Pennycook et al., 2016). If response control were all there is to deliberation or the CRT test, this inference might be valid. However, if deliberation is multifunctional and also entails, for example, response generation in addition to response control, it clearly is not. Put differently, those with low CRT scores might not be less likely to control their intuitive responses and engage in deliberation but rather be simply less likely to complete the deliberate response generation successfully.

There is good empirical evidence for this point too. For example, latency data shows that people who give an erroneous “intuitive” answer on a CRT question, easily take up to 30 s to do so—indicating they are not blindly going for their first hunch (Evans, 2019; Johnson et al., 2016; Stupple et al., 2017). Similarly, Martire et al. (2023) observed that those who endorsed false beliefs more, took significantly longer to solve the CRT questions than non-endorsers, suggesting that the lower CRT scores do not result from a mere failure to deliberate on the problem.

Note that leading dual process theorists such as Stanovich (2011) and Evans (2019) have long included a separate response control and generation function in their taxonomies of reasoning errors and System 2 accounts. Arguably, one reason that these are nevertheless not always differentiated is the popularity of the CRT as a deliberation measure and the way it was originally conceived. As Meyer and Frederick (2023) point out, Frederick (2005) and Kahneman and Frederick (2002) assumed that the basic calculations to arrive at the 5 cents response are so straightforward that anyone who would realize that the intuitively cued response is incorrect and deliberate on the problem, would also solve it. Under this assumption the CRT indeed becomes a pure measure of response control. However, as noted above, by now there is sufficient evidence indicating this is not the case. Meyer and Frederick (2023) have consequently revised their views but the original conceptualization still remains influential.

To avoid confusion, the point is not that response control is not important for deliberation or that the CRT is not a good measure of deliberation. They clearly are. The point is that response control is not the only function of deliberation. The CRT—and deliberation in general—does not only imply the capacity to control intuitive responses but also the capacity to successfully complete the subsequent computations required to arrive at an (alternative) response. This underscores that multifunctionality needs to be integrated in our theoretical and applied accounts of deliberation.

Response generation: As alluded to above, in addition to response control, people may also need to engage in deliberation to generate an answer to a problem. There can be multiple pathways via which response generation is achieved. For example, some authors have stressed the role of hypothetical thinking or mental simulation (Evans, 2007; Evans & Stanovich, 2013). Mental simulation is the process whereby we envision a situation that is not factually the case--what allows us to consider alternative possibilities and hypotheses (hence, the label “hypothetical thinking”). A possible illustration would be the application of a mental “proof-by-contradiction” to solve the bat-and-ball problem, for example. Here one hypothetically assumes that the correct answer is that the ball costs 10 cents. This allows us to conclude that—at a dollar more—the bat will cost $1.10. This in turn leads to the conclusion that together the bat and ball cost $1.20. But this violates the stated premise that the grand total is $1.10. This contradiction allows us to reject the hypothesis that the ball

costs 10 cents and implies that the ball must cost less than 10 cents. Simply repeating the hypothetical thinking process with a smaller value (e.g., assume the ball costs 7 cents, 5 cents, etc.) will then eventually lead to the correct answer (i.e., the answer that avoids the contradiction).

Mental simulation is but one possible route to deliberate response generation in System 2. Some authors have also pointed to the algorithmic nature of deliberation and its role in response generation (Houdé, 2019). In this case deliberation entails retrieving and executing a stepwise sequence of rules. For example, when we have to do long division or multiply multiples of 10 (e.g., "How much is 220 * 30?"), we can use a longdivision or multiplication algorithm (e.g., multiply the non-zero part of the numbers, i.e., 22 * 3 = 66; count the zeros in each factor, i.e., 2; add the same number of zeros to the product, i.e., 6600, De Neys, 2023) to calculate an answer. In the case of the bat-and-ball problem, one could apply an algebraic substitution algorithm, for example (e.g., rewrite problem as linear equations; solve one of the equations for either x or y; Substitute the step 2 solution in the other equation; Solve the new equation obtained using arithmetic operations; Hence, X + Y = 1.10. X = Y + 1; X + X + 1 = 1.10; 2X = 1.10-1; X = 0.10/2; X = .05).

Note that in terms of the wider dual-process interaction between System 1 and 2 (see Figure 1), I hypothesize that whenever we deliberately generate a new response, this response is also represented as an intuitive response in System 1—such that it can compete with other intuitive responses, for example—and modulate System 1 uncertainty. However, my concern here does not pertain to the precise processing specification or the specific instantiation of deliberate response generation (i.e., whether it entails mental simulation, algorithmic thinking or some other process). Although such specification will be important, my point concerns the functional characterization of deliberation. Whatever the best instantiation of response generation turns out to be, it is important that our theories underscore the generative function and factor in that people also deliberate to compute a solution to a problem.

It is equally important to avoid possible misunderstanding about this core function. Critical here is what Evans (2012) has labelled the normative fallacy or the idea that System 2 deliberation would by definition be normatively correct. Clearly, spending time and effort to generate a response does not guarantee its correctness. While engaging in deliberation *can* lead to a correct response, there is no certainty. For instance, when using algorithmic thinking,

we might misinterpret the input or apply an incorrect algorithm, such as mistakenly reading a subtraction problem as a multiplication problem (e.g., 220 * 30 = ? is read as 220 + 30 = ?). Likewise, in a reasoning problem such as the bat-and-ball, one may overlook the "more than" statement and simply read the second premise as "The bat costs $1" (Mata et al., 2017). But even when one applies the correct algorithm to the correct input, one may run out of resources to keep track of the intermediary computations or states (e.g., in the multiplication algorithm we may add an erroneous number of zeros to the product). Hence, although deliberation may often help, it does not guarantee that our deliberately generated responses are sound.

Second, at the other end of the spectrum, it is crucial to stress that deliberation should not be conceived as being necessary for the generation of correct or sound responses per se. On one hand, the reasoning and decision-making field has capitalized on tasks and cases in which intuitive System 1 reasoning cues responses that conflict with logico-mathematical principles. But in many other tasks and cases, mere intuitive processing may obviously be perfectly valid (Evans & Stanovich, 2013; Gigerenzer, 2007). Critically, this even holds for classic reasoning tasks. Indeed, recent studies indicate that people who manage to respond correctly to "trick" problems such as the bat-and-ball problem often can do so intuitively (Bago & De Neys, 2017, 2019). These studies use a two-response paradigm in which participants are asked to first respond as quickly as possible with the first response that comes to mind. Afterwards, they can take all the time they want to reflect on the problem an generate a final response. To make maximally sure that participants do not deliberate during the initial response stage, they are forced to respond within a very tight deadline and while their cognitive resources are burdened with a secondary load task (e.g., memorizing digits or a complex visual pattern). As a side note, the rationale here is that deliberation is operationally defined as time and resource demanding. Hence, by not giving people the bandwidth to reflect, one can experimentally minimize deliberation in the initial, "intuitive" response stage. By now, numerous studies have shown that reasoners who give a correct response in the final response stage, often already give the same correct response in the initial, intuitive stage (Bago & De Neys, 2017, 2019; Buric & Konradova, 2021; Buric & Srol, 2020; Dujmovic et al., 2021; Raoelison et al., 2020; Thompson & Johnson, 2014). Good reasoners do not necessarily need to deliberate to generate a correct response, since often their intuitive response is already accurate.

In sum, we must avoid the misconception that deliberate response generation is always sufficient or necessary for sound reasoning. Yet, it's evident that people often deliberate to calculate a response, and any model of deliberation must incorporate this fundamental function.

Response justification. Even when there is no need to control or generate a response, people may still need to deliberate to justify an intuitively generated answer. One of the features that is associated with intuitive processing is that it is cognitively non-transparent (Bonnefon, 2018; Evans & Stanovich, 2013). We are typically aware of the output or end product of intuitive processing but we often have no awareness of how or why the response was generated. As Bago and De Neys (2019) noted, it is precisely the absence of such processing insight or justification that is one of the reasons to label intuitions as "gut feelings" (Marewski & Hoffrage, 2015; Mega & Volz, 2014). Deliberate processing can allow us to look for explicit reasons or arguments to justify an intuitively generated response. This process is often referred to as rationalization (Evans & Stanovich, 2013; Pennycook et al., 2015) and will be critical to convince others (or ourselves) to adopt our intuitively generated problem solutions (Cushman, 2020; Mercier & Sperber, 2011, 2017).

The role of rationalization in deliberation has long been acknowledged in dual-process research (Evans & Wason, 1976; Wason & Evans, 1975). Evans (2019), for example, recalls how when he was running his experiments on the infamous Wason card selection task, gaze tracking indicated that people mostly looked only at the (incorrect) cards they ended up selecting (i.e., the cards that matched the instances in the rule they needed to test). However, although people instantly focused on these cards, they nevertheless often took up to 30 s to confirm their choices. Evans reasoned that the only plausible explanation for the delay was that people were engaging in deliberation to look for reasons that supported their intuitive choice and convince themselves they were correct.

The problem with this classic dual process conceptualization is that it tends to tie rationalization to rationalization of *incorrect* responses. This has led to quite a negative connotation. Rationalization is considered as a sort of "making up excuses after the facts" that is epiphenomenal and even detrimental to reasoning (Cushman, 2020; Ellis & Schwitzgebel, 2020; Myers & Chater, 2024). Clearly, given that our intuitions are often non-transparent, deliberately looking for justifications *can* lead to biases in that we may generate

arguments to support invalid responses. However, this sketches only half the story. Justification or rationalization is equally important for correct intuitions. A good illustration comes from the two response studies I referred to above. To recap, these studies showed that sound reasoners often generate correct intuitive responses to classic reasoning problems (Bago & De Neys, 2017). However, it has also been observed that these sound intuitors struggle to justify and explain why their answer is correct in the absence of deliberation (Bago & De Neys, 2019; Beauvais et al., 2024). For example, Bago and De Neys asked reasoners after the initial and final response stages to justify their responses. They observed that after the final response stage (i.e., when participants had had the possibility to deliberate about their answer), correct answers were typically also correctly justified (e.g., in case of the bat-and-ball problem, participants would say something like "well, it has to be 5 cents because then at a dollar more the bat will cost $1.05 and the total will be $1.10", Bago & De Neys, 2019). However, although these reasoners often already generated the same correct "5 cents" response as their initial response, they struggled to provide a correct justification at this stage.

As Mercier and Sperber (2017) have stressed, such a justification process in which we look for explicit reasons in support of our intuitions can be critical to efficiently sway others. If you want to convince your peers that your solution to a problem is right, you will be more successful when giving them an explicit, verifiable argument than by simply telling them that you "felt" it was right (Bago & De Neys, 2019). And even when reasoning in isolation (and "having an argument with oneself") a deliberate explication of the reasons behind our intuitive beliefs can be equally helpful (even though it may be less efficient than in a social, group setting, Mercier & Sperber, 2011).

To avoid confusion, it is important to stress that the justification process—as a search for explicit reasons or arguments—does not entail that only reasons consistent with the intuitive response will be searched or considered. Clearly, a justification process that would only look for supporting reasons is self-contradictory. If the search for reasons is restricted to supportive arguments, one postulates an omniscient reasoning engine that knows in advance which arguments are supportive or contradictory and where to look for them. While one can postulate additional machinery or processes that may discard certain arguments once they are retrieved, assuming that a justification process only considers supporting evidence begs the question. Note that I opted for the label "justification" rather than mere "rationalization"

precisely because the dual-process literature typically only ties rationalization to a process in which reasons consistent with the intuitive response are considered (Pennycook et al., 2015).

Considering the interaction between System 1 and 2, as with the other deliberation functions, the outcome of the justification process is assumed to be "fed-back" to System 1 by affecting the activation strength of a corresponding intuitive response. Finding an explicit reason or argument to support an intuitive response will bolster one's confidence in the answer, which can be conceived as an increase in its activation strength (De Neys, 2023). There is indeed evidence that deliberation often tends to increase confidence (Koehler, 1991; Shynkaruk & Thompson, 2006). Clearly, there can be exceptions, for example, in cases where justification fails to find an explicit reason (Evans, 2019). The general point is simply that deliberate justification can also modulate System 1.

I clarified how deliberate processing can allow us to look for explicit reasons or arguments for non-transparent intuitively generated responses. Since deliberately generated responses are by definition transparent, justification is less critical for deliberate responses. But this does not entail that people cannot engage in response justification for deliberately generated responses. Although the reasons behind a deliberately generated response will be clear, we may still engage in additional justification to look for additional supporting or contradictory reasons (e.g., to win an argument or better convince others).

One may also note that my take on justification is orthogonal to the wider, philosophical issue as to whether it is possible to uncover the true causes of intuitive responses or not (Dennett, 1989). That is, given the non-transparent nature of intuitive processing, at the surface, justification may be considered as a tool to "open-up the black box" and identify the precise intuitive computations that led to an intuitive response. But various philosophers and cognitive scientists have questioned whether this is possible (Carruthers, 2011; Chater, 2018; Dennett, 1989). In a nutshell, the idea is that justification is always backward-looking and constitutes an all new, independent construction rather than an accurate reconstruction of the actual reasons that led to our desires and thoughts (Dennett, 1989). Although the question is intriguing, it should be noted that it is ultimately irrelevant for the current functional characterization of deliberation. Whether the generated justification accurately reflects the true underlying intuitive computations or not, what matters is that the justification process generates an explicit reason that we can evaluate and communicate to others (and ourselves, Cushman, 2020).

In sum, the point is that given that intuitive responding is typically non-transparent, deliberate justification will often be critical to explicitate our reasoning process and look for reasons that support (or contradict) our intuitions. Although there is no guarantee that deliberate justification will lead to valid or correct justifications per se, there is no theoretical reason to assume it is bound to lead to biased justifications either. As such, it is an essential component of the reasoning process.

Regulation. People sometimes deliberate to monitor and regulate their thinking, reflecting on how to allocate cognitive resources. For instance, they may decide whether it is worthwhile to continue searching for a justification. In response to specific instructions or incentives, individuals can deliberately choose to invest more or less effort in reasoning (Sirota et al., 2023). Similarly, they can evaluate whether a newly generated response or justification aligns with or contradicts another response or justification they have in mind. These reflective cases of "thinking about thinking" have been extensively studied in the fields of metacognition and meta-reasoning (Ackerman & Thompson, 2017). I group these processes here under the "Regulation" header.

Traditionally, dual-process theorists have emphasized this regulatory function. Early influential models suggested that a central role of System 2 was to monitor System 1 for conflict and to decide when effortful deliberation should be initiated (Kahneman, 2011; Sloman, 1996; Toplak & Stanovich, 2023). The idea was that a reasoner would switch from System 1 to System 2 processing whenever conflict or uncertainty was detected. However, this view is problematic because it creates a circular explanation (De Neys, 2012; Evans, 2019; Stanovich, 2018). To explain when people engage in deliberation, we must assume that they already deliberate to regulate the monitoring process.

Contemporary dual-process models, such as the one illustrated in Figure 1, therefore posit that the decision to engage System 2 is driven by uncertainty monitoring within System 1 (De Neys, 2023; Evans, 2019; Pennycook et al., 2015; Stanovich, 2018). Empirical evidence supports this view. For example, in classic reasoning tasks like the bat-and-ball problem, people who provide incorrect responses ("10 cents!") often exhibit signs of intuitive error detection or monitoring. For instance, they report decreased confidence in their erroneous answers (Bago & De Neys, 2017; Buric & Srol, 2020; Johnson et al., 2016; Thompson & Johnson, 2014). Crucially, this occurs even when deliberation is experimentally minimized,

such as during the initial response phase of the two-response paradigm introduced earlier, suggesting that the error monitoring is intuitive in nature (De Neys, 2023; Pennycook et al., 2015).

The critical misconception I aim to address here is the assumption that regulation is exclusive to deliberation. A viable dual-process model must also include an intuitive, "bottom-up" regulatory mechanism (De Neys, 2012; Evans, 2019; Pennycook, 2017; Stanovich, 2018). However, once System 2 is engaged, individuals can exert additional effort to regulate their thinking. For example, as noted in commentaries on De Neys (2023), deliberative regulation could involve modulating the uncertainty threshold in System 1. If reasoners deliberately lower this threshold, they might begin deliberation sooner (i.e., at lower levels of uncertainty) and persist in deliberation longer (i.e., because more uncertainty reduction is required to drop below the threshold).

Regardless of the specific mechanisms one proposes, it is essential that models of deliberation incorporate this regulatory function. Recognizing that regulation operates at both intuitive and deliberate levels provides a more comprehensive framework for understanding the interplay between System 1 and System 2.

Guiding principles

After having introduced the core functions of the multifunctional System 2, this second section lists basic operating features or principles that any viable, productive model of deliberation should adhere to. I will focus on three issues: 1) different deliberation functions are complimentary and not mutually exclusive, 2) they are non-exclusive with respect to System 1 functioning, and 3) all functions need to be conceived as a collection of shared subprocesses.

Complementarity. Contemporary dual process models that have considered multiple deliberation functions tend to conceive these as mutually exclusive. That is, people either engage in deliberation to control, generate, or justify a response but not a combination of these functions or all three of them. In the taxonomies and flow charts that illustrate the models, there are separate, exclusive paths for each function (e.g., Pennycook et al., 2015; Stanovich, 2009, 2011). Although there might certainly be cases in which only one specific

function gets activated, I see no good empirical or theoretical reason to postulate that this is a necessity or even the norm. Indeed, more often than not, deliberation will involve a combination of multiple functions. For example, one may engage in deliberate response control to refrain from giving a salient intuitive response, then engage in deliberate response generation to compute an alternative response but upon failing to do so, engage in further deliberation to find a justification for the initial intuitive response. Hence, contra mutual exclusivity, complementarity seems to be the more sensible default starting point to build into our deliberation models.

To avoid confusion, this does not imply that deliberation always needs to involve multiple or all functions either. As I noted, there can be cases in which only one function will be activated. For example, when one already generates a correct intuitive response on a classic reasoning task, there may be a need to engage in deliberation for justification but not for control or response generation. The point is simply that it is problematic to uphold a mandatory mutual exclusive operation when conceptualizing the interaction between multiple deliberation functions. A reasoner with a multifunctional System 2 does not necessarily have to choose between either one deliberation function but can engage in multiple functions in the course of a reasoning process.

One way to reconcile this perspective with earlier mutually exclusive models (Pennycook et al., 2015; Stanovich, 2009, 2011) is to recognize that these models primarily focused on a single moment in the deliberation process, rather than its full course. In other words, they specify the possible functions a reasoner might be engaged in at a specific point in time (e.g., System 2 is engaged in either response generation or justification at time x). By acknowledging that a reasoner can iterate through and switch between multiple functions over the time course of a deliberation process, the complementarity principle can be met[1].

Non-exclusivity. There is long-standing debate about whether the difference between intuition and deliberation is qualitative or merely quantitative in nature (e.g., Keren & Shul, 2009; Osman, 2004). In the wake of this debate authors have tried to list defining features of deliberate and intuitive processing (Evans & Stanovich, 2013). These defining features would be exclusive to deliberation and allow one to unequivocally demarcate intuitive and

[1] I am indebted to Gord Pennycook for pointing this out.

deliberate processing. In this context, it is crucial to stress that the core System 2 functions I listed should not—and cannot—be conceived as a defining feature of deliberation. That is, it is not because a reasoner is engaging in response control, generation, justification, or regulation that they are necessarily deliberating. There is clear evidence that any one of these functions can be accomplished by mere intuitive processing too. I already mentioned studies pointing to successful intuitive regulative monitoring (Bago & De Neys, 2017; Buric & Srol, 2020; Johnson et al., 2016; Thompson & Johnson, 2014). This is also obvious for response generation where few would disagree with the point that intuition can generate responses. But this even applies in the more specific case of algorithmic or hypothetical thinking, for example. For algorithmic thinking one may consider habitual scripts (e.g., "Order at restaurant", Bower et al., 1979) in which we routinely follow a sequence of steps (e.g., "Wait to be seated. Sit down. Look at menu. Wait for waiter. Order.") without much deliberate thought. For hypothetical thinking, a simple illustration are cases in which we experience regret (De Neys, 2021). Here we typically engage in counterfactual thinking (e.g., "If only I hadn't betted on black" after bad luck in the casino). We readily start envisaging alternative states of affairs. This process involves mental simulation or hypothetical thinking but it is not cognitively demanding. Experimental studies using load paradigms indicate that people intuitively engage in it and it is precisely refraining from doing so that seems to be demanding (Goldinger et al., 2003).

Likewise, in the last decade, numerous studies in the cognitive control field have established that control can be executed automatically (Abrahamse et al., 2016; Braem et al., 2023; Desender et al., 2013; Jiang et al., 2018; Linzarini et al., 2017; Voudouri et al., 2024). In a recent demonstration, Voudouri et al. (2024) adopted the two-response paradigm referred to above with classic cognitive control tasks such as the Stroop and Flanker. Initial responses needed to be given under stringent time pressure and memorization task load that experimentally minimized possible deliberation. Nevertheless, Voudouri et al. observed that correct final responses were typically already preceded by correct responses in the initial response stage—indicating that the required response control could also be engaged more intuitively.

Even justification does not necessarily require deliberation. For example, as any parent can attest, even young children can readily come up with reasons to justify their behavior in a split-second ("But she hit me first!", "Mom said I could!"). More direct evidence

comes from the two-response justification studies (Bago & De Neys, 2019; Beauvais et al., 2024). In addition to classic reasoning problems, participants in these studies were also given easy, control problems in which the intuitively cued, heuristic response was also correct (e.g., for the bat-and-ball problem such a problem would read "A bat and ball cost $1.10 together. The bat costs $1. How much does the ball cost?"—correct answer: 10 cents). Not surprisingly, even in the initial, intuitive 2-response stage, participants' accuracy was at ceiling on these problems. However, in the vast majority of cases, participants also correctly justified their answer after the initial response stage (e.g., "Well, $1.10 - $1 is 10 cents"). Hence, when response generation does not require complex calculations, justification can be intuitive. Although it is possible that deliberation allows for a more extensive and elaborate justification process, the core process does not seem to be out of reach of the intuitive System 1 (Beauvais et al., 2024; Mercier & Sperber, 2017).

In sum, it is unwarranted and unproductive to conceive the functions that the multi-functional System 2 engages in as being exclusive to deliberation per se. As with the complementarity of functions, non-exclusivity should be the modal norm in our models of deliberation.

Subprocesses. In line with original suggestions by Evans and Stanovich (Evans, 2019; Evans & Stanovich, 2013; Stanovich, 2011), whatever we postulate as core deliberate function(s), it is presumably useful to think of each deliberative function as being comprised of and relying on a set of lower-level subprocesses (e.g., sustained attention, storage, retrieval, etc.). These subprocesses are the shared building blocks of the various deliberation functions. The subprocesses are shared in that all functions depend on and recruit them. For example, in response generation via algorithmic thinking it will be critical to keep track of the results of previous steps while we're executing each new step. Otherwise our conclusions will bound to be biased.

When we're looking for a justification, it will be equally important to store and keep track of previously visited reasons to avoid re-visiting the same reason over and over again. Likewise, attentional control mechanisms will not only be critical for response control to allow us to refrain from readily blurting the first response that comes to mind, but also for response generation via hypothetical thinking. Clearly, whenever we mentally simulate an event, we must be able to prevent our representation of the hypothetical situation to become confused

with our representation of the real world (otherwise we would no longer know what is real and simulated, De Neys, 2021). Keeping the original representation in a sufficiently activated state will recruit attentional control resources. Obviously, storage and attentional control will also be necessary for regulation. If we want to monitor whether a newly generated response or justification conflicts with or contradicts an already generated one, we will need to keep the original response sufficiently activated.

What is important here is to avoid tying either one of the subprocesses exclusively to any specific deliberation function. For example, it is not because a reasoner engages attentional control that this implies they are suppressing a salient intuitive response. They might as well be trying to keep their simulated event from interfering with their representation of the original answer during response generation. Hence, while any one deliberation function should not be tied exclusively to deliberation, any one subprocess should not be uniquely tied to one specific deliberation function. Deliberation entails a combination of multiple functions and multiple shared subprocesses will jointly underlie each function. Figure 2 illustrates this simple idea.

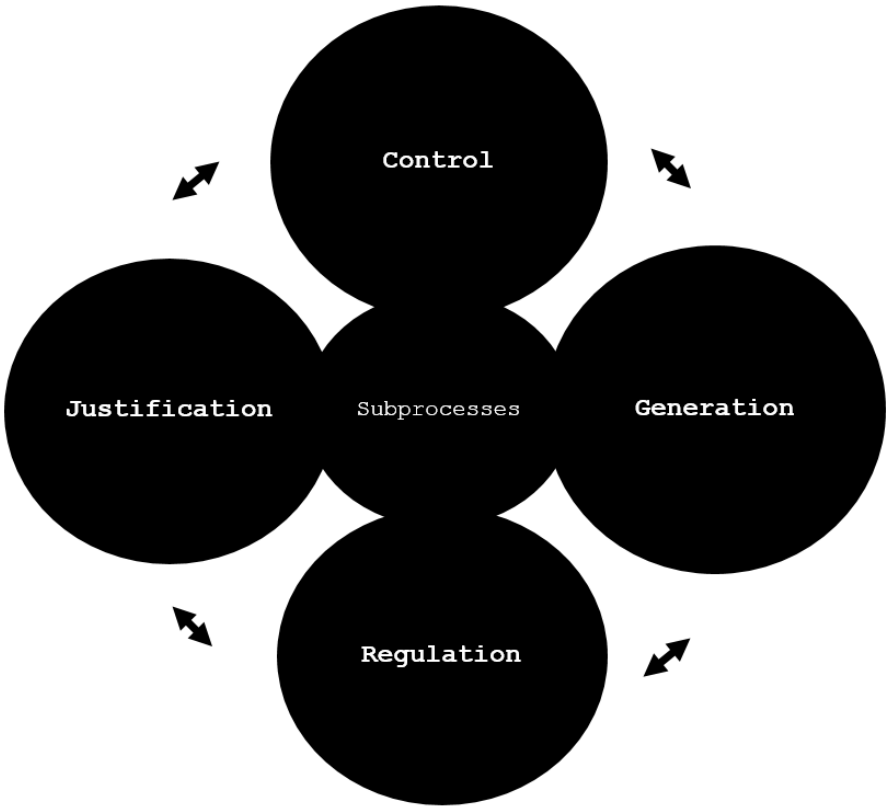

Figure 2. **Conceptual illustration of core deliberation functions**. The node in the center reflects that the functions rely on shared subprocesses. The fact that all function nodes are linked reflects that the functions are not mutually exclusive. Reasoners can engage in and iterate through multiple functions during the course of a deliberation process, as further indicated by the double-headed arrows.

Conclusion

In this paper I tried to present a basic conceptual framework and principles to think of deliberation and avoid problematic misconceptions. The framework was intended as a verbal model of the key features that can help guide the empirical study of deliberation in the coming years. It goes without saying that the framework is a starting point that will need to be further developed. In this last section I point to some intriguing outstanding questions that have hitherto been largely neglected in the literature on deliberation.

One issue is whether the multiple System 2 functions are activated in a specific order or sequence. Evans (2019), for example, suggested that people always start with response justification and will only engage in response generation if response justification fails. Such a

default activation hypothesis model is a viable theoretical option but at present there is no direct empirical evidence to support (or discard) it. However, if there is no fixed default-order, posting multiple System 2 functions also implies that we need a mechanism to explain which function gets activated when deliberation is called upon. That is, how does our reasoning engine know whether to engage in response control, generation, justification, or regulation when we initially decide to deliberate? Is this decision made randomly? One speculative alternative hypothesis is that the choice is determined by the level of experienced uncertainty. I clarified that contemporary dual process models assume that the decision to switch from System 1 to System 2 processing depends on uncertainty monitoring (Ackerman & Thompson, 2017; De Neys, 2023; Pennycook et al., 2015). If the uncertainty crosses a certain threshold, deliberation will be called upon. One could further hypothesize that the level of uncertainty not only determines whether we engage System 2 but also what function we engage in. For example, at lower uncertainty levels we may rather engage justification, higher levels may lead to response control, a further increase may trigger response generation, whereas regulation would only be called upon if uncertainty is maximal. Obviously, this suggestion is purely hypothetical at this point. To my knowledge, there is currently no direct evidence that allows us to decide the issue. This lack of evidence applies even more strongly to intriguing follow-up questions, such as whether the activation order might vary across individuals, specific contexts, or even cultures. It illustrates, however, the value of the conceptual framework that the present paper calls for by pointing to new research questions and shortcomings in our current theorizing. Clearly, if we keep on thinking of deliberation as a monolithic function or mere response control, the switch or activation order question I am pointing out here will remain unexplored.

Another issue is whether the response control function can (or should) be further divided in different subtypes of response control. Currently, the specification focuses on control as inhibitor of a salient intuitive response to give oneself the time for further deliberation and, for example, the generation of a deliberate response. But is this the same control that is needed when after deliberate response generation or justification we realize the initial response is not appropriate? Put differently, is exerting control over an intuitive response to allow further deliberation the same as exerting control over an intuitive response after deliberation? Or does deliberate response generation or justification simply sidesteps the need for further control altogether? That is, imagine that in the bat-and-ball task we have

refrained from immediately responding “10 cents“, engaged in deliberate response generation, and as a result realize that the correct answer is “5 cents”. Does the “10 cents” response automatically lose its appeal at this point or does it continue to require further deliberate effort to control it? Anecdotal evidence suggests it might (e.g., Gould’s famous quote “but she can’t be a bank teller – read the description” about the Linda problem, Gould, 1988 – see also the acquiescence phenomenon, Risen, 2016). Failures of impulse control in addictions also suggest that the deliberate realization of the proper course of activation (e.g., “stop smoking because it is bad for your health”) does not suffice to abandon the intuitive impulse (but see also Meiran et al., 2017). Similarly, learning scientific knowledge does not seem to simply supplant earlier intuitive misconceptions (Brault Foisy et al., 2015; Shtulman & Valcarel, 2012; Shtulman & Young, 2024). But the question has received little systematic attention in case of higher-order reasoning tasks.

This is not only important for theoretical reasons. It has also applied implications for the efficiency of interventions aimed at debiasing or improving people’s reasoning, for example. If response generation is not sufficient for controlling an incorrect intuitive response, then our interventions will not only need to focus on improving deliberate generation (e.g., by teaching/explaining the correct solution strategy) but also on boosting deliberate control capacities. Clearly, my goal here is not to make claims about the optimal design of intervention studies but simply to illustrate why our specific conceptualization of deliberation matters.

Pinpointing the precise interplay between response generation and justification will also need further work. Response justification may be necessary for non-transparent intuitive processes. But since deliberate response generation is by definition transparent, there is no need for an additional justification process in this case. Hence, in theory, deliberate response generation renders response justification obsolete. However, in practice, justification after response generation cannot be excluded. As I already noted, although the reasons behind a deliberately generated response will be clear, we may still engage in justification to look for additional supporting or contradictory reasons. This may explain why people can have multiple changes of mind during deliberation. Recent evidence indeed suggests that in case people change an initial intuitive response after some deliberation, they might nevertheless deliberately switch back to the original response after further reflection (Shivnekar & Srivastava, 2024). For example, in one study Shivnekar and Srivastava (2024) used a variant

of the two-response paradigm in which participants were requested to deliberate for at least one minute after their initial response on a moral reasoning task. The authors observed that in case people changed their initial response after deliberation, they sometimes reverted back to the initial response after further extended deliberation. One explanation is that people initially engage in deliberate response generation to arrive at an alternative response to their initial intuitive hunch but afterwards nevertheless engage in justification of the *initial* response and consequently find it more appealing. In sum, this indicates that more work is needed to delineate generation and justification and their interaction.

As the outstanding issues already indicate, it will be critical to develop a more detailed processing specification of the current verbal model. That is, the model specifies what deliberation is supposed to be doing, what functions it serves. But it does not specify how exactly it does this. For example, we can define justification as the search for explicit reasons, but even if we assume that it relies on executive subprocesses (such as retrieval, storage, etc.) how precisely these are engaged and operate is not clear (Cushman, 2020). Same goes for response generation. The suggestions that it relies on algorithmic or hypothetical thinking are sensible but these remain but high-level verbal characterizations, of course. To date, the current deliberation model (as the more traditional conceptualizations) lacks a more detailed processing account. Such a specification and the development of a proper formal, computational model will obviously be critical for the further development of the field. However, although I readily underwrite the importance of formal model development (Guest & Martin, 2021), I also believe we should not downplay the importance of verbal models. Although this is but a first step in theory development, it remains critical because it determines the kind of questions we will ask and address. For example, as noted above, if we keep on thinking of deliberation as a monolithic function, the question of whether there is a default activation order of the multiple System 2 functions and the other outstanding issues I pointed to will simply not surface.

The list of deliberation functions I discussed here does not necessarily need to be exhaustive. I focused on the most common candidates that have been considered in the literature. In theory, one could explore or conceive additional functions (or further subdivision, see above). However, the key point is that whatever functions one postulates, these will need to respect the guiding “System 2” principles of complementarity and non-exclusivity that I introduced here.

In the introduction I referred to key debates in the dual-process literature such as whether intuition and deliberation are qualitatively distinct or whether they operate in parallel or serially (De Neys, 2012, 2021, 2023; Evans, 2007, 2019; Evans & Stanovich, 2013; Gawronski et al., 2014; Sloman, 1996; Pennycook et al., 2015; Stanovich, 2018). It should be clear that the current perspective and conceptualization is orthogonal to these previous debates or any specific dual-process model instantiation. Whether or not one considers System 1 and 2 to be qualitatively different, whether they operate in parallel or not, the point is that one will still need to specify what deliberation is doing. The question as to how we should think of deliberation is more general in this respect. The present conceptualization is not consequential for these debates per se but it does provide a set of operating principles that any position will need to adhere to if we are to avoid problematic misconceptions in the way we think of deliberation.

It is evident that people sometimes engage in slow, effortful deliberation (Pennycook, 2017). In this perspective, I aimed to outline what occurs during deliberation and highlighted its multiple functions. One further, epistemological question is why people engage in these functions in the first place. I clarified that we can view deliberation as being triggered by uncertainty in System 1. I speculate that people are ultimately motivated to deliberate because its various functions can help to reduce this uncertainty—possibly because uncertainty is experienced as aversive (Kurth, 2023). This does not imply that deliberation will always succeed in reducing uncertainty. For instance, failing to find an explicit justification during deliberation might increase uncertainty. However, I hypothesize that throughout our life experiences we have learned that more often than not deliberation is helpful here. This framing also suggests that while System 1 can engage in uncertainty-reducing functions (consistent with the principle of non-exclusivity), deliberation offers additional advantages. For example, deliberation may enhance success by enabling individuals to track more justifications or responses over a longer time frame. But these assumptions remain speculative and are currently not backed up by empirical evidence. Nevertheless, they underscore the need for a new conceptual perspective. If we continue to view deliberation functions as exclusive, there is little point in conducting studies that compare the efficacy of System 1 and System 2 justification, for example. I believe that it are precisely these sort of newly emerging questions that hold significant promise for advancing the field.

This perspective focused on the role of deliberation in human thinking. However, the conceptualization of deliberation will be equally critical for the development of machine thinking accounts in Artificial Intelligence (AI) research (Bonnefon & Rahwan, 2020; Rahwan et al., 2019). Current work is already applying the dual process framework in AI system design and analysis (Bonnefon & Rahwan, 2020; Hagendorff et al. 2023; Nye et al., 2021; Yax et al., 2024). However, just as in research on human thinking, the conceptualization of what "System 2" precisely entails is not clear. This is especially critical for the field of so-called "explainable AI" (Linardatos et al., 2020). That is, despite the impressive performance of Large Language Models and other AI systems, the inner workings of these systems remain a black box. When AI systems generate answers or provide advice, we often have no insight into how these responses were generated. This poses critical ethical, legal, and application issues (e.g., on whether people should and will trust AI recommendations, Myers & Chater, 2024). Hence, not unlike an intuitive human thinker, it will be critical to have AI systems generate explicit justifications for non-transparent "System 1" computations (Linardatos et al., 2020; Myers & Chater, 2024). In this sense, the need for a clearer conceptualization of System 2 deliberation will be essential for advancing the study of thinking in our machine counterparts as well.

In this context, my hope is that the model and guidelines I sketched will help to debunk problematic thinking about deliberation and boost the study of this core process in the coming years.

ACKNOWLEDGMENTS

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 author 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.

I would like to thank Zoe Purcell, Gord Pennycook, Mirsoslav Sirota, and an anonymous reviewer for valuable comments on a previous version of this manuscript.

REFERENCES

Abrahamse, E., Braem, S., Notebaert, W., & Verguts, T. (2016). Grounding cognitive control in associative learning. Psychological Bulletin, 142, 693.

Ackerman, R., & Thompson, V. A. (2017). Meta-reasoning: Monitoring and control of thinking and reasoning. Trends in cognitive sciences, 21, 607-617.

Bago, B., & De Neys, W. (2017). Fast logic?: Examining the time course assumption of dual process theory. Cognition, 158, 90-109.

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, 3, 257-299.

Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. *Current Directions in Psychological Science, 16(6)*, 351-355.

Beauvais, N., Voudouri, A., Boissin, E., & De Neys, W. (2020). System 2 and cognitive transparency: Deliberation helps to justify sound intuitions during reasoning. Manuscript submitted for publication.

Betsch, T., & Glöckner, A. (2010). Intuition in judgment and decision making: Extensive thinking without effort. Psychological Inquiry, 21(4), 279-294.

Bonnefon, J. F. (2018). The pros and cons of identifying critical thinking with system 2 processing. Topoi, 37, 113-119.

Bonnefon, J. F., & Rahwan, I. (2020). Machine thinking, fast and slow. Trends in Cognitive Sciences, 24(12), 1019-1027.

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. Cognitive psychology, 11(2), 177-220.

Braem, S., Held, L., Shenhav, A., & Frömer, R. (2023). Learning how to reason and deciding when to decide. Behavioral and Brain Sciences, 46, e115.

Brault Foisy, L. M. B., Potvin, P., Riopel, M., & Masson, S. (2015). Is inhibition involved in overcoming a common physics misconception in mechanics?. Trends in Neuroscience and Education, 4, 26-36.

Burič, R., & Konrádová, Ľ. (2021). Mindware instantiation as a predictor of logical intuitions in the Cognitive Reflection Test. Studia Psychologica, 63, 114-128.

Burič, R., & Šrol, J. (2020). Individual differences in logical intuitions on reasoning problems presented under two-response paradigm. Journal of Cognitive Psychology, 32, 460-477.

Carruthers, P. (2011). The opacity of mind: an integrative theory of self-knowledge. Oxford University Press.

Chater, N. (2018). The mind is flat: the illusion of mental depth and the improvised mind. Penguin UK.

Cushman, F. (2020). Rationalization is rational. *Behavioral and Brain Sciences, 43*, e28.

Dennett, D. C. (1989). The intentional stance. MIT Press.

De Neys, W. (2012). Bias and conflict: A case for logical intuitions. Perspectives on Psychological Science, 7, 28-38.

De Neys, W. (2021). On dual and single process models of thinking. Perspectives on Psychological Science, 16, 1412-1427.

De Neys, W. (2023). Advancing theorizing about fast-and-slow thinking. Behavioral and Brain Sciences, 46, E111.

De Neys, W., & Bonnefon, J. F. (2013). The whys and whens of individual differences in thinking biases. Trends in Cognitive Sciences, 17, 172-178.

Desender, K., Van Lierde, E., Van den Bussche, E. (2013). Comparing conscious and unconscious conflict adaptation. PLoS ONE 8(2), e55976.

Dujmović, M., Valerjev, P., & Bajšanski, I. (2021). The role of representativeness in reasoning and metacognitive processes: an in-depth analysis of the Linda problem. Thinking & Reasoning, 27, 161-186.

Draheim, C., Pak, R., Draheim, A. A., & Engle, R. W. (2022). The role of attention control in complex real-world tasks. *Psychonomic Bulletin & Review, 29(4)*, 1143–1197

Evans, J. St. B. T. (2007). *Hypothetical thinking: Dual processes in reasoning and judgement*. Hove, UK: Psychology Press.

Evans, J. St. B. T. (2012). Dual process theories of deductive reasoning: facts and fallacies., In K. Holyoak & R. G. Morrison (Eds.), The Oxford handbook of thinking and reasoning (pp 115-133). New York: Oxford University Press.

Evans, J. St. B. T. (2019). Reflections on reflection: the nature and function of Type 2 processes in dual-process theories of reasoning. Thinking & Reasoning, 25, 383-415.

Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition advancing the debate. Perspectives on Psychological Science, 8, 223–241.

Evans, J. St. B., & Wason, P. C. (1976). Rationalization in a reasoning task. *British Journal of Psychology, 67,* 479-486.

Frederick, S. (2005). Cognitive reflection and decision making. Journal of Economic perspectives, 19, 25-42.

Gawronski, B., Sherman, J. W., & Trope, Y. (2014). Two of what? A conceptual analysis of dual-process theories, In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.). Dual-process theories of the social mind (pp. 3-19.) New York, NY: Guilford Press.

Ghasemi, O., Handley, S., Howarth, S., Newman, I. R., & Thompson, V. A. (2022). Logical intuition is not really about logic. Journal of Experimental Psychology: General.

Gigerenzer, G. (2007). Gut feelings: the intelligence of the unconscious. New York: Viking.

Goldinger, S. D., Kleider, H. M., Azuma, T., & Beike, D. R. (2003). "Blaming the victim" under memory load. Psychological Science, 14(1), 81-85.

Gould, S. J. (1988). The streak of streaks. Retrieved from https://www.nybooks.com/articles/1988/08/18/the-streak-of-streaks/

Guest, O., & Martin, A. E. (2021). How computational modeling can force theory building in psychological science. *Perspectives on Psychological Science, 16(4)*, 789–802.

Hagendorff, T., Fabi, S., & Kosinski, M. (2023). Human-like intuitive behavior and reasoning biases emerged in large language models but disappeared in ChatGPT. Nature Computational Science, 3(10), 833-838.

Hofmann, W., Friese, M., & Strack, F. (2009). Impulse and self-control from a dual-systems perspective. *Perspectives on Psychological Science, 4(2),* 162-176.

Houdé, O. (2019). 3-system Theory of the Cognitive Brain: A Post-Piagetian Approach to Cognitive Development. Oxon, UK: Routledge.

Houdé, O., & Borst, G. (2015). Evidence for an inhibitory-control theory of the reasoning brain. Frontiers in human neuroscience, 9, 122116.

Jiang, J., Correa, C. M., Geerts, J., & van Gaal, S. (2018). The relationship between conflict awareness and behavioral and oscillatory signatures of immediate and delayed cognitive control. NeuroImage, 177, 11-19.

Johnson, E. D., Tubau, E., & De Neys, W. (2016). The doubting System 1: Evidence for automatic substitution sensitivity. Acta Psychologica, 164, 56-64.

Kahneman, D. (2011). Thinking, Fast and Slow. New York, NY: Farrar, Straus and Giroux.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin and D. Kahneman (Eds), *Heuristics of*

Intuitive Judgment: Extensions and Applications (pp. 49-81). New York: Cambridge University Press.

Keren, G., & Schul, Y. (2009). Two is not always better than one: A critical evaluation of two-system theories. Perspectives on Psychological Science, 4, 533-550.

Knight, E. L., McShane, B. B., Kutlikova, H. H., Morales, P. J., Christian, C. B., Harbaugh, W. T., Mayr, U., Ortiz, T. L., Gilbert, K., Ma-Kellams, C., Riečanský, I., Watson, N. V., Eisenegger, C., Lamm, C., Mehta, P. H., & Carré, J. M. (2020). Weak and Variable Effects of Exogenous Testosterone on Cognitive Reflection Test Performance in Three Experiments: Commentary on Nave, Nadler, Zava, and Camerer (2017). Psychological Science, 31(7), 890-897.

Koehler, D. J. (1991). Explanation, imagination, and confidence in judgment. Psychological bulletin, 110, 499-519.

Kurth, C. (2023). Why is system 1/system 2 switching affectively loaded?. Behavioral and Brain Sciences, 46, e128.

Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberate judgments are based on common principles. Psychological review, 118(1), 97-109.

Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable AI: A review of machine learning interpretability methods. Entropy, 23(1).

Linzarini, A., Houdé, O., & Borst, G. (2017). Cognitive control outside of conscious awareness. Consciousness and cognition, 53, 185-193.

Marewski, J. N., & Hoffrage, U. (2015). Modeling and aiding intuition in organizational decision making. Journal of Applied Research in Memory and Cognition, 4, 145–311.

Martire, K. A., Robson, S. G., Drew, M., Nicholls, K., & Faasse, K. (2023). Thinking false and slow: Implausible beliefs and the Cognitive Reflection Test. Psychonomic Bulletin & Review, 30(6), 2387-2396.

Mata, A., Ferreira, M. B., Voss, A., & Kollei, T. (2017). Seeing the conflict: An attentional account of reasoning errors. Psychonomic Bulletin & Review, 24, 1980-1986.

Mega, L. F., & Volz, K. G. (2014). Thinking about thinking: implications of the introspective error for default-interventionist type models of dual processes. Frontiers in Psychology, 5, 864.

Meiran, N., Liefooghe, B., & De Houwer, J. (2017). Powerful instructions: automaticity without practice. *Current Directions in Psychological Science, 26(6)*, 509-514.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. Behavioral and brain sciences, 34, 57-74.

Mercier, H., & Sperber, D. (2017). The Enigma of Reasoning. Cambridge, MA: Harvard University Press.

Meyer, A., & Frederick, S. (2023). The formation and revision of intuitions. *Cognition, 240*, 105380.

Morewedge, C. K., & Kahneman, D. (2010). Associative processes in intuitive judgment. Trends in cognitive sciences, 14(10), 435-440.

Myers, S., & Chater, N. (2024). Interactive explainability: Black boxes, mutual understanding and what it would really mean for AI systems to be as explainable as people. Manuscript submitted for publication.

Nave, G., Nadler, A., Zava, D., & Camerer, C. (2017). Single-dose testosterone administration impairs cognitive reflection in men. Psychological science, 28(10), 1398-1407.

Nye, M., Tessler, M., Tenenbaum, J., & Lake, B. M. (2021). Improving coherence and consistency in neural sequence models with dual-system, neuro-symbolic reasoning. Advances in Neural Information Processing Systems, 34, 25192-25204.

Osman, M. (2004). An evaluation of dual-process theories of reasoning. Psychonomic Bulletin & Review, 11, 988-1010.

Pennycook, G. (2017). A perspective on the theoretical foundation of dual-process models. In W. De Neys (Ed.), Dual Process Theory 2.0 (pp. 5-27). Oxon, UK: Routledge.

Pennycook, G. (2023). A framework for understanding reasoning errors: From fake news to climate change and beyond. In Advances in experimental social psychology (Vol. 67, pp. 131-208). Academic Press.

Pennycook, G., Cheyne, J. A., Koehler, D. J., & Fugelsang, J. A. (2016). Is the cognitive reflection test a measure of both reflection and intuition?. *Behavior research methods, 48*, 341-348.

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.

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J. F., Breazeal, C., ... & Wellman, M. (2019). Machine behaviour. Nature, 568(7753), 477-486.

Raoelison, M., Thompson, V., & De Neys, W. (2020). The smart intuitor: Cognitive capacity predicts intuitive rather than deliberate thinking. Cognition, 204, 104381.

Reber, A., & Allen, R. (2022). The Cognitive Unconscious: The First Fifty Years. Oxford, UK: Oxford University Press.

Risen, J. L. (2016). Believing what we do not believe: Acquiescence to superstitious beliefs and other powerful intuitions. *Psychological review, 123(2)*, 182.

Sirota, M., Juanchich, M., & Holford, D. L. (2023). Rationally irrational: When people do not correct their reasoning errors even if they could. Journal of Experimental Psychology: General, 152, 2052-2073.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. Psychological Bulletin, 119(1), 3.

Shivnekar, R. V., & Srivastava, N. (2024). Measuring vacillations in reasoning. *Judgment and Decision Making, 19*, e15.

Stanovich, K. E. (2009). Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory. In J. Evans & Frankish, L. (Eds.), *In two minds: Dual processes and beyond* (pp. 55-88). Oxford University Press.

Stanovich, K. E. (2011). Rationality and the reflective mind. Oxford, UK: Oxford University Press.

Stanovich, K. E. (2005). The robot's rebellion: Finding meaning in the age of Darwin. Chicago, IL: University of Chicago press.

Stanovich, K. E. (2018). Miserliness in human cognition: the interaction of detection, override and mindware. Thinking & Reasoning, 24, 423-444.

Stanovich, K. E., & Toplak, M. E. (2023). A good architecture for fast and slow thinking, but exclusivity is exclusively in the past. Behavioral and Brain Sciences, 46, e142.

Shtulman, A., & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. Cognition, 124, 209-215.

Shtulman, A., & Young, A. G. (2024). Tempering the tension between science and intuition. Cognition, 243, 105680.

Stupple, E. J., Pitchford, M., Ball, L. J., Hunt, T. E., & Steel, R. (2017). Slower is not always better: Response-time evidence clarifies the limited role of miserly information processing in the Cognitive Reflection Test. PloS one, 12(11), e0186404.

Shynkaruk, J. M., & Thompson, V. A. (2006). Confidence and accuracy in deductive reasoning. Memory & cognition, 34(3), 619-632.

Thompson, V. A. (2014). What intuitions are… and are not. Psychology of learning and motivation, 60, 35-75.

Thompson, V. A., & Johnson, S. C. (2014). Conflict, metacognition, and analytic thinking. Thinking & Reasoning, 20(2), 215-244.

Thompson, V., & Newman, I. R. (2017). Logical intuitions and other conundra for dual process theories. In W. De Neys (Ed.), Dual Process Theory 2.0 (pp. 121-136). Oxon, UK: Routledge.

Thompson, V. A., Turner, J. A. P., & Pennycook, G. (2011). Intuition, reason, and metacognition. Cognitive Psychology, 63, 107–140.

Volz, K. G., & Zander, T. (2014). Primed for intuition?. Neuroscience of Decision Making, 1, 26-34.

Voudouri, A., Bialek, M., & De Neys, W. (2024). Fast & slow decisions under risk: Intuition rather than deliberation drives advantageous choices. Cognition, 250, 105837.

Wason, P. C., & Evans, J. S. B. (1975). Dual processes in reasoning?. *Cognition, 3*, 141-154.

Yax, N., Anlló, H., & Palminteri, S. (2024). Studying and improving reasoning in humans and machines. Communications Psychology, 2(1), 51.