Behavioral and Brain Sciences

cambridge.org/bbs

Advancing theorizing about fast-and-slow thinking

Wim De Neys

LaPsyDE, CNRS, Université Paris Cité, Paris, France
wim.de-neys@u-paris.fr

Target Article

Cite this article: De Neys W. (2023) Advancing theorizing about fast-and-slow thinking. *Behavioral and Brain Sciences* **46**, e111: 1–71. doi:10.1017/S0140525X2200142X

Target Article Accepted: 13 July 2022
Target Article Manuscript Online: 2 September 2022
Commentaries Accepted: 5 December 2022

Keywords:
decision making; deliberation; dual process; heuristics; intuition; reasoning; thinking

Abstract

Human reasoning is often conceived as an interplay between a more intuitive and deliberate thought process. In the last 50 years, influential fast-and-slow dual-process models that capitalize on this distinction have been used to account for numerous phenomena – from logical reasoning biases, over prosocial behavior, to moral decision making. The present paper clarifies that despite the popularity, critical assumptions are poorly conceived. My critique focuses on two interconnected foundational issues: the exclusivity and switch feature. The exclusivity feature refers to the tendency to conceive intuition and deliberation as generating unique responses such that one type of response is assumed to be beyond the capability of the fast-intuitive processing mode. I review the empirical evidence in key fields and show that there is no solid ground for such exclusivity. The switch feature concerns the mechanism by which a reasoner can decide to shift between more intuitive and deliberate processing. I present an overview of leading switch accounts and show that they are conceptually problematic – precisely because they presuppose exclusivity. I build on these insights to sketch the groundwork for a more viable dual-process architecture and illustrate how it can set a new research agenda to advance the field in the coming years.

Sometimes thinking can be hard. As majestically portrayed in Rodin's "The Thinker" sculpture, in these cases it will take us laborious inferencing to arrive at a problem solution. At other times, however, thinking can be surprisingly easy. If you ask an educated adult how much half of $100 is, in what city the Statue of Liberty is located, or whether a toddler should be allowed to drink beer, they can answer in a split second. At least since antiquity, such duality in our mental experiences has led to the idea that there are two types of thinking, one that is fast and effortless, and one that is slower and requires more effort (Frankish & Evans, 2009; Pennycook, 2017). This distinction between what is often referred to as a more intuitive and deliberate mode of cognitive processing – or the nowadays more popular "system 1" and "system 2" labels – lies at the heart of the influential "fast-and-slow" dual-process view that has been prominent in research on human thinking since the 1960s (Evans, 2008; Kahneman, 2011).

It is presumably hard to overestimate the popularity of dual-process models in current-day psychology, economics, philosophy, and related disciplines (Chater & Schwarzlose, 2016; Melnikoff & Bargh, 2018). As De Neys (2021) clarified, they have been applied in a very wide range of fields including research on thinking biases (Evans, 2002; Kahneman, 2011), morality (Greene & Haidt, 2002), human cooperation (Rand, Greene, & Nowak, 2012), religiosity (Gervais & Norenzayan, 2012), social cognition (Chaiken & Trope, 1999), management science (Achtziger & Alós-Ferrer, 2014), medical diagnosis (Djulbegovic, Hozo, Beckstead, Tsalatsanis, & Pauker, 2012), time perception (Hoerl & McCormack, 2019), health behavior (Hofmann, Friese, & Wiers, 2008), theory of mind (Wiesmann, Friederici, Singer, & Steinbeis, 2020), intelligence (Kaufman, 2011), creativity (Barr, Pennycook, Stolz, & Fugelsang, 2015), fake news susceptibility (Bago, Rand, & Pennycook, 2020), and even machine thinking (Bonnefon & Rahwan, 2020). In addition, the dual-process framework is regularly featured in the popular media (Lemir, 2021; Shefrin, 2013; Tett, 2021) and has inspired policy recommendations on topics ranging from economic development (World Bank Group, 2015), over carbon emissions (Beattie, 2012), to the corona-virus pandemic (Sunstein, 2020).

The present paper tries to clarify that despite the popularity, a lot of the current day use of dual-process models is poorly conceived. Foundational assumptions are empirically questionable and/or conceptually problematic. I argue that a core underlying problem is the exclusivity feature or the tendency to conceive intuition and deliberation as generating unique responses such that one type of response is exclusively tied to deliberation and is assumed to be beyond the reach of the intuitive system. For example, influential dual-process accounts of biases in logical reasoning rely on exclusivity when attributing flawed thinking to a failure to correct an intuitively generated response with a deliberate response (Evans & Stanovich, 2013;

Kahneman, 2011). Likewise, dual-process accounts of moral and prosocial reasoning rely on it to explain how intuitive emotional responses prevent us from taking the consequences of our actions into account (e.g., Greene, 2013; Greene & Haidt, 2002) or to clarify why people behave selfishly rather than cooperate (e.g., Rand et al., 2012). In section 1 I review the empirical evidence in key fields and will show that although the exclusivity assumption might be appealing, there is no solid ground for it.

In section 2, I focus on a conceptual consequence of the exclusivity feature. Any dual-process model needs a switch mechanism that allows us to shift between intuitive and deliberate processing. Given that we can use two types of reasoning, there might be cases in which either one will be more or less beneficial. But how do we know that we can rely on an intuitively cued problem solution or need to engage in costly further deliberation? And when do we switch back to the intuitive processing mode once we start deliberating? I review popular traditional dual-process accounts for the switch issue and show that they are conceptually problematic – precisely because they presuppose exclusivity. In section 3, I build on this insight and recent theoretical advances to sketch a more viable general dual-process architecture that can serve as theoretical groundwork to build future dual-process models in various fields. Finally, in the closing section, I use the model to identify new and outstanding questions that should advance the field in the coming years.

Before moving to the main sections, it might be a good idea to clarify my use of the nomenclature. I adopt the fast-and-slow dual-process label as a general header to refer to models that posit an interaction between intuitive and deliberate reasoning processes. Dual-process theories are sometimes opposed to single-model theories. Both single- and dual-process theories focus on the interaction between intuition and deliberation. But they differ concerning the question as to whether the difference between the two types of processing should be conceived as merely quantitative or qualitative in nature (target article Introduction; see De Neys, 2021, for a recent review). My argument here is completely orthogonal to this issue (see sect. 4.8). My criticism and recommendations equally apply to single- and dual-process models. I stick to the dual-process label simply because it is more widely adopted.

There are also a wide range of labels that are being used to refer to the two types of reasoning that are posited by dual-process models (e.g., type 1/2, system 1/2, heuristic/analytic thinking, associative/rule-based thinking, automatic/reflective, intuitive/deliberate, etc.). I will stick here to the traditional labels "intuitive" and "deliberate" processing as well as the nowadays more popular "system 1" and "system 2" processing. The system term can sometimes refer to a specific subtype of dual-process models (Gawronski & Creighton, 2013). Here it is used in a generic, general sense. As in Kahneman (2011), system 1 and 2 processing can be interpreted as synonyms for the type of effortless intuiting and effortful deliberating that are traditionally contrasted in dual-process theories.

Wim De Neys is a CNRS research director affiliated with the LaPsyDE lab at the Université Paris Cité (Paris, France). His work focuses on the interaction between intuition and deliberation in the reasoning and decision-making field. He has published over 100 peer-reviewed articles.

1. Exclusivity in dual-process models

As briefly introduced, the exclusivity feature refers to the tendency to associate intuitive and deliberate processing with the computation of unique responses. System 1 is believed to be responsible for generating a response X and system 2 is responsible for generating an alternative response Y. Critically, here is the underlying exclusivity: Generation of the alleged deliberate response is by definition believed to be beyond the capacity of the intuitive system 1.

This simple exclusive dichotomization is appealing. System 1 quickly provides us with one type of response. If we want to generate the alternative response, we will necessarily need to switch to effortful deliberation. By combining this with the human tendency to minimize cognitive effort ("cognitive miserliness," e.g., Stanovich & West, 2000) one has a seemingly simple account of a wide range of mental processes. To illustrate, below I sketch in more detail how popular dual-process models in various fields are relying on the exclusivity assumption. I focus on the dual-process model of logical, moral, and prosocial reasoning because these have been among the most influential applications and allow me to demonstrate the generality of the findings. I present a brief introduction of the paradigmatic model in each field and then move to a discussion of the empirical evidence.

1.1 Logical, moral, and prosocial dual-process exclusivity

1.1.1 Logical reasoning bias

One of the first fields in the cognitive sciences in which dual-process models were popularized is research on "biases" in logical reasoning (e.g., Evans, 2016; Kahneman, 2000, 2011; Wason, 1960; Wason & Evans, 1975). Since the 1960s numerous studies started showing that people readily violate the most elementary logical, mathematical, or probabilistic rules when a task cues an intuitive response that conflicts with these principles.[1]

For example, imagine that we have two trays with red and white marbles. There's a small tray with 10 marbles of which one is red. There is also a large tray holding 100 marbles of which nine are red. You can draw one marble from either one of the trays. If the marble is red, you win a nice prize. Which tray should you draw from to maximize your chances of winning? From a logical point of view, it is clear that the small tray gives you a 10% chance of drawing a red marble (1/10) whereas the large tray gives you only a 9% (9/100) chance. However, people often prefer to draw from the large tray because they intuitively tend to use the absolute number of red marbles as a shortcut or "heuristic" to guide their inferences (Epstein, 1994). Obviously, there are indeed more red marbles in the large tray than in the small tray (i.e., nine vs. one). In case there would be the same number of white marbles in both trays, the simple absolute number focus would lead to a correct judgment. However, in the problem in question, there are also a lot more white marbles in the large tray. If you take the ratio of red and white marbles into account it is crisp clear that you need to draw from the small tray. Unfortunately, the available evidence suggests that in situations in which an intuitive association cues a response that conflicts with more logical considerations (e.g., the role of denominators or ratios), people seem to neglect the logical principle and opt for the intuitively cued conclusion (Kahneman, 2011).[2] Hence, our intuitions often seem to lead us astray and bias our judgment.

The dual-process framework presents a simple and elegant explanation for the bias phenomenon (Evans, 2008; Kahneman,

https://doi.org/10.1017/S0140525X2200142X Published online by Cambridge University Press

2011). In general, dual-process theorists have traditionally highlighted that taking logical principles into account typically requires demanding system 2 deliberation (e.g., Evans, 2002, 2008; Evans & Over, 1996; Kahneman, 2011; Stanovich & West, 2000). Because human reasoners have a strong tendency to minimize demanding computations, they will often refrain from engaging or completing the slow deliberate processing when mere intuitive processing has already cued a response (Evans & Stanovich, 2013; Kahneman, 2011). Consequently, most reasoners will simply stick to the intuitive response that quickly came to mind and fail to consider the logical implications. It will only be the few reasoners who have sufficient resources and motivation to complete the deliberate computations and override the initially generated intuitive response, who will manage to reason correctly and give the logical response (Stanovich & West, 2000).

This illustrates how the bias account critically relies on the exclusivity assumption. Taking logical principles in classic reasoning tasks into account is uniquely linked to deliberation. Because this is out of reach of the intuitive system, sound reasoning will require us to switch from system 1 to demanding system 2 processing – something that few will manage to accomplish. To avoid confusion, it is important to stress here that the exclusivity assumption does not entail that system 1 is always biased and system 2 always leads to correct answers. Dual-process theorists have long argued against such a simplification (Evans, 2011; Evans & Stanovich, 2013). Clearly, nobody will disagree that educated adults can intuitively solve a problem such as "Is 9 more than 1?" or "How much is 2 + 2?." The hypothesis concerns situations in which the two systems are assumed to be generating conflicting responses. More generally, as any scientific theory, dual-process models make their assertions within a specific application context. For the dual-process model of logical reasoning, the application context concerns situations in which an intuitively cued problem solution conflicts with a logico-mathematical norm. The classic "heuristics and biases" tasks in the field (such as the earlier ratio bias problem with the two trays) all capitalize on such conflict and are designed such that they cue a salient conflicting intuitive heuristic response that is pitted against logical considerations. It is in such conflict cases that avoiding biased thinking is expected to require switching to system 2 deliberation.

1.1.2 Dual-process model of moral reasoning

The influential dual-process model of moral cognition focuses on situations in which utilitarian and deontological considerations lead to conflicting moral judgments (e.g., is it acceptable to sacrifice one human life to save five others?). From a utilitarian point of view, one focuses on the consequences of an action. Harming an individual can be judged acceptable if it prevents comparable harm to a greater number of people. One performs a cost–benefit analysis and chooses the greater good. Hence, from a utilitarian perspective, it can be morally acceptable to sacrifice someone's life to save others. Alternatively, the moral perspective of deontology focuses on the intrinsic nature of an action. Here harming someone is considered wrong regardless of its potential benefits. From a deontological point of view, sacrificing one life to save others is never acceptable. In a nutshell, the dual-process model of moral reasoning (Greene, 2013; Greene & Haidt, 2002) has associated utilitarian judgments with deliberate system 2 processing and deontological judgments with intuitive system 1 processing. The core idea is that giving a utilitarian response to moral dilemmas requires that one engages in system 2 thinking and allocates cognitive resources to override an intuitively cued system 1 response that primes us not to harm others (Greene, 2007; Paxton, Ungar, & Greene, 2012). Hence, here too the exclusivity assumption is key: Utilitarian reasoning is assumed to be out of reach of the intuitive system and requires a switch to costly effortful processing.

1.1.3 Dual-process model of prosocial reasoning

Finally, the dual-process model of prosocial reasoning or human cooperation focuses on situations in which self-interest can conflict with the group interest (e.g., get more money yourself or share more with others). Some authors have claimed that making prosocial choices requires deliberate system 2 control of our intuitive selfish impulses (e.g., DeWall, Baumeister, Gailliot, & Maner, 2008; Knoch, Pascual-Leone, Meyer, Treyer, & Fehr, 2006; Martinsson, Myrseth, & Wollbrant, 2014). Alternatively, others have argued that system 1 cues prosocial choices and it is only after deliberation that we will seek to maximize our self-interest (e.g., Rand, 2019; Rand et al., 2012; Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003). However, despite the differences concerning which behavior is assumed to be favored by deliberation and intuition, both views are built on the same underlying exclusive dual-process logic: Intuition will favor one type of behavior whereas making the competing choice will require slow, deliberate system 2 processes to control and correct the initial intuitive impulse (Hackel, Wills, & Van Bavel, 2020; Isler, Yilmaz, & Maule, 2021).

To be clear, just like the dual-process model of logical reasoning, dual-process models of prosocial (and moral) reasoning also have a specific application context. As with logical reasoning, this context concerns prototypical cases in which the two systems are assumed to be generating conflicting responses. For example, dual-process models of prosocial choice focus on anonymous decision settings (i.e., the identity of the decision maker and recipient are never revealed and they only interact one single time, e.g., Rand et al., 2012). Clearly, even models that posit that prosocial (vs. selfish) decisions require system 2 processing would not dispute that the prosocial decision to share with one's offspring, for example, can be made completely intuitively. Similarly, the dual-process model of moral reasoning focuses on moral dilemmas that cue a strong moral transgression (e.g., killing). In some cases, the deontological option might be so trivial (e.g., is it acceptable to tell a white lie to save five people?) that it will not give rise to a proper conflict. In these non-conflict cases it would not be expected that a utilitarian judgment necessarily requires system 2 processing (Greene, Sommerville, Nystrom, Darley, & Cohen, 2001).

Note that the empirical evidence I will review in the following section always concerns the prototypical test and application context that the dual-process models traditionally envisaged. The fundamental problem I will raise is that – in contrast to widely publicized initial reports – even in these cherished prototypical contexts there is no solid empirical ground for the exclusivity assumption. For completeness, I start by discussing the traditional evidence that has been cited in support of the exclusivity assumption and then move on to a discussion of more recent counter-evidence.

1.2 Empirical evidence

Why have dual-process models ever assumed the exclusivity feature? What empirical evidence was there to support it? The undisputed starting point here is that deliberation is defined as being more time and effort-demanding than system 1 processing. Hence, if the exclusivity assumption holds, one would expect that the alleged system 2 response will take longer than the intuitive system 1 response. Likewise, generation of the alleged system 2 response should be more likely among those higher in cognitive capacity (and motivation to use this capacity). This would be consistent with the idea that the alleged system 2 response indeed requires slow, effortful deliberation. The introduction of traditional dual-process models in various fields has typically been accompanied by correlational studies that supported these predictions (Trémolière, De Neys, & Bonnefon, 2019). For example, from logical, over moral, to prosocial reasoning, various studies showed that people who give the alleged deliberate response indeed tend to take more time to answer and score higher on standard cognitive ability/disposition tests than people who give the alleged intuitive response (e.g., De Neys, 2006a, 2006b; Greene et al., 2001; Moore, Clark, & Kane, 2008; Paxton et al., 2012; Rand et al., 2012; Stanovich & West, 1998, 2000).

In addition to correlational studies, dual-process proponents have also pointed to experimental evidence coming from cognitive constraint paradigms in which people are forced to respond under time-pressure or secondary cognitive load (e.g., concurrent memorization). The rationale here is again that deliberation requires more time and cognitive resources than system 1 processing. Consequently, depriving people of these resources by forcing them to respond quickly or while they are performing a capacity demanding secondary task, should make it less likely that the exclusive system 2 response can be generated. Across logical, moral, and prosocial reasoning studies, dual-process proponents have indeed shown that these constraints often hinder the production of the alleged deliberate responses (e.g., Conway & Gawronski, 2013; De Neys, 2006b; Evans & Curtis-Holmes, 2005; Rand et al., 2012, 2014; Trémolière, De Neys, & Bonnefon, 2012). In sum, the point of this short overview is that dual-process theorists have not made their claims in an empirical vacuum. There are past findings that are consistent with the exclusivity assumption.

However, a first problem is that over the years these initial positive findings have not always been confirmed. Recent studies and large-scale replication efforts have pointed to negative findings and null-effects (e.g., Baron, 2017; Baron & Gürçay, 2017; Białek & De Neys, 2017; Bouwmeester et al., 2017; Grossman & Van der Weele, 2017; Gürçay & Baron, 2017; Robison & Unsworth, 2017; Tinghög et al., 2016). Available meta-analyses suggest that if there is an effect, it is very small. For example, Rand (2019) found that experimental manipulations that limited deliberation (and/or favored intuition) led on average to an increase of 3.1% prosocial choices (see also Kvarven et al., 2020). Likewise, in one of the largest studies to date on reasoning bias, Lawson, Larrick, and Soll (2020) found that experimental constraints on a wide range of classic bias tasks led on average to a 9.4% performance decrease (from 62 to 52% accuracy). As Lawson et al. put it, this suggests that the alleged deliberate response can often be generated intuitively. Even when deliberation is prevented, the alleged deliberate response is still frequently observed. Hence, although there is indeed some evidence that deliberation pushes responses in the expected dual-process direction (e.g., more alleged system 2 responses) it is becoming clear – contra the exclusivity assumption – that generation of the alleged unique system 2 response does often not require deliberation and is not uniquely tied to system 2.

Critically, studies adopting new experimental paradigms have presented further direct evidence against the exclusivity assumption (De Neys & Pennycook, 2019). Perhaps most illustrative are studies with the two-response paradigm (Thompson, Turner, & Pennycook, 2011). In this paradigm, participants are asked to give two consecutive answers to a problem. First, they have to answer as quickly as possible with the first response that comes to mind. Immediately afterward, they are shown the problem again and can take all the time they want to reflect on it and give a final answer. To make maximally sure that the initial answer is generated intuitively, it typically has to be generated under time-pressure and/or cognitive load (Bago & De Neys, 2017; Newman, Gibb, & Thompson, 2017). As with the cognitive constraint paradigms above, the rationale is that this will deprive participants of the very resources they need to engage in proper deliberation. Consequently, the paradigm gives us a good indication of which response can be generated intuitively and deliberately (Bago & De Neys, 2017, 2020; Raoelison, Thompson, & De Neys, 2020; Thompson et al., 2011).

Under the exclusivity assumption, it is expected that people who generate the alleged system 2 response as their final response will initially have generated the system 1 response in the first, intuitive response stage. That is, in the prototypical dual-process test situation in which both systems are expected to cue a conflicting response, it is assumed that slow deliberation will need to correct and override the intuitively generated fast system 1 response. For example, in a classic bias task, it is hypothesized that people will initially generate the biased system 1 response but that sound reasoners will consequently be able to correct this once they are allowed to take the time to deliberate. To illustrate, take the infamous cognitive reflection test (e.g., "A bat and ball cost $1.10 together. The bat costs $1 more than the ball. How much does the ball cost?," Frederick, 2005). Here it is expected that sound reasoners will reason correctly precisely because they will take the time to reflect on their first hunch ("10 cents") which allows them to realize that it is incorrect. It is this demanding deliberation or "reflection" that is assumed to be crucial for generation of the correct answer ("5 cents"). However, two-response studies with these and other classic bias tasks have shown that this is typically not the case. Those reasoners who give the correct response as their final response after deliberation often already generate this same correct response at the initial, intuitive response stage (e.g., Bago & De Neys, 2017, 2019a; Burič & Konrádová, 2021; Burič & Šrol, 2020; Dujmović, Valerjev, & Bajšanski, 2021; Raoelison et al., 2020; Thompson & Johnson, 2014). Hence, sound reasoners do not need to deliberate to correct an initial response, their initial response is already correct.

This same pattern has been observed during moral (Bago & De Neys, 2019b; Vega, Mata, Ferreira, & Vaz, 2021) and prosocial (Bago, Bonnefon, & De Neys, 2021; Kessler, Kivimaki, & Niederle, 2017) reasoning. People who generate the alleged system 2 response (e.g., utilitarian moral decision or selfish prosocial choice) typically already generate this same decision as their intuitive response in the initial response stage. Hence, pace the exclusivity assumption, the alleged system 2 response is often already generated intuitively.

Related evidence comes from studies with the conflict detection paradigm (e.g., De Neys & Pennycook, 2019). This paradigm focuses specifically on those participants who give the alleged system 1 response. The studies contrast people's processing of classic prototypical problems (i.e., "conflict problems") in which systems 1 and 2 are expected to cue different responses and control "no-conflict" problems in which both systems are expected to cue the same response. For example, in a logical reasoning task such as the introductory ratio bias problem, a control problem could be one in which participants have to choose between a small tray with one red marble and a large tray with 11 (instead of nine) red marbles. In this case both the absolute number of red marbles (nine vs. one) and the ratios (11/100 vs. 1/10) favor the large tray. In a moral reasoning study, a no-conflict control problem could ask whether it is acceptable to kill five people to save the life of one person (instead of killing one to save five). Both utilitarian and deontological considerations will converge here in that the action is not permissible.

By and large, conflict detection studies have found that on various processing measures, reasoners who give the alleged system 1 response typically show sensitivity to the presence of conflict with the alleged system 2 response. For example, they take longer and are less confident when solving classic "conflict" versus control "no-conflict" problems (e.g., Białek & De Neys, 2016; Frey, Johnson, & De Neys, 2018; Gangemi, Bourgeois-Gironde, & Mancini, 2015; Mata, 2020; Šrol & De Neys, 2021; Vartanian et al., 2018; see De Neys, 2017, for a review but also Travers, Rolison, & Feeney, 2016; or Mata, Ferreira, Voss, & Kollei, 2017, for negative findings). Hence, even people who give the alleged system 1 response seem to be processing the alleged system 2 response. Critically, this conflict sensitivity is also observed when potential system 2 processing is knocked out with experimental constraint manipulations (e.g., Bago & De Neys, 2017, 2019b; Białek & De Neys, 2017; Burič & Konrádová, 2021; Burič & Šrol, 2020; Johnson, Tubau, & De Neys, 2016; Pennycook, Trippas, Handley, & Thompson, 2014; Thompson & Johnson, 2014). In line with the two-response findings, this indicates that the alleged unique system 2 response is also being processed intuitively.

In sum, although the idea that intuitive and deliberate processing are cueing unique responses is appealing in its simplicity, taken together, the empirical evidence reviewed here indicates that there is no strong empirical ground for it. In the most influential dual-process applications, the alleged system 2 response does not seem to be out of reach of the intuitive system 1. Rather than positing unique responses in systems 1 and 2, it appears that system 1 can often handle both responses.

To avoid confusion, it is important to stress here that the above conclusion does not argue against the idea that deliberation *can* lead to generation of the alleged system 2 response. For example, the meta-analyses I referred to often suggest that there is evidence for a small effect in the expected dual-process direction (i.e., more alleged system 2 responses after deliberation). Also, the two-response data consistently indicate that there are cases in which an initial, intuitively generated response is replaced with the alleged system 2 response after deliberation. The point is that this is rare. More often than not, the alleged deliberate response tends to be generated intuitively. Exclusive deliberate generation of the alleged system 2 response seems to be the exception rather than the rule. This implies that any model in which generation of this response is exclusively or predominantly tied to the operation of the deliberate system will have poor empirical fit.

A possible general argument against the reviewed empirical evidence contra the exclusivity assumption is that we can never be sure that the study designs prevented all possible deliberation. For example, it might be that the two-response studies still allowed some minimal deliberation during the initial response generation. It might be this minimal deliberation that drives the generation of the "alleged" system 2 response during the initial response stage. Here it should be noted that the two-response studies adopted the same constraint methodology and logic as the initial studies that were used to argue in favor of the exclusivity assumption. Moreover, whereas traditional studies used either time-pressure or load manipulations, the two-response studies have combined both to further restrict potential deliberate intrusion (e.g., Bago & De Neys, 2017). In addition, control studies indicate that making the constraints even more challenging by increasing the load and decreasing the deadlines typically does not alter the results (e.g., Bago & De Neys, 2017, 2019b; Bago et al., 2021), suggesting that deliberation was successfully minimized in the design. Nevertheless, the point still stands that no matter how challenging the test conditions might be, we can never be completely sure that participants did not deliberate. The problem here is that the dual-process framework does not give us an unequivocal threshold (i.e., longer than x seconds or less than x amount of load implies deliberation) that allows us to universally demarcate intuition and deliberation (Bago & De Neys, 2019a; De Neys, 2021). Ultimately, this implies that exclusivity cannot be empirically falsified. As long as one keeps on observing alleged system 2 responses under constraints, one can always argue that the constraints were not challenging enough. The general point is that the cognitive constraint evidence needs to be interpreted within practical, relative boundaries (Bago & De Neys, 2019a). In sum, although empirical evidence can question exclusivity and can point to a lack of strong supporting evidence, it can never rule it out completely. Therefore, in the next section, I will focus on a conceptual critique that underscores that positing exclusivity is fundamentally problematic for a dual-process model.

2. Switch issue

Although it might not be necessary to generate the alleged "system 2 response" per se, we sometimes clearly do engage in deliberation. Given that we can use two types of reasoning, there might be cases in which either one will be more or less beneficial. For example, in situations in which intuitive and deliberate processing are expected to cue the same response (e.g., the "no-conflict" problems I referred to earlier), there is no need to waste precious resources by engaging in costly deliberation. But how do we know that we can rely on an intuitively cued problem solution or need to revert to deliberation? And when we do decide to engage in deliberation, at what point do we decide it is safe to switch back to the mere intuitive processing mode?

Of course, there are some situations in which this is straightforward. One concerns cases in which we are faced with an entirely new problem we haven't seen before and our intuitions are not cueing a response. Here, all we can do to arrive at an answer is to engage in deliberation. Likewise, there will be cases in which the decision is made for us. That is, in some situations we get external feedback that indicates that an intuitively cued response is problematic. Generally speaking, these are cases of expectancy violations. For example, imagine your superior told you that you are getting a new colleague named Sue. Given their name, you'd readily expect that Sue is female. If your officemate subsequently tells you that the new colleague is a man, you'll presumably be surprised. Your system 1 has built up an expectation that is not met in the face of feedback. This expectancy violation will cue deliberation (Did you mishear the name? Was your colleague mistaken? Are Sue's parents Johnny Cash fans[3]? etc.). Unfortunately, the expectancy violation mechanism only works in case you're actually getting feedback. In many situations this will not be available or we want reasoners to operate (and avoid mistakes) without external supervision. Hence, reasoners need an internal mechanism that signals a need to switch between mere intuitive and deliberate processing.

My point is that traditional dual-process models have failed to present a viable internal switch mechanism. Popular accounts are conceptually problematic and this can be directly tied to the exclusivity assumption. I'll clarify that as long as we posit exclusivity, it will always be hard for a dual-process model to explain how reasoners can ever reliably determine whether there is a need to switch between intuitive system 1 and deliberate system 2 processing. I start by giving an overview of the dominant traditional switch views to clearly illustrate the problem.

2.1 Traditional switch accounts

2.1.1 Conflict monitoring system 2

Dual-process models are typically – what is being referred to as – "default-interventionist" in nature (Evans & Stanovich, 2013; Kahneman, 2011). This implies that they posit a serial processing architecture. The idea is that we rely on system 1 by default and only turn on the costly deliberate system to intervene when it is needed. It is this feature that brings about the switch question, of course. The traditional solution is to assume that system 2 is monitoring the output of system 1 and will be activated in case of conflict between the two systems (Kahneman, 2011; Stanovich & West, 2000). Hence, system 2 will intervene on system 1 whenever the system 1 output conflicts with more deliberate system 2 considerations. This idea is appealing in its simplicity. However, on second thought it is clear that it readily leads to a paradox (De Neys, 2012; Evans, 2019). To detect that our system 1 intuition conflicts with unique deliberate system 2 considerations, we would already need to engage system 2 first to compute the system 2 response. Unless we want to posit an all-knowing homunculus, system 2 cannot activate itself. Hence, the decision to activate system 2 cannot rely on the activation of system 2. The prototypical conflict monitoring system 2 account simply begs the question here (De Neys, 2012).

2.1.2 Low-effort deliberation

A popular variant of the simple conflict monitoring system 2 position – or a workaround – is to posit that the monitoring relies on low-effort deliberation and not on full-fledged demanding system 2 processing (De Neys & Glumicic, 2008; Kahneman, 2011). Whenever system 1 is cueing a response it will be passed on to system 2 which is by default in this non-demanding, low-effort mode. If the low-effort deliberation detects a conflict between system 1 and 2 processing, it will trigger deeper, high-effort deliberation (De Neys & Glumicic, 2008; Kahneman, 2011). Unfortunately, this simply pushes the explanatory burden one step forward. Clearly, if the low-effort mode suffices to generate a response against which the intuitive response can be contrasted, there is no need to postulate a unique high-effort deliberation (and to assume that the alleged system 2 response can only be computed by those highest in cognitive capacity, for example). In this case, everyone – even those lowest in cognitive capacity – should be able to generate the non-demanding deliberate response and it should not be considered unique to system 2. However, in case we assume that generating the deliberate response does require proper demanding system 2 processing, we are back at square one and we cannot explain how the low-effort system 2 processing detects conflict with the high-effort deliberate response in the first place. Hence, although it might sound appealing, the low-effort deliberation position does not present a viable processing mechanism.

2.1.3 System 3

One of the core problems of the conflict monitoring system 2 account is that system 2 is assumed to both generate a unique deliberate response and monitor for conflict between systems 1 and 2 to make the switch decision. It serves multiple functions: response generation and monitoring/switching. One suggested solution is to attribute the monitoring and switch decision to a third type of system or process (i.e., system 3 or type 3 processing, e.g., Evans, 2009; Houdé, 2019). Hence, system 2 computes a deliberate response and system 3 compares the output of systems 1 and 2. System 3 itself operates automatically and does not require the limited cognitive resources that system 2 needs. In case system 3 detects an output conflict, it will intervene, call for more deliberation and block the system 1 response. However, this solution still begs the question and leads to an infinite regression. To decide whether the system 1 output conflicts with the system 2 output, system 2 needs to be activated to compute a response, of course. Even an automatically operating system 3 cannot know whether there is a conflict between systems 1 and 2 without engaging system 2 first.

2.1.4 Parallel solution

A radically different solution to explain how we know that our intuition can be trusted or we need to engage in deliberation is to simply assume that systems 1 and 2 operate in parallel (Epstein, 1994; Sloman, 1996). In contrast to the dominant serial view, parallel dual-process models assume that intuitive and deliberate thought processes are always activated simultaneously when we are faced with a reasoning problem. Hence, just like intuitive processing, system 2 is always on. We always activate both reasoning systems from the start. Consequently, we also do not need a mechanism to decide whether or not we need to engage in deliberation and switch system 2 on.

The key problem is that the parallel account throws out the cognitive advantage of a dual-process model (De Neys, 2012). That is, nobody contests that system 1 will often converge with system 2 and can cue sound decisions. Hence, in these cases there is no need to burden our precious cognitive resources with demanding system 2 activation. Consequently, a parallel model will often be wasting scarce resources in situations where it is not needed. From a cognitive economy point of view, this is highly implausible. Furthermore, in case the parallel system 1 and 2 computations do lead to conflicting responses, the fast system 1 will need to wait until the slow system 2 has computed its response to register the conflict and decide which response to favor. But if the fast system 1 always waits for system 2, we lose the capacity to reason and act fast. On the contrary, if the fast system 1 does not wait for system 2, how are we to know that the system 1 response is valid and does not conflict with system 2? Hence, just like its serial competitors, the parallel account leads to conceptual inconsistencies and fails to present a working processing account.

To avoid confusion, note that the problem for the parallel account is not the parallel activation of systems 1 and 2 per se but the postulated *continuous* parallel activation of both systems. That is, the serial default-interventionist account also assumes that once system 2 is activated, system 1 remains activated and that the two systems will be running in parallel at this point.

https://doi.org/10.1017/S0140525X2200142X Published online by Cambridge University Press

The key difference is that the serial model posits that there needs to be an initial phase in which people do not deliberate yet – and it is this feature that brings about the switch problem. One might be tempted to argue that a parallel model does not necessarily need to assume that system 2 is always on. When there is no longer a need for deliberation, system 2 could be switched off to avoid wasting resources and it may be turned on again whenever it is needed. But at this point, one will have re-introduced the switch issue and will need to explain how this decision is made. That is, such a "parallel" model throws out its conceptual advantage over the serial model (i.e., no need for a switch mechanism) and faces the same difficulties as its rivals.

Relatedly, one may argue that even if system 2 is always on, it doesn't always have to run to completion. Maybe it only provides some quick partial computations that suffice to generate a response and check whether it conflicts with the cued system 1 answer. Note that under this reading, the parallel model boils down to the low-effort-deliberation account (see sect. 2.1.2) and will face the same problems: If low-effort or partial system 2 processing already allows generating an accurate proxy of the complete system 2 response, there is no need to assume that computation of the alleged unique system 2 response is demanding and necessarily requires time and effort. But if more extensive system 2 processing is necessary, it is not clear how the partial deliberations may ever reliably signal conflict.

2.1.5 Stuck-in-system 1 or no switch account

Finally, a last alternative possibility is to assume that people do not detect there is a need to engage system 2 and always stay in system 1 mode. In this "no switch" model, reasoners simply never internally switch from system 1 to system 2 themselves. People can use system 2 but only in case system 1 does not cue a response or they are externally told to do so. Whenever system 1 cues a response they are bound to blindly rely on the intuitively cued problem solution. Hence, the account solves the switch question by positing that reasoners never switch. Such a model can explain why people often give the alleged system 1 response (e.g., why they are biased in the case of logical reasoning): They simply fail to detect there is a need to activate system 2 (e.g., Evans & Stanovich, 2013; Kahneman, 2011; Morewedge & Kahneman, 2010; Stanovich & West, 2000). Note that although the account might be questioned on empirical grounds (e.g., see the conflict detection findings in sect. 1), in contrast to the other accounts I reviewed it is at least conceptually coherent. It does not beg the question or introduce a homunculus. The problem, however, is that it only models half the story.

The "no switch" model allows us to account for the behavior of people who give the alleged system 1 response, but it turns a blind eye to those who do give the alleged system 2 response. Indeed, although it might be rarer, there are always reasoners who arrive at the alleged system 2 response themselves. In general, the fact that there are two types of responses is a key motivation to posit an (exclusive) dual-process model in the first place. Hence, one still needs to explain how these "system 2" responders managed to detect there was a need to engage system 2. Consequently, even in the stuck-in-system 1 account, the switch issue inevitably rears up its head again.

2.2 Toward a working switch solution

The overview pointed to the fundamental conceptual problems that plague popular switch accounts in traditional dual-process models. How can we avoid this conceptual muddle and arrive at a viable switch account? Any solution will have two necessary core components. First, we need to postulate that the internal switch decision is itself intuitive in nature. The switch decision needs to rely on mere system 1 processing. System 1 decides whether system 2 is activated or not. This avoids the paradox of assuming that to decide whether to engage in costly system 2 deliberation you already need to engage system 2 (De Neys, 2012; Evans, 2019; Stanovich, 2018). Second, and more controversially, we will need to discard the exclusivity feature. If we agree that system 1 takes the switch decision, the billion-dollar question then becomes how exactly it does this. What informs the decision within system 1? My point is that solving this puzzle forces us to get rid of exclusivity. Instead of allocating unique responses to each system, we need to assume that the alleged system 2 response can also be cued by system 1. Hence, system 1 will be generating different types of responses or intuitions. One of these will be the traditional alleged system 1 response (e.g., a biasing heuristic, deontological, or prosocial intuition), the other one will be the traditional alleged system 2 response (e.g., logical, utilitarian, or selfish intuition). In case both intuitions cue the same response, the response can be given without further system 2 deliberation. In case the two intuitions cue conflicting responses, system 2 will be called upon to intervene.

With these building blocks in hand, it is possible to present a conceptually coherent switch account. It will be conflict between competing intuitions within system 1 that will function as the trigger to switch on system 2. But clearly, by definition, the account can only work if the alleged system 2 response is not exclusively calculated by system 2. If exclusivity is maintained, there is no way for system 1 to be reliably informed about potential conflict with the exclusive system 2 response. An exclusive model is bound to fall prey to the same conceptual pitfalls that plague the traditional switch accounts.

To avoid confusion, the point is not that exclusivity is impossible per se. Non-exclusivity is not a necessary prerequisite for a dual-process model. The point concerns the necessary conceptual coupling between the exclusivity and switch features. A dual-process model may posit exclusivity, but it will pay the price at the switch front. To remain coherent, a dual-process model that posits exclusivity will also need to postulate that reasoners have no internal mechanism that allows them to switch from system 1 to system 2 themselves (i.e., the stuck-in-system 1 position). One cannot have their exclusive cake and eat it here.

The good news is that the empirical evidence reviewed in section 1 indicates that the elementary conditions for the above switch mechanism may often be met. In key dual-process applications there is evidence that the alleged system 2 response can indeed be processed more intuitively. Hence, the required building blocks for a coherent switch mechanism seem to be in place. However, although positing non-exclusivity might provide the building blocks, it clearly does not suffice to arrive at a workable model. For example, one may wonder why reasoners often still opt for the alleged system 1 response if the alternative response is also intuitively available? Relatedly, what exactly determines system 2 engagement? Does the mere generation of two conflicting intuitions suffice per se? Does the amount of conflict matter? Furthermore, we do not only need to explain when reasoners will engage system 2 but also when they will stop doing so. That is, once we have activated system 2 it doesn't stay activated forever. At what point does a reasoner decide it is safe to revert back to system 1 processing then? In the following section,

https://doi.org/10.1017/S0140525X2200142X Published online by Cambridge University Press

I sketch a general architecture that allows us to address these issues.

3. Working model

The model I develop here builds on emerging ideas from various authors working in a range of dual-process application fields (e.g., Bago & De Neys, 2019b, 2020; Bago et al., 2021; Baron & Gürçay, 2017; De Neys & Pennycook, 2019; Evans, 2019; Pennycook, Fugelsang, & Koehler, 2015; Reyna, Rahimi-Golkhandan, Garavito, & Helm, 2017; Stanovich, 2018; Thompson & Newman, 2017; Trippas & Handley, 2017[4]). Because these ideas often entail some revision of traditional dual-process models they are sometimes collectively referred to as dual-process theory 2.0 (De Neys, 2017). The current model presents a personal integration and specification of what I see as key features. I focus on a general, field-independent specification that can serve as a basic architecture for future models across various fields.

The model has four core components which I will introduce in more detail below. Figure 1 presents a schematic illustration.

3.1 Intuitive activation

The first component (illustrated in Fig. 1.1) reflects the starting point that system 1 can be conceived as a collection of intuitively cued responses. For convenience, I focus on the critical case in which two competing intuitions are being cued. These are labeled as intuition 1 (I1) and intuition 2 (I2). These can be the alleged system 1 and alleged system 2 responses but in general, they can be any two intuitions that cue a different response. Each intuition is simply identified by the response it cues.

At each point in time, an intuition is characterized by its activation level or strength. The strength can change over time. Once an intuition is generated it can grow, peak, and decay. The y-axis in Figure 1.1 represents the intuition strength, the x-axis represents time. The peak activation strength of an intuition reflects how automatized or instantiated the underlying knowledge structures are (i.e., how strongly it is tied to its eliciting stimulus, e.g., Stanovich, 2018). The stronger an intuitive response is tied to its eliciting stimulus, the higher the resulting activation strength. This implies that not all intuitions will be created equal. Some might be stronger than others.

But where do these intuitions and strength differences come from? Although it is not excluded that some intuitive associations might be innate, the working model postulates that intuitive responses primarily emerge through an automatization or learning process. Throughout development, any response might initially require exclusive deliberation but through repeated exposure and practice this response will become compiled and automatized (e.g., Shiffrin & Schneider, 1977). Note that although such a claim is uncontroversial for the alleged system 1 response in traditional dual-process models (e.g., Evans & Stanovich, 2013; Rand et al., 2012), it is assumed here that it also applies to the alleged system 2 response. The rationale is that in most dual-process fields, adult reasoners have typically already been exposed to the system 2 response through education and daily life experience. For example, the ratio principle in the introductory ratio bias task is explicitly taught during elementary and secondary education (e.g., fractions). Likewise, children will have had many occasions to experience that selfish behavior has often negative consequences (e.g., if you don't share with your little brother your mom and dad will be mad, your brother will be less likely to share with you in the future, etc.). Hence, through repeated exposure and practice an original system 2 response may gradually become automatized and will be generated intuitively (De Neys, 2012). But because not every response will have been equally well automatized or instantiated, strength differences may arise, and not every eliciting stimulus will cue the associated response equally well in system 1.

Note that the eliciting stimulus can be any specific problem feature. For example, when solving the ratio bias problem with the marbles and trays, the absolute number information (e.g., "1 red marble in small tray, 9 red in large tray") might give rise to one intuition (e.g., "pick large") and the ratio information (e.g., "1 out of 10 red vs. 9 out of 100 red") might give rise to a conflicting one (e.g., "pick small"). In a moral reasoning problem, the information that an action will result in harm (e.g., a person will die) can cue a deontological intuition (e.g., "action not acceptable") and the subsequent information that it may prevent more harm (e.g., "if nothing done, 5 people will die") an utilitarian one (e.g., "action acceptable"). Hence, the intuition 1 (I1) and intuition 2 (I2) labels in the illustration simply refer to the temporal order in which the intuitions accidentally happened to be cued. They bear no further implications concerning the nature of the intuition per se.

3.2 Uncertainty monitoring

The second component of the model is what we can refer to as an uncertainty monitoring process. The idea is simply that system 1 will continuously calculate the strength difference between activated intuitions. This results in an uncertainty parameter U. The more similar in strength the competing intuitions are, the higher the resulting experienced uncertainty. Once the uncertainty reaches a critical threshold (represented by d in Fig. 1.2), system 2 will be activated. However, in case one intuition clearly dominates the other in strength, the resulting uncertainty will be low and the deliberation threshold will not be reached. In that case, the reasoner will remain in system 1 mode and the dominant intuition can lead to an overt response without any further deliberation.

This explains why postulating non-exclusivity and assuming that the traditionally alleged system 2 response can also be generated intuitively does not imply that reasoners will always opt for the alleged system 2 response. For different individuals and situations, the strength of the competing intuitions can differ. Sometimes the alleged system 1 intuition will dominate. Consequently, although the presence of a competing intuition that cues the alleged system 2 response will result in some uncertainty, this may not be sufficient to engage system 2. In the case of logical reasoning bias, for example, this explains why some reasoners may detect that their dominant intuitive answer is questionable but nevertheless will fail to engage in further deliberation to double-check and correct it.

A possible mathematical representation of the uncertainty parameter is: $U = 1 - |I1 - I2|$. U stands for uncertainty and can range from 0 to 1. I1 and I2 represent the strength of the respective intuitions. The strength can also range between 0 and 1. The vertical bars (|) denote we calculate the absolute difference. Hence, the more similar the activation strength, the smaller the absolute difference and the higher the uncertainty will be.

A simple analogy might clarify the basic idea. Imagine that as part of a lunch combo, a local cafeteria offers its customers a choice between two desserts: ice cream or a cupcake. John is

https://doi.org/10.1017/S0140525X2200142X Published online by Cambridge University Press

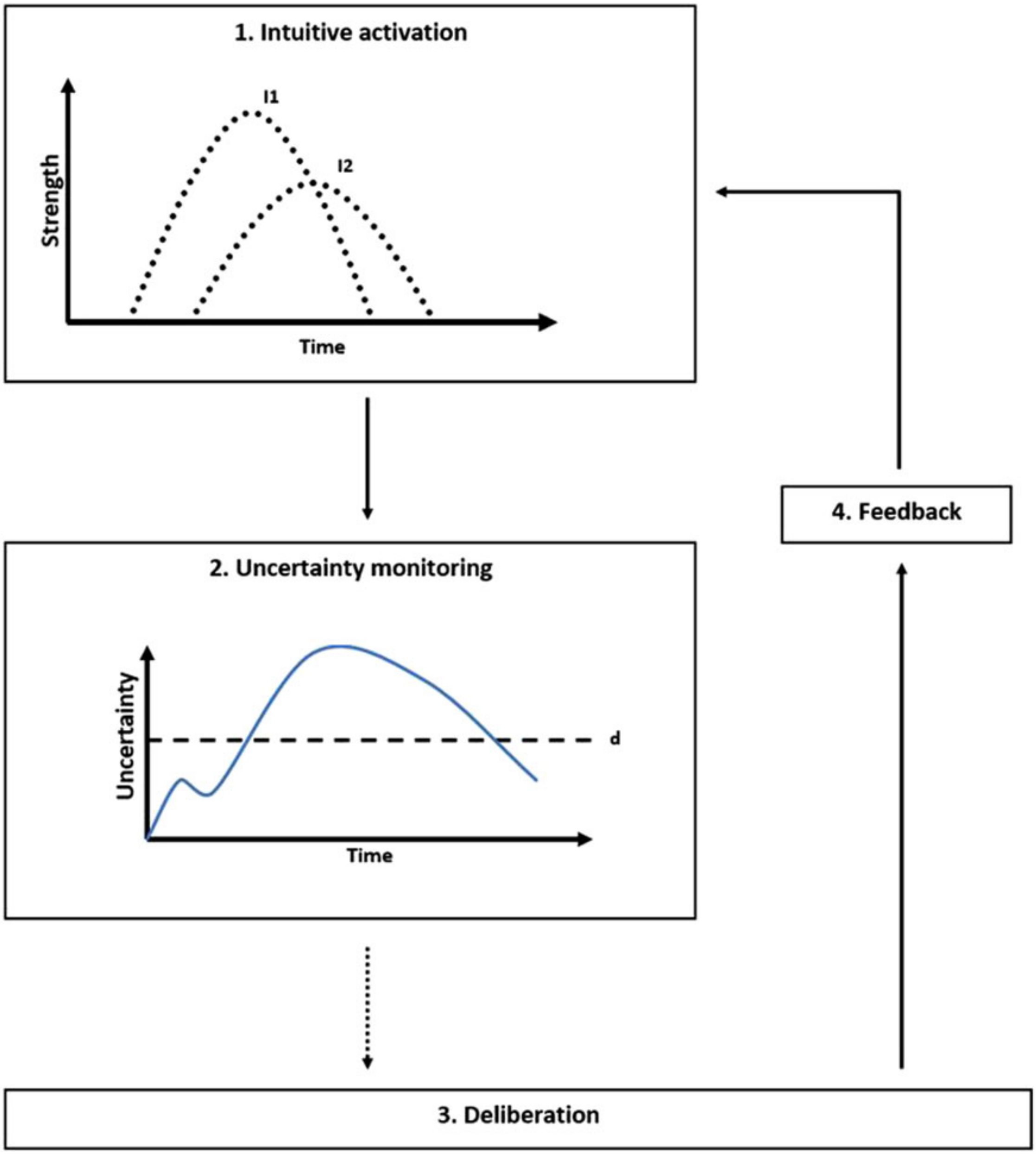

Figure 1. Schematic illustration of the working model's core components. I1, intuition 1; I2, intuition 2; d, deliberation threshold. The dashed arrow indicates the optional nature of the deliberation stage.

fond of ice cream but really dislikes cupcakes. Hence, John will readily choose the ice cream without giving it any further reflection. Steve, however, likes both equally well. When presented with the two options, Steve's decision will be harder and require deeper deliberation. For example, he might try to remember what he had last time he ate at the cafeteria and decide to give the other option a try. Or he might try to look for arguments to help him make a decision (e.g., "The cupcake has blueberries in it this week. Blueberries are healthy. Better take the cupcake."). Just like the strength of our food preferences, the activation strength of our intuitions will similarly determine whether or not we will deliberate about our response.

Note that although I focus on two competing intuitions, the monitoring also applies in case there is only one or no intuition cued. For example, if a reasoner is being faced with an entirely new problem for which system 1 does not cue a response, the absolute difference factor will equal 0 (i.e., the intuition strength equals 0), the resulting uncertainty will be maximal (e.g., $U = 1 - 0$), and system 2 will be called upon to compute and answer. If a problem only cues one single intuition (or both intuitions cue the same response), the difference factor will equal its strength (e.g., 0.8). Consequently, if the strength is high, the uncertainty will be low (e.g., $U = 1 - 0.8$) and the cued response can be selected without further deliberation. Conversely, a weaker intuition will result in a higher uncertainty, which increases the likelihood that the deliberation threshold is crossed, system 2 is activated and the reasoner engages in additional deliberation about the problem. Finally, one may also envisage cases in which more than two intuitions are simultaneously activated. If there is one intuition that clearly dominates, the strength difference will be high and no further

https://doi.org/10.1017/S0140525X2200142X Published online by Cambridge University Press

deliberation will be engaged. In case the differences are more diffuse, deliberation will likewise be triggered.

It is important to recap that uncertainty monitoring is a core system 1 process. It operates effortlessly without any system 2 supervision. For illustrative purposes, it is represented as a separate box in Figure 1. It can be functionally isolated but at an implementation level there is no need to postulate a different type of system or processing. It should also be clear that deliberation is always optional; it will only be engaged when the uncertainty monitoring deliberation threshold is reached. This is represented in Figure 1 by the dashed arrow between the uncertainty monitoring and deliberation component.

3.3 Deliberation

The third component is system 2 activation. It is at this stage (and this stage only) that the reasoners will engage in slow, demanding deliberation. Deliberation can take many forms. For example, one classic function is its role as response inhibitor (e.g., De Neys & Bonnefon, 2013; Evans & Stanovich, 2013). Here attentional control resources will be allocated to the active suppression of one of the competing intuitions. In addition, some authors have pointed to the algorithmic nature of deliberation and its role in the generation of new responses (e.g., Houdé, 2019). In this case system 2 allows us to retrieve and execute a stepwise sequence of rules. For example, when we have to multiply multiples of 10 (e.g., "How much is 220×30?"), we can use a multiplication algorithm (e.g., multiply the non-zero part of the numbers, i.e., $22 \times 3 = 66$; count the zeros in each factor, i.e., 2; add the same number of zeros to the product, i.e., 6,600) to calculate an answer. While we're executing each step we need to memorize the results of the previous steps which will burden our attentional resources. When system 1 does not readily cue an intuitive response, such algorithmic system 2 deliberation allows us to generate an answer.

Likewise, some authors have also pointed to the role of deliberation in a justification or rationalization process (Bago & De Neys, 2020; Evans, 2019; Evans & Wason, 1976; Pennycook et al., 2015; see also Cushman, 2020; Mercier & Sperber, 2011). In this case we will deliberate to look for an explicit argument to support an intuition. This explains why engagement of system 2 does not imply that the alleged system 2 response will be generated. Reasoners can also use their cognitive resources to look for a justification for the alleged system 1 intuition (e.g., the incorrect "heuristic" intuition in logical reasoning tasks). More generally, this underscores the argument that system 2 engagement does not "magically" imply that the resulting response will be "correct," "rational," or "normative" (De Neys, 2020; Evans, 2009, 2019). It simply implies that a reasoner will have taken the time and resources to explicitly deliberate about their answer.

Clearly, none of these roles need to be mutually exclusive. Deliberation might entail a combination of response suppression, generation, justification, or additional processes. Whatever the precise nature of deliberation may be, what is critical for the current purpose is the outcome or result. The key point is that deliberation will always operate on system 1 in that it will modulate the strength of the different activated intuitions in system 1 (or generate a new intuitive response altogether). Consequently, although it is possible to have system 1 activation without system 2 activation, the reverse is not true. During deliberation, the effortless system 1 remains activated and deliberation will operate on its strength representations. As I will explain in more detail below, it is this feature that provides us with a mechanism to stop system 2.

3.4 Feedback

A last component of the model is what we can refer to as a feedback loop. A reasoning process does not stop at the point that one starts to deliberate. Traditionally, dual-process models have mainly focused on the question as to how we can know when to engage system 2. The question as to how we know we can stop system 2 engagement has received far less attention. Clearly, a viable switch account requires us to address both questions. When we activate the effortful system 2, at some point we will need to revert back to system 1. Hence a working dual-process model needs to specify when system 2 will be switched on *and* off. Put bluntly, we not only need to know what makes us think (Pennycook et al., 2015) but also what makes us stop thinking.

The simple idea I put forward here is that of a feedback loop. System 2 operates on the strength representations in system 1 such that the outcome of system 2 processing is fed back into system 1. Hence, because deliberation will act on the strength representations, it will also affect the uncertainty parameter. For example, if we deliberately suppress one of two competing intuitions, this will decrease its activation level. Because of this decrease, the activation difference with the non-suppressed intuition will increase. As a result, the uncertainty parameter will decrease. At the point that the uncertainty falls below the deliberation threshold, system 2 deliberation will be switched off and the reasoner will return to mere system 1 processing.

In other words, in essence, the critical determinant of system 2 engagement is the uncertainty parameter. As soon as it surpasses the deliberation threshold, the reasoner will start deliberating. System 2 deliberation will extend for as long as the uncertainty remains above the threshold. As soon as the uncertainty drops below the threshold, deliberation stops, and the reasoner will revert to mere system 1 processing. Hence, it is the uncertainty parameter that determines the extent of deliberation. Figure 2 tries to illustrate this core idea. The figure sketches a situation in which initially only system 1 is activated and two intuitions are generated, a first intuition (I1) and slightly later a second intuition (I2). The activation strength of the two intuitions gradually increases. Initially, there is a large activation difference between I1 and I2 and consequently, the U parameter will be low. However, at a certain point I1 plateaus whereas I2 is still increasing. Consequently, their activation strength becomes more similar, U will increase, and the deliberation threshold will be crossed. At this point (t1), system 2 will be activated. This activation will modulate the strength through deliberate suppression, rationalization, and so on. This may decrease or increase the activation strengths and uncertainty parameter. As long as the uncertainty parameter remains above the threshold, system 2 activation will be extended (represented by the gray bar in Fig. 2). At a certain point (t2 in the figure), the activation difference will be sufficiently large again such that the uncertainty falls below the threshold and the reasoner switches back to pure system 1 processing.

To avoid confusion, it is important to stress that deliberation does not necessarily need to lead to a decreased uncertainty (or "conflict resolution") per se. Deliberation can also increase uncertainty and lead to more deliberation. For example, one can think of a situation in which initially a single weak intuitive response is cued. This leads to high uncertainty and system 2 engagement. Subsequently, algorithmic processing leads to the generation of a new, competing response. This response will also be represented

https://doi.org/10.1017/S0140525X2200142X Published online by Cambridge University Press

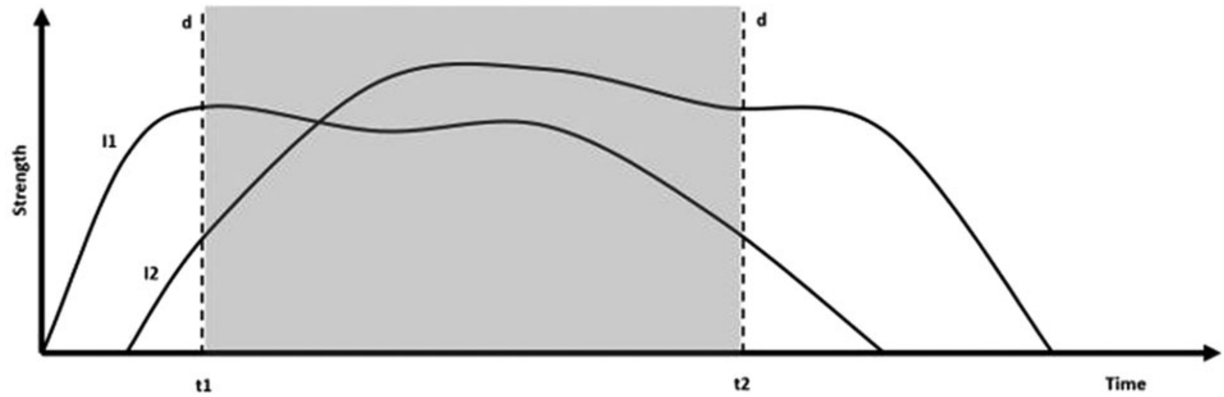

Figure 2. Illustration of the idea that the strength interplay of conflicting intuitions determines uncertainty and the extent of deliberation. I1, intuition 1; I2, intuition 2; d, deliberation threshold; t1 and t2, time points at which the deliberation threshold is crossed. The gray area represents the time during which system 2 deliberation will be engaged.

in system 1 and have a specific strength. Depending on the specific activation levels, the net result might very well be more rather than less uncertainty which will lead to further deliberation. Alternatively, imagine that during logical reasoning on a classic bias task, a reasoner generates both a logically correct and incorrect ("heuristic") intuition. The heuristic intuition is only slightly stronger than the logically correct one and the resulting uncertainty triggers system 2 deliberation. During deliberation the reasoner looks for a justification for the heuristic intuition but does not find one. As a result, its strength will decrease making it even more similar to the logical intuition. Consequently, the uncertainty will increase and deliberation will be boosted rather than stopped. These are illustrative examples but they underscore the core point that there is no necessary coupling between deliberation and uncertainty reduction or resolution per se. The point is that the feedback mechanism guarantees that deliberation *can* reduce uncertainty and thereby stop system 2 engagement.

In the full model sketch in Figure 1, the feedback component – just like the uncertainty monitoring component – is represented in a separate box. Just as with the uncertainty monitoring component, it can be functionally isolated but there is no need to postulate a different type of system or type of processing. Feedback results from system 2 processing but the critical updating of the system 1 representations itself occurs automatically and does not require additional cognitive resources. In this sense it is a system 1 process. At the same time, the feedback component also underscores that in practice, thinking always involves a continuous interaction between system 1 and system 2 activation. At a specific isolated point in time we'll be either in system 2 mode or not but this split-up is always somewhat artificial. In practice, reasoning involves a dynamic interaction between the two systems. System 1 can call for system 2 activation which will operate on system 1 which can lead to more or less system 2 operation which will further affect system 1 operations. This dynamic interaction is represented by the flow arrows in Figure 1.

3.5 Working guidelines

The combined intuitive activation, uncertainty monitoring, deliberation, and feedback components sketch the basic architecture of a dual-process model that can explain how people switch between system 1 and system 2 thinking. The model sketch also allows us to delineate some more general principles that a working dual-process model needs to respect: First, the model needs to be default-interventionist in nature. The idea of a parallel model in which systems 1 and 2 are always activated simultaneously is both empirically and conceptually problematic. A dual-process model should not assume that system 2 is always on. There will always need to be a processing stage in which the reasoner remains in mere system 1 mode. Second, because system 2 cannot always be on, the model needs to specify a switch mechanism that allows us to decide when system 2 will be turned on (and off). Third, while it is critical that there is a state in which system 1 is activated without parallel system 2 activation, the reverse does not hold. During system 2 activation, system 1 always remains activated. System 2 necessarily operates on the system 1 representations. This modulation ultimately allows us to stop deliberating. Fourth, a viable internal switch account implies that the model will be non-exclusive. As soon as we posit exclusive responses that are out of reach of the intuitive system, it will be impossible for the reasoner to accurately determine whether there is a need to generate the exclusive deliberate response when they are in the intuitive processing mode. In case exclusivity is nevertheless maintained, the model necessarily posits that there is no reliable internal switch mechanism.

If these features or principles are not met, the model will not "work" and cannot qualify as a proper dual-process model that allows us to explain how intuition and deliberation interact. As such, the model sketch may help to separate the wheat from the chaff when evaluating future dual-process accounts.

4. Prospects

I referred to the architecture I presented as a working model. This label serves two goals. On the one hand, it stresses that the model "works" in that it presents a viable account that avoids the conceptual pitfalls that plague traditional dual-process models. However, on the other, the "working" also refers to its preliminary status – the model is a work-in-progress. The current specification is intended as a first, high-level verbal description of the core processes and operating principles. Clearly, the model will need to be further fleshed out, fine-tuned, and developed at a more fine-grained processing level. In this section I point to critical outstanding questions that will need to be addressed. These queries have remained largely neglected in the dual-process field. As such, the section can also illustrate the models' potential to

https://doi.org/10.1017/S0140525X2200142X Published online by Cambridge University Press

identify and generate new research questions and set the research agenda in the coming years.

4.1 Uncertainty parameter specification

The working model specifies the uncertainty parameter U as the absolute strength difference between competing intuitions (i.e., $U = 1 - |I1 - I2|$). This is most likely an oversimplification. For example, the current model does not take the absolute activation level into account. That is, two weak intuitions that have the same strength level (e.g., both have activation level 0.1 out of 1) are assumed to result in the same level of uncertainty as two strong intuitions that have the same strength level (e.g., both have activation level 0.9 out of 1). If two intuitions have trivially small activation levels, one may wonder whether potential conflict requires or warrants deliberation. It is not unreasonable to assume that we would primarily allocate our precious cognitive resources to the most highly activated or most intense conflicts. One way to account for this feature would be to incorporate the absolute strength level into the U parameter. For example, by multiplying the absolute difference with the individual strength levels such that $U = (1 - |I1 - I2|) \times I1 \times I2$. Under this specification, conflict between stronger intuitions will be weighted more heavily and result in more uncertainty.

Likewise, one may wonder whether the variability of the strength levels is taken into account. Imagine two situations in which upon generation of competing intuitions the uncertainty parameter reaches the deliberation threshold after 1 second. In the first case, the intuition strength levels gradually change such that the U parameter gradually increases until the deliberation threshold is reached. With every unit of time, the uncertainty smoothly increases. Contrast this with a case whereby the strength levels are highly variable and constantly shoot up and down. For example, imagine that initially the uncertainty steeply rises but after a couple of milliseconds it steeply drops, then rises again, drops, and then rises again before it ultimately crosses the threshold. In theory, this variability may be informative. Strength instability might signal an increased need for deliberation. Such a feature could be integrated into the model by factoring strength variability into the U parameter such that, for example, $U = (1 - |I1 - I2|) \times V(I1) \times V(I2)$. The V factor then simply reflects the variability of the strength level over an elapsed period of time (e.g., standard signal deviation). Consequently, more variability will result in more uncertainty and faster deliberation engagement.

In the same sense, in theory, the uncertainty may be impacted by the intuition rise time or strength slope. That is, imagine two intuitive responses that have the exact same peak strength level at a certain point in time. However, it took the first response twice as long to reach that level as the second response. In other words, the slope of the strength function of the first intuition will be much lower than that of the second intuition (i.e., the second one is steeper). Is this factored into the uncertainty equation? Or is the slope simply invariant (i.e., intuitive strength always rises at a fixed rate)? These are open queries but illustrate how the working model generates new research questions that have hitherto remained unexplored in the dual-process field.

Currently, these suggestions or hypotheses remain purely speculative. The absolute strength level, strength variability, slope, and other factors might or might not affect the uncertainty parameter. This remains to be tested and empirically verified. The point is that, in theory, the model can be updated to account for these refinements, and pinpointing the precise signal or strength characteristics that affect the experienced uncertainty should be a promising avenue for further research.

4.2 Nature of non-exclusive system 1 and 2 responses

In a non-exclusive model there is no unique, exclusive response in system 2 that can only be generated through deliberation. Any response that can be computed by system 2 can also be computed by system 1. However, it is important that this equivalence is situated at the response or outcome level. Generating a logically correct response in bias tasks, making a utilitarian decision during moral reasoning, or deciding between a selfish or prosocial decision in a cooperation task, can all be done intuitively. But this does not imply that the intuitive and deliberate calculation of the responses is generated through the same mechanism or has the same features. Indeed, given that one is generated through a fast automatic process and one through a slow deliberate process, by definition, the processing mechanisms will differ. To illustrate, consider one is asked how much "3×10" is. For any educated adult, the answer "30" will immediately pop up through mere intuitive processing. An 8-year-old who starts learning multiplication will initially use a more deliberate addition strategy (e.g., 3 times 10 equals $10 + 10 + 10$; $10 + 10$ equals 20, plus 10 is 30). Both strategies will result in the same answer, but they are generated differently and do not have the same features. For example, the intuitive strategy might allow the adult to respond instantly but when asked for a justification even adults might need to switch to a more deliberate addition strategy ("well, it's 30 because $10 + 10 + 10$ is thirty"). Hence, non-exclusivity does not entail that there is no difference between intuition and deliberation. The point is that intuition and deliberation can cue the same response.

However, it will be important to pinpoint how exactly the non-exclusive system 1 and system 2 responses differ. For example, one of the features that is often associated with deliberation is its cognitive transparency (Bonnefon, 2018; Reber & Allen, 2022). Deliberate decisions can typically be justified; we can explain why we opt for a certain response after we reflected on it. Intuitive processes often lack this explanatory property: People tend to have little insight into their intuitive processes and do not always manage to justify their "gut-feelings" (Marewski & Hoffrage, 2015; Mega & Volz, 2014). Hence, one suggestion is that non-exclusive system 1 and 2 responses might differ in their level of transparency (e.g., De Neys, 2022). For example, in one of their two-response studies on logical reasoning bias, Bago and De Neys (2019a) also asked participants to justify their answers after the initial and final response stages. Results showed that reasoners who gave the correct logical response in the final response stage typically managed to justify it explicitly. However, although reasoners frequently generated the same correct response in the initial response phase, they often struggled to justify it. Bago and De Neys (2019b) observed a similar trend during moral reasoning; although the alleged utilitarian system 2 response was typically already generated in the intuitive response stage, sound justifications of this response were more likely after deliberation in the final response stage. Hence, a more systematic exploration of the role of deliberation in response explicitation or justification seems worthwhile.

Likewise, one may wonder what the exact problem features are that system 1 reasoning exploits to generate the alleged system 2 response. For example, it has been suggested that computation

https://doi.org/10.1017/S0140525X2200142X Published online by Cambridge University Press

of correct intuitive responses during deductive reasoning may rely on surface features that closely co-vary with the logical status of a conclusion rather than logical validity per se (Ghasemi, Handley, Howarth, Newman, & Thompson, 2022; Hayes et al., 2022; Meyer-Grant et al., 2022). In this sense, intuitive logical reasoning would serve to calculate a proxy of logical reasoning but not actual logical reasoning. These questions concerning the precise nature of non-exclusive system 1 intuitions should help to fine-tune the model in the coming years.

4.3 System 2 automatization

The working model posits that the critical emergence of a non-exclusive "alleged system 2" intuition within system 1 typically results from a developmental learning or automatization process. Through repeated exposure and practice, the system 2 response will gradually become automatized and will be elicited intuitively (De Neys, 2012; Stanovich, 2018). The basic idea that an originally deliberate response may be automatized through practice, is theoretically sound (e.g., Shiffrin & Schneider, 1977) and well-integrated in traditional dual-process models (e.g., Evans & Stanovich, 2013; Rand et al., 2012).

However, although the automatization idea might not be unreasonable, there is currently little direct evidence to support it (De Neys & Pennycook, 2019). This points to a need for developmental research to test the emergence of these new intuitions (e.g., Raoelison, Boissin, Borst, & De Neys, 2021). Likewise, individual differences in the strength of intuitions might be linked to differences in response automatization. People might differ in the extent to which they have automatized the system 2 operations. To test this idea more directly, one may envisage training studies in which the activation level or automatization is further boosted through practice. Although there have been some recent promising findings in this respect (Boissin, Caparos, Raoelison, & De Neys, 2021; Purcell, Wastell, & Sweller, 2020), a more systematic exploration is key. Such work may have critical applied importance. Rather than training people to deliberate better to suppress faulty or unwanted intuitions, we might actually help them to boost the desired intuition directly within system 1 (e.g., Milkman, Chugh, & Bazerman, 2009).

Emerging evidence in the logical reasoning field also suggests that spontaneous differences in the strength of sound "logical" intuitions might be associated with individual differences in cognitive capacity (Raoelison et al., 2020; Schubert, Ferreira, Mata, & Riemenschneider, 2021; Thompson, 2021; Thompson, Pennycook, Trippas, & Evans, 2018). That is, people higher in cognitive capacity might have automatized the logical operations better and developed more accurate intuitions (Thompson et al., 2018). Consequently, rather than predicting how good one is at deliberately correcting faulty intuitions, cognitive capacity would predict how likely it is that a correct intuition will dominate from the outset in the absence of deliberation (Raoelison et al., 2020). Although promising, this finding will require further testing (e.g., Thompson & Markovits, 2021) and generalization to different fields.

4.4 Deliberation issues

The deliberation component of the working model will also need further development. I noted that deliberation can take many forms. It will be important to specify these and their potential interaction in more detail. For example, one may wonder about the link between suppression and justification. Do we ever suppress an intuitive response without a justification? That is, do we need an explicit argument or reason to discard an intuitive response, or is such justification independent of the suppression process and does it follow (rather than precede) suppression (Evans, 2019)? More critically perhaps, how are deliberative processes instantiated? For example, does the suppression process imply an active suppression of a target intuition per se or rather a boosting of the activation level of the competing intuition? Alternatively, it has been argued that deliberate suppression can be conceived as a mere response delay (Martiny-Huenger, Bieleke, Doerflinger, Stephensen, & Gollwitzer, 2021). Under this interpretation, the activation level of a dominant intuition automatically decays if it is not acted upon (i.e., does not result in an overt response). Hence, as long as the reasoner refrains from responding, the mere passive passing of time will guarantee that the activation level of an initially dominant intuition will fall below its competitor. Consequently, it would be the act of refraining from responding rather than the suppression of a dominant intuition itself that would be demanding. This illustrates how more work is needed to specify the precise instantiation of deliberation.

Another question concerns the gradual or discrete nature of deliberation engagement (Dewey, 2021, 2022). In the current model specification, I focused on the extent of deliberation. The longer the uncertainty parameter remains above the threshold, the longer we will remain deliberating. But in addition to the question as to how long we will keep deliberating for, one may also wonder how hard we will deliberate. How much of our cognitive resources do we allocate to the task at hand? Do we always go all-in, in an all-or-nothing manner or do we set the amount of allocated resources more gradually? In theory, the amount of deliberation might be determined by the uncertainty parameter. For example, the higher the uncertainty parameter (above the threshold), the more resources will be allocated. This issue will need to be determined empirically (e.g., see Dewey, 2022) but again illustrates how the current working model leads to new questions and can guide future research.

Finally, one can also question whether the cost of deliberation is factored into our decision to revert to system 1 processing. Imagine that even when we are engaging all our available resources, we still do not manage to resolve a conflict between competing intuitions. What do we do when we do not readily find a solution to a problem? We cannot deliberate forever so at a certain point we need to stop deliberation even when the uncertainty might not have been resolved. Here we presumably need to take the opportunity cost of deliberation into account (e.g., Boureau, Sokol-Hessner, & Daw, 2015; Sirota, Juanchich, & Holford, 2022). Although in a typical experimental study participants only need to focus on the specific reasoning task at hand, in a more ecologically valid environment we always face multiple tasks or challenges. Resources spent on one task, cannot be spend on another one. If another task is more pressing or more rewarding, we may deliberately decide to stop allocating cognitive resources to the current target task. In theory, this opportunity factor may affect the uncertainty parameter. That is, one consequence of not being able to solve a problem is that we may lose interest in it and shift to a different challenge. This may be instantiated by an overall lowering of the activation strength of the intuitions or the inclusion of an opportunity cost factor into the U parameter calculation, for example, which may both decrease the experienced uncertainty. Hence, bluntly put, the longer a

https://doi.org/10.1017/S0140525X2200142X Published online by Cambridge University Press

deliberation process takes, the less we may bother about it. These suggestions are speculative but they illustrate how research on the opportunity cost of deliberation can be integrated into the model.

4.5 Multiple, one, or no intuitions

The current model focuses on the paradigmatic case in which a reasoner is faced with two competing intuitions. As I noted, in theory, the model can be extended to situations in which no, one, or more than two intuitions are cued. In the latter case, the uncertainty parameter might focus on the absolute difference or strength variability of the different intuitions. The more similar in strength they are, the higher the uncertainty. In case there is no intuitive response cued, its strength will obviously be zero. Consequently, the uncertainty will be maximal and the reasoner will be obliged to look for a deliberate response. However, note that in practice, these cases have received little or no empirical testing in dual-process studies. For example, rather than variability per se, uncertainty might be determined by the distance between the strongest intuition and its competitors. Imagine that in a first case three competing intuitions have strength levels 0.9, 0.1, and 0.1, and in a second case 0.9, 0.9, and 0.1. In both cases the average strength deviation (e.g., standard deviation) will be the same but uncertainty and need for deliberate judication might be higher in the second case. Likewise, although it is generally assumed in the dual-process literature that the absence of an intuitive cue will necessarily imply activation of system 2 (e.g., Evans & Stanovich, 2013; Kahneman, 2011; Stanovich, 2011), this activation might also depend on the perceived opportunity cost of deliberation (Shenhav, Prater Fahey, & Grahek, 2021). Future dual-process research will need to pay more empirical attention to these a-typical cases.

Finally, the working model's uncertainty monitoring account also applies when only one intuition is cued. In this case the difference factor will equal the intuition's strength. If the strength is high, the uncertainty will be low and the cued response can be selected without further deliberation. A weaker intuition will result in a higher uncertainty, which increases the likelihood that the deliberation threshold is crossed, and system 2 is called upon. Here the working model fits well with recent accounts that examine the role of metacognition in reasoning (i.e., so-called metareasoning, e.g., Ackerman & Thompson, 2017; see also Baron, 1985, for a related older suggestion). The basic idea is that an intuitive response is always accompanied by an intuitive confidence judgment (i.e., the so-called feeling of rightness, Ackerman & Thompson, 2017). This confidence level would then determine deliberation engagement (i.e., the lower the confidence, the higher the deliberation probability). In essence, this process serves the same role as the uncertainty monitoring in the current working model and it might be worthwhile to integrate the accounts further.

4.6 Links with other fields

Some of the challenges that the working model tries to address show interesting similarities and connections with ongoing developments in other fields such as work on the automatic triggering of cognitive control (e.g., Algom & Chajut, 2019), mental effort allocation (e.g., Kool & Botvinick, 2018; Shenhav et al., 2021), or computational modeling of changes-of-mind in perceptual decision making (e.g., Stone, Mattingley, & Rangelov, 2022; Turner, Feuerriegel, Andrejević, Hester, & Bode, 2021). Although these fields have typically focused on lower-level tasks than dual-process models of reasoning – and have remained somewhat isolated from this literature – the working model might allow us to integrate both which can offer some guidance for the further development of dual-process models of higher-order cognition.[5]

For example, research on the engagement of cognitive control in tasks such as the Stroop (e.g., name the ink color in which a color word is written), has indicated that various processes that had long been considered the hallmark of deliberate controlled processing can also operate automatically (e.g., Desender, Van Lierde, & Van den Bussche, 2013; Jiang, Correa, Geerts, & van Gaal, 2018; Linzarini, Houdé, & Borst, 2017). These findings have resulted in broader theoretical advances that indicate how core control mechanisms can also be achieved through low-level associative mechanisms (Abrahamse, Braem, Notebaert, & Verguts, 2016; Algom & Chajut, 2019; Braem & Egner, 2018). Hence, as in the dual-process literature, there seems to be a tendency to move from an exclusive to a non-exclusive view on elementary control processes (e.g., see also Hassin, 2013, for a related point on conscious and unconscious processing).

Likewise, the field of mental effort allocation has long studied the motivational aspects of deliberate control (e.g., Kool & Botvinick, 2018; Shenhav et al., 2017, 2021). Here, the decision to engage effortful controlled processing in a cognitive task is modeled as a function of the likelihood that allocation of control will result in the desired outcome and the weighing of the costs and benefits of allocating control to the task. Such a framework might be highly relevant for the integration of an opportunity cost factor in dual-process models of reasoning (Sirota et al., 2022).

In the same vein, research on so-called changes-of-mind (Evans, Dutilh, Wagenmakers, & van der Maas, 2020; Resulaj, Kiani, Wolpert, & Shadlen, 2009; Turner et al., 2021; Van Den Berg et al., 2016) can be inspirational. Scholars in this field try to explain when and how participants will revise perceptual decisions (e.g., whether or not a stimulus was perceived). For example, you initially might infer that an "X" was briefly flashed on screen but milliseconds later revise this answer and decide it was a "Y." Various computational models that make differential assumptions about whether an increase in the activation level of one decision automatically implies an activation decrease of its competitor or whether such activation necessarily decays over time, have been developed and can be contrasted (e.g., Pleskac & Busemeyer, 2010; Usher & McClelland, 2001). Integration of this modeling work might be useful for the further fine-grained specification of the intuitive activation component of the working model.

4.7 Computation issues

The present working model is intended to serve as a first, verbal model of core processes and operating principles. It does not present a computational model that specifies how the operations are calculated and what processes ultimately underlie system 1 or the generation of intuitions. However, such a specification or integration is not impossible. For example, Oaksford and Hall (2016) showed how a probabilistic Bayesian approach might in theory be used to model conflict between competing intuitions and the generation of "logical" (or alleged system 2) intuitions in classic reasoning tasks. Oaksford and Hall gave the example of a base-rate neglect task in which base-rate information (e.g., a sample with 995 men and 5 women) can conflict with information

https://doi.org/10.1017/S0140525X2200142X Published online by Cambridge University Press

provided by a stereotypical description (e.g., a randomly drawn individual from the sample is described as someone who likes shopping). Traditionally it is assumed that the description will cue an incorrect intuitive response (i.e., the randomly drawn individual is most likely female) and that taking the base-rate information into account will require system 2 deliberation. Oaksford and Hall demonstrated how both might be done intuitively in system 1 by an unconscious sampling of probability distributions. In a nutshell, probabilities are represented as probability density functions in the model (e.g., Clark, 2013). Different cues in the problem information (e.g., base-rates and the description) will give rise to a probability distribution of possible values. The first cue that is encountered (e.g., base-rates) will give rise to a prior distribution. The second cue (e.g., description) will modify this to a posterior probability distribution. A decision is then made by sampling values from these distributions. In essence, this unconscious process of probability distribution sampling would ultimately underlie system 1 processing. Although such an account would need to be generalized to other tasks and domains, it indicates that a more fine-grained computational account is not a mere promissory note. In theory, the underlying computational model can be specified and tested. This remains an important challenge for the current working model. At the same time, it also underscores the value of a verbal working model. If our theories maintain that a response is out of reach of the intuitive system, there is no point in trying to model how such a response can be intuitively instantiated either.

4.8 Dual schmosses?

This paper pointed out that there is little empirical and conceptual support for foundational dual-process assumptions and presented a revised working model to address these challenges. However, given the empirical and conceptual dual-process issues, one might be tempted to draw a radically different conclusion. That is, rather than to try building a more credible version of the framework, shouldn't we simply abandon the dual-process enterprise of splitting cognition into a fast and slow system altogether? This critique can be read and targeted at multiple levels. First, various scholars have long questioned dual-process models (e.g., Gigerenzer & Regier, 1996; Keren & Schul, 2009; Melnikoff & Bargh, 2018; Osman, 2004). Often this is accompanied by a call to switch to so-called single-process models (e.g., Kruglanski & Gigerenzer, 2011; Osman, 2004). As I noted in the Introduction, both single- and dual-process models focus on the interaction between intuition and deliberation. But whereas single-process proponents believe there is only a quantitative difference between intuition and deliberation (i.e., the difference is one of degree, not kind), dual-process theorists have traditionally argued for a qualitative view on this difference (e.g., see Keren & Schul, 2009, and De Neys, 2021, for reviews). Bluntly put, whereas the qualitative view sees intuition and deliberation as running on different engines, the quantitative view entails they run on one and the same engine that simply operates at different intensities. My main argument was orthogonal to this specific issue and I therefore used the fast-and-slow dual-process label as a general header that covers both the qualitative and quantitative interpretation. The simple reason is that single-process models also differentiate between intuitive and deliberate processing and posit that some responses require more deliberation than others (e.g., Kruglanski & Gigerenzer, 2011). At one point we may be at the intuitive end of the processing scale and will need to decide whether we need to move to the more deliberate end, and invest more time and resources (e.g., whether or not we hit the gas pedal and let the engine run at full throttle). Hence, quantitative single-process models face the same switch issue as their qualitative rivals. Any solution will require them to drop exclusivity and postulate that responses that can be computed when we're at the deliberate extreme of the processing scale, can also be computed when we're at the intuitive end. In short, the issues outlined here are not solved by simply moving from a qualitative to a quantitative single-process view on intuition and deliberation.

Another possible general critique of the dual-process approach has to do with the specific reading of the "system" label (e.g., Oaksford & Chater, 2012). Dual-process models are also being referred to as dual system models. These labels are often used interchangeably (as in the present paper) but sometimes they are used to refer to a specific subclass of models. For example, some dual-process models are more specific in their scope, others more general (Gawronski & Creighton, 2013). The more specific models are developed to account for specific phenomena or tasks, the more general ones are intended to be more integrative and apply to various phenomena. Some authors use the system label to specifically refer to the latter, more general models (e.g., Smith & DeCoster, 2000; Strack & Deutsch, 2004). One critique of dual-process models has to do with this general "system" interpretation. One may argue that although intuitive and deliberate processing in various domains might bear some phenomenological family resemblance, they ultimately share no common core. For example, "system 1" processing in moral reasoning might have nothing to do with "system 1" processing during prosocial decision making or logical reasoning. Hence, rather than positing a general intuitive and deliberate processing type, we may have subsets of more intuitively and deliberately operating processes that are at play in different tasks. This is a valid point but it is ultimately independent of the issue addressed here. That is, even if there are domain- or task-specific intuitive and deliberate processes at play, we still need to explain how we switch from one to the other in the specific task at hand. Hence, the classic "system" view is not the problem here. This does help to underscore that the processing details (e.g., the precise value of the deliberation threshold) of the working model may vary across domains (or even tasks). The point is that its core principles (e.g., non-exclusivity, monitoring, feedback component) will need to apply if we want to account for the switch process in any of these individual domains (or tasks).

Finally, one may also wonder whether the central dual-process switch issue is simply an instantiation of the more general challenge of deciding when to stop a calculation. That is, imagine that all human cognition is deliberative in nature. Even in this case where there is never an intuition/deliberation switch decision to make, we would still need to decide whether to keep on calculating or stop and make a stab at the answer in the light of the deliberate calculations we already made. As I noted (sect. 4.6), this "stop" question is specifically examined in work on mental effort allocation and might be especially useful to integrate an opportunity cost factor into the working model (e.g., Sirota et al., 2022). However, is this all we need? I believe it is important to highlight that dual-process models typically focus on a slightly different situation. That is, rather than deciding whether or not to spend (more) resources to get to an answer per se, they deal with cases in which a plausible, salient answer is intuitively cued from the outset before we spend any effort at all. The question is whether there is a need to go beyond this first hunch. Do we need to start deliberating if we are instantly repulsed by a moral option, feel that it's better to share with others than to make more ourselves, or have a positive first impression of a job candidate? Whether such a switch decision can be accounted for by the same mechanism as the general calculation or deliberation "stopping" machinery is ultimately an empirical question. At the very least it will require us to examine and account for the switching in the specific situations that dual-process models envisage. Developing a revised account that provides a viable specification of the postulated intuition/deliberation switch mechanism should always be useful here. Clearly, if the dual-process model doesn't specify a switch account yet, there is no point in contrasting it with other "switch" approaches. Hence, even if one questions the idea that we can distinguish more intuitive and deliberate processing in human cognition and favors an alternative account, it is paramount to request dual-process theorists to develop the best possible specification of the core "fast-and-slow" switch mechanism. The point is that this will allow for a more informative contrast with possible rival accounts. Put simply, if we want to know whether NFL players have a better physique than basketball players, we should test them against NBA players rather than players from the local recreational team. To avoid any confusion, my point is not that the current working model provides the best possible dual-process specification (or that it's the LeBron James of dual-process theory), but that it is sensible to strive for the best possible version of the framework.

5. Conclusion

In the last 50 years dual-process models of thinking have moved to the center stage in research on human reasoning. These models have been instrumental for the initial exploration of human thinking in the cognitive sciences and related fields (Chater, 2018; De Neys, 2021). However, it is time to rethink foundational assumptions. Traditional dual-process models have typically conceived intuition and deliberation as generating unique responses such that one type of response is exclusively tied to deliberation and is assumed to be beyond the reach of the intuitive system. I reviewed empirical evidence from key dual-process applications that argued against this exclusivity feature. I also showed how exclusivity leads to conceptual complications when trying to explain how a reasoner switches between intuitive and deliberate reasoning. To avoid these complications, I sketched an elementary non-exclusive working model in which it is the activation strength of competing intuitions within system 1 that determines system 2 engagement.

It will be clear that the working model is a starting point that will need to be further developed and specified. However, by avoiding the conceptual paradoxes that plague the traditional model, it presents a more viable basic architecture that can serve as theoretical groundwork to build future dual-process models in various fields. In addition, it should at the very least force dual-process theorists to specify more explicitly how they address the switch issue. In the absence of such specification, dual-process models might continue to provide an appealing narrative but will do little to advance our understanding of the interaction between intuitive and deliberate – fast-and-slow – thinking. It is in this sense that I hope that the present paper can help to sketch the building blocks of a more judicious dual-process future.

Financial support. Writing of this paper was supported by a grant from Agence Nationale de la Recherche (DIAGNOR ANR-16-CE28-0010-01).

Competing interest. None.

Notes

1. I will use "logical" as a general header to refer to logical, probabilistic, and mathematical principles and reasoning.
2. For a recent illustration consider the widespread mistaken belief that Covid-19 vaccines are unsafe because more vaccinated than unvaccinated people are hospitalized (neglecting that the group of vaccinated people is far larger in most Western countries, e.g., Devis, 2021).
3. See the legendary Johnny Cash song "A boy named Sue" (Cash, 1969).
4. This does not imply that these authors agree with or can be held accountable for the claims made here. I simply want to acknowledge that my theorizing does not come out of the blue and was inspired by the thinking of multiple scholars.
5. For example, vice versa this could also help to scale-up models focusing on more elementary low-level cognition tasks to higher-level reasoning about morality, cooperation, and logic.

References

Abrahamse, E., Braem, S., Notebaert, W., & Verguts, T. (2016). Grounding cognitive control in associative learning. *Psychological Bulletin*, *142*, 693.
Achtziger, A., & Alós-Ferrer, C. (2014). Fast or rational? A response-times study of Bayesian updating. *Management Science*, *60*, 923–938.
Ackerman, R., & Thompson, V. A. (2017). Meta-reasoning: Monitoring and control of thinking and reasoning. *Trends in Cognitive Sciences*, *21*, 607–617.
Algom, D., & Chajut, E. (2019). Reclaiming the Stroop effect back from control to input-driven attention and perception. *Frontiers in Psychology*, *10*, 1683.
Bago, B., Bonnefon, J. F., & De Neys, W. (2021). Intuition rather than deliberation determines selfish and prosocial choices. *Journal of Experimental Psychology: General*, *150*, 1081–1094.
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. (2019a). The smart system 1: Evidence for the intuitive nature of correct responding on the bat-and-ball problem. *Thinking & Reasoning*, *3*, 257–299.
Bago, B., & De Neys, W. (2019b). The intuitive greater good: Testing the corrective dual process model of moral cognition. *Journal of Experimental Psychology: General*, *48*, 1782–1801.
Bago, B., & De Neys, W. (2020). Advancing the specification of dual process models of higher cognition: A critical test of the hybrid model view. *Thinking & Reasoning*, *26*, 1–30.
Bago, B., Rand, D. G., & Pennycook, G. (2020). Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of Experimental Psychology: General*, *149*, 1608–1613.
Baron J. (1985). *Rationality and intelligence*. Cambridge University Press.
Baron, J. (2017). Utilitarian vs. deontological reasoning: Method, results, and theory. In J.-F. Bonnefon & B. Trémolière (Eds.), *Moral inferences* (pp. 137–151). Psychology Press.
Baron, J., & Gürçay, B. (2017). A meta-analysis of response-time tests of the sequential two-systems model of moral judgment. *Memory & Cognition*, *45*, 566–575.
Barr, N., Pennycook, G., Stolz, J. A., & Fugelsang, J. A. (2015). Reasoned connections: A dual-process perspective on creative thought. *Thinking & Reasoning*, *21*, 61–75.
Beattie, G. (2012). Psychological effectiveness of carbon labelling. *Nature Climate Change*, *2*, 214–217.
Białek, M., & De Neys, W. (2016). Conflict detection during moral decision-making: Evidence for deontic reasoners' utilitarian sensitivity. *Journal of Cognitive Psychology*, *28*, 631–639.
Białek, M., & De Neys, W. (2017). Dual processes and moral conflict: Evidence for deontological reasoners' intuitive utilitarian sensitivity. *Judgment and Decision Making*, *12*, 148–167.
Boissin, E., Caparos, S., Raoelison, M., & De Neys, W. (2021). From bias to sound intuiting: Boosting correct intuitive reasoning. *Cognition*, *211*, 104645.
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*, 1019–1027.
Boureau, Y. L., Sokol-Hessner, P., & Daw, N. D. (2015). Deciding how to decide: Self-control and meta-decision making. *Trends in Cognitive Sciences*, *19*, 700–710.
Bouwmeester, S., Verkoeijen, P. P., Aczel, B., Barbosa, F., Bègue, L., Brañas-Garza, P., … Wollbrant, C. E. (2017). Registered replication report: Rand, Greene, and Nowak (2012). *Perspectives on Psychological Science*, *12*, 527–542.

Braem, S., & Egner, T. (2018). Getting a grip on cognitive flexibility. *Current Directions in Psychological Science*, *27*, 470–476.
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.
Cash, J. (1969). *A boy named Sue*. Columbia Records.
Chaiken, S., & Trope, Y. (Eds.) (1999). *Dual-process theories in social psychology*. Guilford Press.
Chater, N. (2018). Is the type 1/type 2 distinction important for behavioral policy? *Trends in Cognitive Sciences*, *22*, 369–371.
Chater, N., & Schwarzlose, R. F. (2016). Thinking about thinking: 28 years on. *Trends in Cognitive Sciences*, *20*, 787.
Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*, 181–204.
Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: A process dissociation approach. *Journal of Personality and Social Psychology*, *104*, 216.
Cushman, F. (2020). Rationalization is rational. *Behavioral and Brain Sciences*, *43*, e28.
De Neys, W. (2006a). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science*, *17*, 428–433.
De Neys, W. (2006b). Automatic-heuristic and executive-analytic processing in reasoning: Chronometric and dual task considerations. *Quarterly Journal of Experimental Psychology*, *59*, 1070–1100.
De Neys, W. (2012). Bias and conflict a case for logical intuitions. *Perspectives on Psychological Science*, *7*, 28–38.
De Neys, W. (Ed.) (2017). *Dual process theory 2.0*. Routledge.
De Neys, W. (2020). Morality, normativity, and the good system 2 fallacy. *Diametros*, *17*, 1–6. doi: 10.33392/diam.1447
De Neys, W. (2021). On dual and single process models of thinking. *Perspectives on Psychological Science*, *16*, 1412–1427.
De Neys, W. (2022). The cognitive unconscious and dual process theories of reasoning. In A. S. Reber & R. Allen (Eds.), *The cognitive unconscious: The first fifty years*. Oxford University Press.
De Neys, W., & Bonnefon, J. F. (2013). The whys and whens of individual differences in thinking biases. *Trends in Cognitive Sciences*, *17*, 172–178.
De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition*, *106*, 1248–1299.
De Neys, W., & Pennycook, G. (2019). Logic, fast and slow: Advances in dual-process theorizing. *Current Directions in Psychological Science*, *28*, 503–509.
Desender, K., Van Lierde, E., & Van den Bussche, E. (2013). Comparing conscious and unconscious conflict adaptation. *PLoS ONE* *8*(2), e55976.
Devis, D. (2021). Why are there so many vaccinated people in hospital?. Retrieved from https://cosmosmagazine.com/health/covid/why-are-there-so-many-vaccinated-people-in-hospital/
DeWall, C. N., Baumeister, R. F., Gailliot, M. T., & Maner, J. K. (2008). Depletion makes the heart grow less helpful: Helping as a function of self-regulatory energy and genetic relatedness. *Personality and Social Psychology Bulletin*, *34*, 1653–1662.
Dewey, A. R. (2021). Reframing single- and dual-process theories as cognitive models: Commentary on De Neys (2021). *Perspectives on Psychological Science*, *16*, 1428–1431.
Dewey, A. R. (2022). Metacognitive control in single- vs. dual-process theory. Manuscript submitted for publication.
Djulbegovic, B., Hozo, I., Beckstead, J., Tsalatsanis, A., & Pauker, S. G. (2012). Dual processing model of medical decision-making. *BMC Medical Informatics and Decision Making*, *12*, 94.
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.
Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, *49*, 709–724.
Evans, J. S. B. (2002). Logic and human reasoning: An assessment of the deduction paradigm. *Psychological Bulletin*, *128*(6), 978–996.
Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255–278.
Evans, J. S. B. (2016). Reasoning, biases and dual processes: The lasting impact of Wason (1960). *The Quarterly Journal of Experimental Psychology*, *69*, 2076–2092.
Evans, J. S. B. (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. S. B., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning*, *11*, 382–389.
Evans, J. S. B., & Over, D. E. (1996). *Rationality and reasoning*. Psychology Press.
Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition advancing the debate. *Perspectives on Psychological Science*, *8*, 223–241.
Evans, J. S. B., & Wason, P. C. (1976). Rationalization in a reasoning task. *British Journal of Psychology*, *67*, 479–486.
Evans, J. S. B. T. (2009). How many dual process theories do we need? One, two or many? In J. St. B. T. Evans & K. Frankish (Eds). *In two minds: Dual processes and beyond* (pp. 1–32). Oxford University Press.
Evans, J. S. B. T. (2011). Dual-process theories of reasoning: Contemporary issues and developmental applications. *Developmental Review*, *31*, 86–102.
Evans, N. J., Dutilh, G., Wagenmakers, E. J., & van der Maas, H. L. (2020). Double responding: A new constraint for models of speeded decision making. *Cognitive Psychology*, *121*, 101292.
Frankish, K., & Evans, J. St. B. T. (2009). The duality of mind: An historical perspective. In J. St. B. T. Evans & K. Frankish (Eds.), *In Two minds: Dual processes and beyond* (pp. 1–29). Oxford University Press.
Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*, 25–42.
Frey, D., Johnson, E. D., & De Neys, W. (2018). Individual differences in conflict detection during reasoning. *Quarterly Journal of Experimental Psychology*, *71*, 1188–1208.
Gangemi, A., Bourgeois-Gironde, S., & Mancini, F. (2015). Feelings of error in reasoning – In search of a phenomenon. *Thinking & Reasoning*, *21*(4), 383–396.
Gawronski, B., & Creighton, L. A. (2013). Dual-process theories. In D. E. Carlston (Ed.), *The Oxford handbook of social cognition* (pp. 282–312). Oxford University Press.
Gervais, W. M., & Norenzayan, A. (2012). Analytic thinking promotes religious disbelief. *Science*, *336*, 493–496.
Ghasemi, O., Handley, S., Howarth, S., Newman, I. R., & Thompson, V. A. (2022). Logical intuition is not really about logic. *Journal of Experimental Psychology: General*, *151*(9), 2009–2028.
Gigerenzer, G., & Regier, T. (1996). How do we tell an association from a rule? Comment on Sloman (1996). *Psychological Bulletin*, *119*, 23–26.
Greene, J. (2013). *Moral tribes: Emotion, reason and the gap between us and them*. Penguin Press.
Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences*, *6*, 517–523.
Greene, J. D. (2007). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences*, *11*, 322–323.
Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*, 2105–2108.
Grossman, Z., & Van der Weele, J. J. (2017). Dual-process reasoning in charitable giving: Learning from non-results. *Games*, *8*, 36.
Gürçay, B., & Baron, J. (2017). Challenges for the sequential two-system model of moral judgement. *Thinking & Reasoning*, *23*, 49–80.
Hackel, L. M., Wills, J. A., & Van Bavel, J. J. (2020). Shifting prosocial intuitions: Neurocognitive evidence for a value-based account of group-based cooperation. *Social Cognitive and Affective Neuroscience*, *15*(4), 371–381.
Hassin, R. R. (2013). Yes it can: On the functional abilities of the human unconscious. *Perspectives on Psychological Science*, *8*, 195–207.
Hayes, B. K., Stephens, R. G., Lee, M. D., Dunn, J. C., Kaluve, A., Choi-Christou, J., & Cruz, N. (2022). Always look on the bright side of logic? Testing explanations of intuitive sensitivity to logic in perceptual tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(11), 1598–1617. doi: 10.1037/xlm0001105
Hoerl, C., & McCormack, T. (2019). Thinking in and about time: A dual systems perspective on temporal cognition. *Behavioral and Brain Sciences*, *42*, e244.
Hofmann, W., Friese, M., & Wiers, R. W. (2008). Impulsive versus reflective influences on health behavior: A theoretical framework and empirical review. *Health Psychology Review*, *2*, 111–137.
Houdé, O. (2019). *3-System theory of the cognitive brain: A post-Piagetian approach to cognitive development*. Routledge.
Isler, O., Yilmaz, O., & Maule, J. A. (2021). Religion, parochialism and intuitive cooperation. *Nature Human Behaviour*, *5*, 512–521.
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. (2000). A psychological point of view: Violations of rational rules as a diagnostic of mental processes. *Behavioral and Brain Sciences*, *23*, 681–683.
Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
Kaufman, S. B. (2011). Intelligence and the cognitive unconscious. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 442–467). 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.
Kessler, J., Kivimaki, H., & Niederle, M. (2017). Thinking fast and slow: Generosity over time. Retrieved from http://assets.wharton.upenn.edu/~juddk/papers/KesslerKivimakiNiederle_GenerosityOverTime.pdf
Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science*, *314*, 829–832.
Kool, W., & Botvinick, M. (2018). Mental labour. *Nature Human Behaviour*, *2*, 899–908.

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

Kvarven, A., Strømland, E., Wollbrant, C., Andersson, D., Johannesson, M., Tinghög, G., … Myrseth, K. O. R. (2020). The intuitive cooperation hypothesis revisited: A meta-analytic examination of effect size and between-study heterogeneity. *Journal of the Economic Science Association*, *6*, 26–42.

Lawson, M. A., Larrick, R. P., & Soll, J. B. (2020). Comparing fast thinking and slow thinking: The relative benefits of interventions, individual differences, and inferential rules. *Judgment and Decision Making*, *15*, 660–684.

Lemir, J. (2021). *This book is not about baseball but baseball teams swear by it*. Retrieved from https://www.nytimes.com/2021/02/24/sports/baseball/thinking-fast-and-slow-book.html.

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.

Martinsson, P., Myrseth, K. O. R., & Wollbrant, C. (2014). Social dilemmas: When self-control benefits cooperation. *Journal of Economic Psychology*, *45*, 213–236.

Martiny-Huenger, T., Bieleke, M., Doerflinger, J., Stephensen, M. B., & Gollwitzer, P. M. (2021). Deliberation decreases the likelihood of expressing dominant responses. *Psychonomic Bulletin & Review*, *28*, 139–157.

Mata, A. (2020). Conflict detection and social perception: Bringing meta-reasoning and social cognition together. *Thinking & Reasoning*, *26*(1), 140–149.

Mata, A., Ferreira, M. B., Voss, A., & Kollei, T. (2017). Seeing the conflict: An attentional account of reasoning errors. *Psychonomic Bulletin & Review*, *24*(6), 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.

Melnikoff, D. E., & Bargh, J. A. (2018). The mythical number two. *Trends in Cognitive Sciences*, *22*, 280–293.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*, 57–74.

Meyer-Grant, C. G., Cruz, N., Singmann, H., Winiger, S., Goswami, S., Hayes, B. K., & Klauer, K. C. (2022). Are logical intuitions only make-believe? Reexamining the logic-liking effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0001152

Milkman, K. L., Chugh, D., & Bazerman, M. H. (2009). How can decision making be improved? *Perspectives on Psychological Science*, *4*, 379–383.

Moore, A. B., Clark, B. A., & Kane, M. J. (2008). Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgment. *Psychological Science*, *19*, 549–557.

Morewedge, C. K., & Kahneman, D. (2010). Associative processes in intuitive judgment. *Trends in Cognitive Sciences*, *14*(10), 435–440.

Newman, I. R., Gibb, M., & Thompson, V. A. (2017). Rule-based reasoning is fast and belief-based reasoning can be slow: Challenging current explanations of belief-bias and base-rate neglect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*, 1154.

Oaksford, M., & Chater, N. (2012). Dual processes, probabilities, and cognitive architecture. *Mind & Society*, *11*, 15–26.

Oaksford, M., & Hall, S. (2016). On the source of human irrationality. *Trends in Cognitive Sciences*, *20*, 336–344.

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

Paxton, J. M., Ungar, L., & Greene, J. D. (2012). Reflection and reasoning in moral judgment. *Cognitive Science*, *36*, 163–177.

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). Routledge.

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.

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

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review*, *117*, 864–901.

Purcell, Z. A., Wastell, C. A., & Sweller, N. (2020). Domain-specific experience and dual-process thinking. *Thinking & Reasoning*, *27*, 239–267.

Rand, D. G. (2019). Intuition, deliberation, and cooperation: Further meta-analytic evidence from 91 experiments on pure cooperation. Available at SSRN 3390018. Retrieved from http://dx.doi.org/10.2139/ssrn.3390018

Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature*, *489*, 427–430.

Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., & Greene, J. D. (2014). Social heuristics shape intuitive cooperation. *Nature Communications*, *5*, 1–12.

Raoelison, M., Boissin, E., Borst, G., & De Neys, W. (2021). From slow to fast logic: The development of logical intuitions. *Thinking & Reasoning*, *27*, 599–622.

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 University Press.

Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changing your mind: A computational mechanism of vacillation. *Nature*, *461*(7261), 263.

Reyna, V. F., Rahimi-Golkhandan, S., Garavito, D. M. N., & Helm, R. K. (2017). The fuzzy-trace dual-process model. In W. De Neys (Ed.), *Dual process theory 2.0* (pp. 90–107). Routledge.

Robison, M. K., & Unsworth, N. (2017). Individual differences in working memory capacity and resistance to belief bias in syllogistic reasoning. *Quarterly Journal of Experimental Psychology*, *70*, 1471–1484.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, *300*, 1755–1758.

Schubert, A. L., Ferreira, M. B., Mata, A., & Riemenschneider, B. (2021). A diffusion model analysis of belief bias: Different cognitive mechanisms explain how cognitive abilities and thinking styles contribute to conflict resolution in reasoning. *Cognition*, *211*, 104629.

Shefrin, H. (2013). Advice for CFOs: Beware of fast thinking. Retrieved from https://www.wsj.com/articles/SB10001424127887324299104578531561852942612

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience*, *40*, 99–124.

Shenhav, A., Prater Fahey, M., & Grahek, I. (2021). Decomposing the motivation to exert mental effort. *Current Directions in Psychological Science*, *30*, 307–314.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*, 127–190.

Sirota, M., Juanchich, M., & Holford, D. L. (2022). Rationally irrational: When people do not correct their reasoning errors even if they could. Manuscript submitted for publication.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*(1), 3–22.

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, *4*, 108–131.

Šrol, J., & De Neys, W. (2021). Predicting individual differences in conflict detection and bias susceptibility during reasoning. *Thinking and Reasoning*, *27*, 38–68.

Stanovich, K. (2011). *Rationality and the reflective mind*. Oxford University Press.

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

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General*, *127*, 161.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences*, *23*, 645–665.

Stone, C., Mattingley, J. B., & Rangelov, D. (2022). On second thoughts: Changes of mind in decision-making. *Trends in Cognitive Sciences*, *26*, 419–431.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*, 220–247.

Sunstein, C. R. (2020). The cognitive bias that makes us panic about coronavirus. Retrieved from https://www.bloomberg.com/opinion/articles/2020-02-28/coronavirus-panic-caused-by-probability-neglect

Tett, G. (2021). Mood and emotion are driving market swings. Retrieved from https://www.ft.com/content/89c95e78-ec7f-4d06-9ea8-a9b19a5ed6da

Thompson, V., & Newman, I. (2017). Logical intuitions and other conundra for dual process theories. In W. De Neys (Ed.), *Dual process theory 2.0* (pp. 121–136). Routledge.

Thompson, V. A. (2021). Eye-tracking IQ: Cognitive capacity and strategy use on a ratio-bias task. *Cognition*, *208*, 104523.

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

Thompson, V. A., & Markovits, H. (2021). Reasoning strategy vs cognitive capacity as predictors of individual differences in reasoning performance. *Cognition*, *217*, 104866.

Thompson, V. A., Pennycook, G., Trippas, D., & Evans, J. S. B. (2018). Do smart people have better intuitions? *Journal of Experimental Psychology: General*, *147*, 945–961.

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

Tinghög, G., Andersson, D., Bonn, C., Johannesson, M., Kirchler, M., Koppel, L., & Västfjäll, D. (2016). Intuition and moral decision-making – The effect of time pressure and cognitive load on moral judgment and altruistic behavior. *PLoS ONE*, *11*, e0164012.

Travers, E., Rolison, J. J., & Feeney, A. (2016). The time course of conflict on the cognitive reflection test. *Cognition*, *150*, 109–118.

Trémolière, B., De Neys, W., & Bonnefon, J. F. (2012). Mortality salience and morality: Thinking about death makes people less utilitarian. *Cognition*, *124*, 379–384.

https://doi.org/10.1017/S0140525X2200142X Published online by Cambridge University Press

Trémolière, B., De Neys, W., & Bonnefon, J. F. (2019). Reasoning and moral judgment: A common experimental toolbox. In L. J. Ball & V. A. Thompson (Eds.), *International handbook of thinking and reasoning* (pp. 575–590). Psychology Press.
Trippas, D., & Handley, S. (2017). The parallel processing model of belief bias: Review and extensions. In W. De Neys (Ed.), *Dual process theory 2.0*. Routledge.
Turner, W., Feuerriegel, D., Andrejević, M., Hester, R., & Bode, S. (2021). Perceptual change-of-mind decisions are sensitive to absolute evidence magnitude. *Cognitive Psychology*, *124*, 101358.
Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*, 550–592.
Van Den Berg, R., Anandalingam, K., Zylberberg, A., Kiani, R., Shadlen, M. N., & Wolpert, D. M. (2016). A common mechanism underlies changes of mind about decisions and confidence. *eLife*, *5*, e12192.
Vartanian, O., Beatty, E. L., Smith, I., Blackler, K., Lam, Q., Forbes, S., & De Neys, W. (2018). The reflective mind: Examining individual differences in susceptibility to base rate neglect with fMRI. *Journal of Cognitive Neuroscience*, *30*, 1011–1022.
Vega, S., Mata, A., Ferreira, M. B., & Vaz, A. R. (2021). Metacognition in moral decisions: Judgment extremity and feeling of rightness in moral intuitions. *Thinking & Reasoning*, *27*, 124–141.
Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, *12*, 129–140.
Wason, P. C., & Evans, J. S. B. (1975). Dual processes in reasoning? *Cognition*, *3*, 141–154.
Wiesmann, C. G., Friederici, A. D., Singer, T., & Steinbeis, N. (2020). Two systems for thinking about others' thoughts in the developing brain. *Proceedings of the National Academy of Sciences*, *117*, 6928–6935.
World Bank Group (2015). *World development report: Mind, society, and behavior*. World Bank.

Open Peer Commentary

We know what stops you from thinking forever: A metacognitive perspective

Rakefet Ackerman[a] and Kinga Morsanyi[b]

[a]Faculty of Data and Decision Sciences, Technion – Israel Institute of Technology, Haifa, Israel and [b]Centre for Mathematical Cognition, Loughborough University, Loughborough, UK
ackerman@technion.ac.il; technion.ac.il
k.e.morsanyi@lboro.ac.uk; lboro.ac.uk

doi:10.1017/S0140525X22003065, e112

Abstract

This commentary addresses omissions in De Neys's model of fast-and-slow thinking from a metacognitive perspective. We review well-established meta-reasoning monitoring (e.g., confidence) and control processes (e.g., rethinking) that explain mental effort regulation. Moreover, we point to individual, developmental, and task design considerations that affect this regulation. These core issues are completely ignored or mentioned in passing in the target article.

This commentary addresses several major omissions in De Neys's "working model." We predominantly focus on gaps in the conceptualization of the "switch feature" and stopping deliberative processes (S2).

Metacognitive research deals with the monitoring and control of thinking processes (Nelson & Narens, 1990). More than 30 years of research have dealt with the processes that inform subjective assessments of success (e.g., confidence) and the subsequent decisions (e.g., to rethink, see Fiedler, Ackerman, & Scarampi, 2019). Of particular relevance is the *meta-reasoning* framework (Ackerman & Thompson, 2017), which is mentioned briefly in section 4.4. By using well-established metacognitive concepts, this framework opens the "black box" of mental effort regulation. It details monitoring and control processes that take place in the early intuitive reasoning stages (S1) separately from the deliberative stages (S2), including processes discussed in the target article and more.

First, the processes covered by the "switch feature" are discussed in length in the literature initiated by Thompson, Prowse Turner, and Pennycook (2011) using the two-response paradigm with *feeling of rightness* judgment (FOR, mentioned in sect. 4.4; Ackerman & Thompson, 2017). FOR is the metacognitive judgment that accompanies the initial response that comes to mind. It has been considered to *trigger the switch between S1 and S2* and found to predict S2 engagement (e.g., Thompson et al., 2013).

A further issue is that the proposed model is incomplete in that the alleged "switch mechanism" is considered to depend entirely on the relative activation levels of competing intuitions and the mysterious "deliberation threshold." In fact, a variety of *situational and personal factors* have been found to affect metacognitive control decisions, such as reasoning time and response choice. Specifically, task design, such as instructions to reason logically (e.g., Ferreira, Garcia-Marques, Sherman, & Sherman, 2006; Morsanyi, Primi, Chiesi, & Handley, 2009), cognitive load (De Neys, 2006; Morsanyi, Busdraghi, & Primi, 2014), and time pressure (Sidi, Shpigelman, Zalmanov, & Ackerman, 2017), as well as individual characteristics, such as thinking dispositions (Cacioppo, Petty, Feinstein, & Jarvis, 1996), cognitive ability (e.g., Stanovich & West, 2000), task-relevant knowledge (e.g., Chiesi, Primi, & Morsanyi, 2011; Stanovich & West, 2008), and anxiety levels (e.g., Beilock & DeCaro, 2007; Primi, Donati, Chiesi, & Morsanyi, 2018) affect reasoning time and response choice. Thus, any model explaining the "switch feature" should incorporate and account for the contextual and individual factors that influence the reasoning process.

Second, the target article discusses stopping deliberative processes (S2) and reverting to S1. An overlooked issue, though, is *when to stop S2 and provide a response*. Within the metacognitive literature, several models address stopping effortful thinking: The discrepancy reduction models (Nelson & Narens, 1990), the region of proximal learning (Metcalfe & Kornell, 2005), and the diminishing criterion model (DCM, Ackerman, 2014; see Ackerman, Yom-Tov, & Torgovitsky, 2020, for a review). According to the most recent model, the DCM, stopping thinking efforts is guided by a combination of two stopping criteria: (a) Confidence in each considered answer is compared to a desired confidence level. Importantly, this stopping criterion dynamically drops as people deliberate longer, reflecting compromising on expected success. (b) A time limit for thinking about each task item, beyond which people are reluctant to think any further (see also Hawkins & Heathcote, 2021).

Third, based on the suggested model, "System 2 deliberation will extend for as long as the uncertainty remains above the threshold" (target article, sect. 3.4, para. 3). Thus, under substantial uncertainty people are *doomed to think forever*. Nevertheless, a totally overlooked aspect is when *people opt out* (e.g., "I don't know") or turn to external help (see Ackerman, 2014; Undorf, Livneh, & Ackerman, 2021). In particular, considering children and novices brings to the fore that people looking at unfamiliar problems may not have any available heuristics to activate. Developmentally, there is a blurry line between deliberative and intuitive processes (Osman & Stavy, 2006) in that responses that can be given quasi automatically by adults may require cognitive effort for children (Morsanyi & Handley, 2008) and may become established by learning (Fischbein, 1987; Gauvrit & Morsanyi, 2014). De Neys briefly considers lack of S1 response (sect. 2.1.5). Another possibility is that people may activate a series of distantly related heuristics, but none of these would be sufficiently strong to offer an answer. In contrast, according to the DCM, when people get to a pre-set time limit, they may prefer opting out over providing a low confidence response. This topic was discussed in metacognitive research already in the 1990s (Koriat & Goldsmith, 1996) and was further developed since then (see Undorf et al., 2021). Thus, there are processes that prevent people from thinking forever.

Fourth, De Neys asks in the Introduction "how do we know that we can rely on an intuitively cued problem solution" (target article, para. 4 in the Introduction) and mentions that "the internal switch decision is itself intuitive in nature" (target article, para. 4 in the Introduction). In metacognitive terms, these intuitions are based on heuristic cues that underlie all metacognitive judgments (Koriat, 1997). Metacognitive judgments *combine* an extensive amount of features (Undorf & Bröder, 2021), including individual self-perceptions and beliefs ("beyond my expertise"), task characteristics (time pressure), and item characteristics (conclusion believability) that may influence, and sometimes mislead, metacognitive judgments (see Ackerman, 2019). Given the wide-spread biases in judgments like FOR and confidence (Thompson et al., 2013), considering potential misleading factors must be incorporated in any model of switch and stopping mechanisms.

Finally, from a developmental perspective, adults have a larger repertoire of heuristics and better ability to integrate them into their cognitive and metacognitive processes than children (Koriat, Ackerman, Adiv, Lockl, & Schneider, 2014). However, in the proposed model, the more heuristics are considered, the longer the thinking process that deals with potential conflicts among them. This contrasts with the traditional role assigned to reasoning heuristics – that they offer immediately available (and highly compelling) responses *immediately* (e.g., Evans, 2006), which is why they are considered to be adaptive and essential parts of the cognitive architecture.

In sum, the proposed model ignores well-established bodies of literature that address the central issues it was meant to cover. Particularly, metacognitive research offers switch and stopping rules, heuristic processes, individual characteristics, and developmental trajectories required for describing the complex processes underlying reasoning.

Financial support. This work was funded by the Israel Science Foundation (grant reference: 234/18) to R.A. and by the Economic and Social Research Council, UK (grant reference: ES/W002914/1) to K.M.

Competing interest. None.

References

Ackerman, R. (2014). The diminishing criterion model for metacognitive regulation of time investment. *Journal of Experimental Psychology: General*, *143*(3), 1349–1368.

Ackerman, R. (2019). Heuristic cues for meta-reasoning judgments: Review and methodology. *Psychological Topics*, *28*(1), 1–20.

Ackerman, R., & Thompson, V. (2017). Meta-reasoning: Monitoring and control of thinking and reasoning. *Trends in Cognitive Sciences*, *21*(8), 607–617.

Ackerman, R., Yom-Tov, E., & Torgovitsky, I. (2020). Using confidence and consensuality to predict time invested in problem solving and in real-life web searching. *Cognition*, *199*, 104248.

Beilock, S. L., & DeCaro, M. S. (2007). From poor performance to success under stress: Working memory, strategy selection, and mathematical problem solving under pressure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(6), 983–998.

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, *119*, 197–253.

Chiesi, F., Primi, C., & Morsanyi, K. (2011). Developmental changes in probabilistic reasoning: The role of cognitive capacity, instructions, thinking styles, and relevant knowledge. *Thinking & Reasoning*, *17*(3), 315–350.

De Neys, W. (2006). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science*, *17*(5), 428–433.

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

Ferreira, M. B., Garcia-Marques, L., Sherman, S. J., & Sherman, J. W. (2006). Automatic and controlled components of judgment and decision making. *Journal of Personality and Social Psychology*, *91*, 797–813.

Fiedler, K., Ackerman, R., & Scarampi, C. (2019). Metacognition: Monitoring and controlling one's own knowledge, reasoning and decisions. In R. J. Sternberg & J. Funke (Eds.), *Introduction to the psychology of human thought* (pp. 89–111). Heidelberg University Publishing.

Fischbein, E. (1987). *Intuition in science and mathematics*. Reidel.

Gauvrit, N., & Morsanyi, K. (2014). The equiprobability bias from a mathematical and psychological perspective. *Advances in Cognitive Psychology*, *10*, 119–130.

Hawkins, G. E., & Heathcote, A. (2021). Racing against the clock: Evidence-based versus time-based decisions. *Psychological Review*, *128*(2), 222–263.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*, 349–370.

Koriat, A., Ackerman, R., Adiv, S., Lockl, K., & Schneider, W. (2014). The effects of goal-driven and data-driven regulation on metacognitive monitoring during learning: A developmental perspective. *Journal of Experimental Psychology: General*, *143*(1), 386–403.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, *103*(3), 490–517.

Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language*, *52*(4), 463–477.

Morsanyi, K., Busdraghi, C., & Primi, C. (2014). Mathematical anxiety is linked to reduced cognitive reflection: A potential road from discomfort in the mathematics classroom to susceptibility to biases. *Behavioral and Brain Functions 2014*, 10, 31.

Morsanyi, K., & Handley, S. J. (2008). How smart do you need to be to get it wrong? The role of cognitive capacity in the development of heuristic-based judgment. *Journal of Experimental Child Psychology*, *99*, 18–36.

Morsanyi, K., Primi, C., Chiesi, F., & Handley, S. J. (2009). The effects and side-effects of statistics education. Psychology students' (mis-)conceptions of probability. *Contemporary Educational Psychology*, *34*, 210–220.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 26, pp. 125–173). Academic Press.

Osman, M., & Stavy, M. O. R. (2006). Development of intuitive rules: Evaluating the application of the dual-system framework to understanding children's intuitive reasoning. *Psychonomic Bulletin & Review*, *13*, 935–953.

Primi, C., Donati, M., Chiesi, F., & Morsanyi, K. (2018). Are there gender differences in cognitive reflection? Invariance and differences related to mathematics. *Thinking & Reasoning*, *24*, 258–279.

Sidi, Y., Shpigelman, M., Zalmanov, H., & Ackerman, R. (2017). Understanding metacognitive inferiority on screen by exposing cues for depth of processing. *Learning and Instruction*, *51*, 61–73.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*(5), 645–665.

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology*, *94*(4), 672–695.

Thompson, V., Prowse Turner, J., Pennycook, G., Ball, L., Brack, H., Ophir, Y., & Ackerman, R. (2013). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition*, *128*, 237–251.

Thompson, V. A., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology*, *63*(3), 107–140.

Undorf, M., & Bröder, A. (2021). Metamemory for pictures of naturalistic scenes: Assessment of accuracy and cue utilization. *Memory & Cognition*, *49*(7), 1405–1422.
Undorf, M., Livneh, I., & Ackerman, R. (2021). Metacognitive control processes in question answering: Help seeking and withholding answers. *Metacognition & Learning*, *16*, 431–458.

"Switching" between fast and slow processes is just reward-based branching

George Ainslie

Department of Veterans Affairs, Coatesville, PA, USA
ga@picoeconomics.org
www.picoeconomics.org

doi:10.1017/S0140525X22002990, e113

Abstract

Shortcuts to goals are rewarded by faster attainment and punished by more frequent failure, so selection of the various kinds – heuristics, cached sequences (habits or *macros*), gut instincts – depends on reward history just like other kinds of choice. The speeds of shortcuts lie on continua along with speeds of deliberation, and these continua have no obvious separation points.

This target article (TA) follows on De Neys's recent proposal "that trying to answer the core single vs dual process model debate is pointless for empirical scientists" so it is "time to move on" (De Neys, 2021, Introduction). What he proposes in the TA is "a more viable dual process architecture" (target article Abstract), which is "orthogonal [to] whether the difference between the two types of processing should be conceived as merely quantitative or qualitative." Nevertheless, he argues for two qualitatively different processes, perhaps characterized by the 14 different properties he listed in the earlier article (fast, effortless, affective, automatic… vs. slow, effortful, affectless, controlled…; De Neys, 2021, p. 4), which he calls here simply fast and slow. He demonstrates the flaws in dual-process theories' usual assumptions: that the two processes must operate separately ("exclusivity"), and that there must be a "switch feature… by which a reasoner can decide to shift between more intuitive and deliberate processing"; but he pulls back from other authors' proposal that "we simply abandon the dual process enterprise." His refutation of the authors who have favored a single, quantitatively based decision process is just to point out that "some responses require more deliberation than others," which would not seem to require a dichotomy.

Dual-process models have admittedly been popular over the years, beginning with Plato's wild versus well-behaved chariot horses. In addition to De Neys's fast versus slow examples, choice making has been described as passionate versus reasonable, impulsive versus reflective, myopic versus far-sighted, hot versus cool, and model-free versus model-based, among others. De Neys also includes as fast the products of "automatization," by which repeated sequences of choices "will be elicited intuitively." In addition, brain imaging has found evidence for steep-discounting versus shallow-discounting brain centers (McClure et al., 2004; van den Bos & McClure, 2013).

However, as De Neys himself concludes, there is no operation that system 2 can perform that system 1 cannot, and "thinking always involves a continuous interaction between system 1 and system 2 application" (target article, sect. 3.4, para. 5). Other authors have pointed out obvious problems with the dual approach: If there are distinct systems, there must be more than two of them, because the properties attributed to the two systems do not reliably occur together (Zbrodoff & Logan, 1986); in particular, automatic processes may or may not be affectively arousing (Ainslie, 2021). Furthermore, the listed properties such as effort, affect, and speed itself are themselves continua. If the two whole lists of properties really constitute discrete systems, there should be natural breaks in the continua from fast to slow, and the breaks should occur at equivalent levels in the n dimensions. As a negative example, the only obvious break in transparency would be too-fast-to-introspect versus not-too-fast-to-introspect, which would not define different kinds.

Most "type 1" processing in humans comprises sequences that have been automatized, macros (or habits) that call up other macros. In language, a squiggly line is interpreted as a letter, a sequence of letters is interpreted as a word, a series of words forms a concept (or cliché). All highly automatized; but if I was to find an anomaly – no, it should be "were to find an anomaly" – my ear would be quick to re-set it. This should not require a distinct system. Even if I stopped to ponder the use of the subjunctive, I would just be trying out sequences I had previously automatized. Likewise, as my calculation proceeds from 2 + 2 through, say 8 + 8, to 64 + 64, and so forth, at some points my mind will pause to find component automatizations; but is there a point where the pause divides two systems?

The strongest case for separate processes might be based on the activities of separate sites in the brain, but even here true separation is doubtful. The dorsolateral striatum (putamen) is differentially active when repeated connections have been cached to form macros, whereas the dorsomedial striatum (caudate) is more active during flexible behavior; but their functioning has been observed to be integrally combined (Dolan & Dayan, 2013; Keramati et al., 2016). Similarly, the existence of separate steep and shallow reward discount centers in the brain is controversial (Kable & Glimcher, 2007; Lempert et al., 2019). If there do exists anatomically separate response-selection systems in the brain, the best candidates would be those for motivational salience and (supposedly separate) reward, governing the attraction of attention and behavioral approach/avoidance, respectively (Berridge & Robinson, 1998). But even here, salience and behavior selection are correlated with activity in mostly the same brain regions (Kim et al., 2021); and when even threatening stimuli are voluntarily gated out, attention to them must have been weighed in the common marketplace of reward (see Ainslie, 2009).

The professed scope of the TA's model is universal, but except for its reference to cupcakes its examples are cognitive searches for correct solutions to puzzles, rather than choices among competing rewards. Accordingly, "the peak activation strength of an intuition reflects how automatized or instantiated the underlying knowledge structures are (i.e., how strongly it is tied to its eliciting stimulus)." This rather Pavlovian convention hampers the model's application to goal-directed activities. By contrast, it is feasible to model the selection of all learnable processes which can replace each other using the amount and timing of their contingent reward (Ainslie, 2017). The sources of reward – consumption goods, ethical goods, social cues, puzzle solutions, signal detections, emotions, the satisfaction of urges – as well as their speeds of onset, are miscellaneous. It should not matter that some of their subroutines involve particular parts of the brain (for instance the amygdala – Aqino et al., 2020 – or hippocampus – Gauthier & Tank, 2018), as long as their weights are ultimately comparable to each other. Likewise, the weighing process may or may not involve a specific site, such as the orbitofrontal cortex (Bartra et al., 2013; Levy & Glimcher, 2012), a set of interacting sites (Krönke et al., 2020), or no identifiable dwelling place (Dohmatob, Dumas, & Bzdok, 2020). In reward research, the adoption of millisecond-specific electroencephalography (for instance, Sambrook et al., 2018) promises to give precise evidence about branching to fast, slow, and intermediate processes.

Financial support. This material is the result of work supported with resources and the use of facilities at the Department of Veterans Affairs Medical Center, Coatesville, PA, USA. The opinions expressed are not those of the Department of Veterans Affairs or of the US Government.

Competing interest. None.

References

Ainslie, G. (2009). Pleasure and aversion: Challenging the conventional dichotomy. *Inquiry: A Journal of Medical Care Organization, Provision and Financing*, *52*(4), 357–377. http://dx.doi.org/10.1080/00201740903087342

Ainslie, G. (2017). *De gustibus disputare:* Hyperbolic delay discounting integrates five approaches to choice. *Journal of Economic Methodology 24*(2), 166–189. http://dx.doi.org/10.1080/1350178X.2017.1309748

Ainslie, G. (2021). Reply to commentaries to "willpower with and without effort." *Behavioral and Brain Sciences 44*, E57. https://doi.org/10.1017/s0140525x21000029

Aquino, T. G., Minxha, J., Dunne, S., Ross, I. B., Mamelak, A. N., Rutishauser, U., & O'Doherty, J. P. (2020). Value-related neuronal responses in the human amygdala during observational learning. *Journal of Neuroscience*, *40*(24), 4761–4772.

Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage*, *76*, 412–427.

Berridge, K. C., & Robinson, T. (1998). What is the role of dopamine in reward: Hedonic impact, reward learning, or incentive salience. *Brain Research Reviews*, *28*, 309–369.

De Neys, W. (2021). On dual- and single-process models of thinking. *Perspectives on Psychological Science*, *16*(6), 1412–1427.

Dohmatob, E., Dumas, G., & Bzdok, D. (2020). Dark control: The default mode network as a reinforcement learning agent. *Human Brain Mapping*, *41*(12), 3318–3341.

Dolan, R. J., & Dayan, P. (2013). Goals and habits in the brain. *Neuron*, *80*(2), 312–325.

Gauthier, J. L., & Tank, D. W. (2018). A dedicated population for reward coding in the hippocampus. *Neuron*, *99*(1), 179–193.

Kable, J. W., & Glimcher, P. W. (2007) The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience 10*, 1625–1633.

Keramati, M., Smittenaar, P., Dolan, R. J., & Dayan, P. (2016). Adaptive integration of habits into depth-limited planning defines a habitual-goal-directed spectrum. *Proceedings of the National Academy of Sciences*, *113*(45), 12868–12873.

Kim, H., Nanavaty, N., Ahmed, H., Mathur, V. A., & Anderson, B. A. (2021). Motivational salience guides attention to valuable and threatening stimuli: Evidence from behavior and functional magnetic resonance imaging. *Journal of Cognitive Neuroscience*, *33*(12), 2440–2460.

Krönke, K. M., Wolff, M., Shi, Y., Kräplin, A., Smolka, M. N., Bühringer, G., & Goschke, T. (2020). Functional connectivity in a triple-network saliency model is associated with real-life self-control. *Neuropsychologia*, *149*, 107667.

Lempert, K. M., Steinglass, J. E., Pinto, A., Kable, J. W., & Simpson, H. B. (2019). Can delay discounting deliver on the promise of RDoC?. *Psychological Medicine*, *49*(2), 190–199.

Levy, D. J., & Glimcher, P. W. (2012). The root of all value: A neural common currency for choice. *Current Opinion in Neurobiology*, *22*(6), 1027–1038.

McClure, S. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2004). The grasshopper and the ant: Separate neural systems value immediate and delayed monetary rewards. *Science (New York, N.Y.) 306*, 503–507.

Sambrook, T. D., Hardwick, B., Wills, A. J., & Goslin, J. (2018). Model-free and model-based reward prediction errors in EEG. *NeuroImage*, *178*, 162–171.

van den Bos, W., & McClure, S. M. (2013). Towards a general model of temporal discounting. *Journal of the Experimental Analysis of Behavior*, *99*(1), 58–73.

Zbrodoff, N. J., & Logan, G. D. (1986). On the autonomy of mental processes: A case study of arithmetic. *Journal of Experimental Psychology: General*, *115*(2), 118.

Individual differences and multi-step thinking

Jonathan Baron

Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA
jonathanbaron7@gmail.com

doi:10.1017/S0140525X2200320X, e114

Abstract

Deliberative thinking often consists of several steps, each involving a switch decision. These decisions may be influenced by confidence in the thinking done so far. Individuals may differ in their tolerance of low confidence and thus may arrive at unjustified high confidence too soon, either from trusting their intuition or by bolstering an initially favored conclusion.

Here I elaborate on some of the points that De Neys discussed only briefly. These concern the nature of deliberative thinking, and the sources of individual differences. This comment is largely a summary of some assertions that I have elaborated elsewhere, and some of them are speculative (Baron, 2019; Baron, Isler, & Yilmaz, in press).

On the first point, the division of thinking into these two systems fits well in explaining most of the laboratory experiments discussed, where "deliberation" simply means that some controlled (i.e., not automatic or immediately intuitive) deliberation is going on. But this deliberation often involves a series of steps, each of which may draw on some automatic/intuitive processes (Ackerman & Thompson, 2017). The steps may be understood as consisting of search and inference (Baron, 1985). The goal is to find the best possibility, the best answer to the question that led to the thinking. Each step may involve the addition or deletion of a possibility from a short list of candidates (which may start with none, in the case of stumper problems), the search for of relevant evidence or arguments bearing on the strength of the various possibilities, and, in many cases, the search for additional goals.

At each step, the thinker makes inferences about the strength of each possibility. Thus, at each step, the think must make another "switch" decision, namely, whether to produce the current strongest possibility as the answer or to continue searching and making inferences. This cycle of search, inference, and deciding whether to continue is clear in such ordinary tasks as consumer purchases, some of which may require days or weeks of deliberation (e.g., buying a house). It is also part of most real-life problem solving, where one possibility is something like, "give up trying to fix it yourself and call the electrician." And it is part of thinking about moral and political issues, thinking that often occurs on the scale of years.

We can think of the switch decision at the end of each step (including the first, which is the focus of the target article) as based on a summary measure of "confidence" in the results so far (essentially the "feeling of rightness" described by Ackerman and Thompson, 2017). Confidence will be high when one possibility is very strong and the others are weak. Strength of the favored option is itself a function of how the thinking done so far was done, and how the thinker responds to the various determinants. Individuals may differ in how they respond, for example:

(1) A thinker trusts her intuition. Her confidence in the initial intuitive response may be high enough to stop at the end of the first step, thus not making the switch that De Neys discusses.
(2) A thinker accepts the standards of "actively open-minded thinking" (AOT; Baron, 2019; Baron et al., in press). Possibilities will not be considered strong unless a search has been made for other possibilities and for evidence both favoring and opposing the initially favored possibility (and for possible goals that were neglected so far).
(3) A thinker begins with low strength but suffers from "uncertainty aversion." Uncertainty in this case can be result of not doing much thinking. To remove the uncertainty, the thinker searches for evidence favoring the initial intuition, bolstering it, so that its strength is artificially high. This bolstering leads to the sorts of apparent failures of system 2 that are noted in the target article.

Thinking does not always have to proceed to get a conclusion with high confidence. It is often reasonable to stop thinking just because thinking is not making progress or because the answer is not worth more time and effort. At this point, a honest answer to a question about confidence, without self-deception, would be that confidence is low. Scientists, when speaking to the public, often qualify their statements with expressions of low confidence. People with uncertainty aversion could think that the scientists are bad thinkers; these people think that good thinkers should always be confident (and that is also why they are inclined to bolster their own confidence, when that is needed).

Alternatively, when thinking is not making progress, it is often reasonable to "outsource" it: for example, consult a professional. In matters like politics, most people outsource their thinking to trusted sources. The problem then becomes how they determine who is trustworthy.

Individual differences can result in part from acceptance/rejection of AOT as a standard, and trying to conform to it (or not). Rejection of this standard consists of myside bias (confirmation bias, looking for support for an initially favored conclusion) and "uncertainty aversion," which is a belief that uncertainty itself is undesirable. These two properties work together. One way to avoid uncertainty is to try to bolster initial conclusions so that confidence will increase. Shynkaruk and Thompson (2006) found support for such bolstering. Subjects judged the validity of each of 12 syllogisms intuitively (within 10 s) and then deliberatively, rating their confidence in each judgment. Of interest, many subjects showed increased confidence after deliberation even though they did not change their (incorrect) answer. Other evidence, reviewed by De Neys, indicates that deliberation can serve to rationalize initial conclusions, an example of myside bias.

It would be nice to put all the pieces together through studies of individual differences in tasks like that used by Shynkaruk and Thompson.

Competing interest. None.

References

Ackerman, R., & Thompson, V. A. (2017). Meta-reasoning: Monitoring and control of thinking and reasoning. *Trends in Cognitive Sciences*, *21*(8), 607–617.
Baron, J. (1985). *Rationality and intelligence*. Cambridge University Press.
Baron, J. (2019). Actively open-minded thinking in politics. *Cognition*, *188*, 8–18 (Special issue on Cognition and politics, edited by Steven Sloman and Elke Weber).
Baron, J., Isler, O., & Yilmaz, O. (in press). Actively open-minded thinking and the political effects of its absence. In V. Ottati & C. Stern (Eds.), *Divided: Open-mindedness and dogmatism in a polarized world*. Oxford University Press.
Shynkaruk, J. M., & Thompson, V. A. (2006). Confidence and accuracy in deductive reasoning. *Memory and Cognition*, *34*(3), 619–632.

Learning how to reason and deciding when to decide

Senne Braem[a], Leslie Held[a], Amitai Shenhav[b] and Romy Frömer[b,c]

[a]Department of Experimental Psychology, Universiteit Gent, Gent, Belgium; [b]Department of Cognitive, Linguistic and Psychological Sciences, Brown University, Providence, RI, USA and [c]Centre for Human Brain Health, School of Psychology, University of Birmingham, Birmingham, UK
senne.braem@ugent.be; https://users.ugent.be/~sbraem/
leslie.held@ugent.be
amitai_shenhav@brown.edu; https://www.shenhavlab.org
r.froemer@bham.ac.uk

doi:10.1017/S0140525X22003090, e115

Abstract

Research on human reasoning has both popularized and struggled with the idea that intuitive and deliberate thoughts stem from two different systems, raising the question how people switch between them. Inspired by research on cognitive control and conflict monitoring, we argue that detecting the need for further thought relies on an intuitive, context-sensitive process that is learned in itself.

Research on reasoning about moral dilemmas or logical problems has traditionally dissociated fast, intuitive modes of responding from slow, deliberate response strategies, often referred to as system 1 versus system 2. For example, when deciding to take the plane versus train, our system 1 might make us decide to take the former because of its speed, whereas our system 2 could lead to deliberations on its environmental impact and decide for the train. De Neys proposes a new working model wherein both intuitive and deliberate reasoning are thought to originate from initial "system 1"-intuitions whose activations build up over time and potentially trigger an uncertainty signal. When this uncertainty signal reaches a certain threshold, it can trigger the need for deliberate reasoning, upon which deliberate thought or "system 2," is called upon to further resolve the reasoning problem. Here, we question the need for assuming a separate, deliberate system, that is activated only conditional upon uncertainty detection. Although we are sympathetic to the idea that uncertainty is being monitored and can trigger changes in the thought process, we believe these changes may result from adaptations in decision boundaries (i.e., deciding when to decide) or other control parameters, rather than invoking qualitatively different thought strategies.

Research on cognitive control often focuses on how goal-directed control processes can help us correct, inhibit, or switch away from interfering action tendencies, such as those originating from overtrained associations (Diamond, 2013; Miller & Cohen, 2001). For example, when deciding between the train or plane, our prior habit of taking the plane might trigger the same decision at first, while our current goal of being more environmentally friendly should lead us to the train. Importantly, recent theories on cognitive control have emphasized how these goal representations and control processes should not be considered as separate "higher" order processes studied in isolation, but that they are deeply embedded in the same associative network that hosts habits and overtrained responses. That is, goals and control functions can be learned, triggered, and regulated, by the same learning principles that govern other forms of behavior (Abrahamse, Braem, Notebaert, & Verguts, 2016; Braem & Egner, 2018; Doebel, 2020; Lieder, Shenhav, Musslick, & Griffiths, 2018; Logan, 1988). For example, much like the value of simple actions, the value of control functions can be learned (Braem, 2017; Bustamante, Lieder, Musslick, Shenhav, & Cohen, 2021; Grahek, Frömer, Prater Fahey, & Shenhav, 2023; Otto, Braem, Silvetti, & Vassena, 2022; Shenhav, Botvinick, & Cohen, 2013; Yang, Xing, Braem, & Pourtois, 2022). This way, similar to De Neys's suggestion that we can learn intuitions for the alleged systems 1 and 2 responses (or habitual versus goal-directed responses), we argue that people also learn intuitions for different control functions or parameters (see below).

One popular way to study the dynamic interaction between goal-directed and more automatic, habitual response strategies is through the use of evidence accumulation models. In these models, decisions are often thought to be the product of a noisy evidence accumulation process that triggers a certain response once a predetermined decision boundary is reached (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Ratcliff, Smith, Brown, & McKoon, 2016; Shadlen & Shohamy, 2016). However, this accumulation of evidence does not qualitatively distinguish between the activation of intuitions versus goal-directed or "controlled" deliberation. Instead, both processes start accumulating evidence at the same time, although potentially from different starting points (e.g., biased toward previous choices or goals) or at different rates (e.g., Ulrich, Schröter, Leuthold, & Birngruber, 2015). Depending on how high a decision maker sets their decision boundary, that is, how cautious versus impulsive they are, the goal-directed process will sometimes be too slow to shape, or merely slow down, the decision. These models have been successfully applied to social decision-making problems (e.g., Hutcherson, Bushong, & Rangel, 2015; Son, Bhandari, & FeldmanHall, 2019).

In line with the proposal by De Neys, we agree that competing evidence accumulation processes could trigger an uncertainty signal (e.g., directional deviations in drift rate), once uncertainty reaches a certain threshold, similar to how it has been formalized in the seminal conflict monitoring theory (Botvinick, Braver, Barch, Carter, & Cohen, 2001), itself inspired by Berlyne (1960). However, in our view, the resolution of said signal does not require the activation of an independent system but rather induces controlled changes in parameter settings. Thus, unlike activating a system 2 that provides answers by using a different strategy, cognitive control changes the parameters of the ongoing decision process (for a similar argument, see Shenhav, 2017). For example, it could evoke a simple increase in decision boundary, allowing for the evidence accumulation process to take more time before making a decision (e.g., Cavanagh et al., 2011; Frömer & Shenhav, 2022; Ratcliff & Frank, 2012). The second-order parameters that determine these adaptive control processes (e.g., how high one's uncertainty threshold should be before calling for adaptations, or how much one should increase their boundary) do not need to be made in the moment, but can be learned (e.g., Abrahamse et al., 2016).

Although we focused on the boundary as closely mapping onto fast and slow processing, we believe other process parameters can be altered too. For example, the response to uncertainty may require or could be aided by directed attention (Callaway, Rangel, & Griffiths, 2021; Jang, Sharma, & Drugowitsch, 2021; Smith & Krajbich, 2019), the memory of previous computations (Dasgupta & Gershman, 2021), learned higher-order strategies (Griffiths et al., 2019; Wang, 2021), or the parsing of a problem into different (evidence accumulation) subprocesses (Hunt et al., 2021). Moreover, a decision maker might even mentally simulate several similar decisions to evaluate one's (un)certainty before making a response (e.g., by covertly solving the same problem multiple times, Gershman, 2021). In sum, we argue that both intuitive and deliberate reasoning result from similar evidence accumulation processes whose parameter adjustments rely on conflict monitoring and learning from previous experiences.

Financial support. This work was supported by an ERC Starting grant awarded to S.B. (European Union's Horizon 2020 research and innovation program, grant agreement 852570), a fellowship by the FWO (Fonds Voor Wetenschappelijk Onderzoek–Research Foundation Flanders 11C2322N) awarded to L.H., and grant R01MH124849 from the National Institute of Mental Health awarded to A.S.

Competing interest. None.

References

Abrahamse, E., Braem, S., Notebaert, W., & Verguts, T. (2016). Grounding cognitive control in associative learning. *Psychological Bulletin*, *142*, 693–728.
Berlyne, D. (1960). *Conflict, arousal, and curiosity*. McGraw-Hill.
Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, *113*(4), 700.
Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624.
Braem, S. (2017). Conditioning task switching behavior. *Cognition*, *166*, 272–276.
Braem, S., & Egner, T. (2018). Getting a grip on cognitive flexibility. *Current Directions in Psychological Science*, *27*(6), 470–476.
Bustamante, L., Lieder, F., Musslick, S., Shenhav, A., & Cohen, J. (2021). Learning to overexert cognitive control in a Stroop task. *Cognitive, Affective, & Behavioral Neuroscience*, *21*(3), 453–471.
Callaway, F., Rangel, A., & Griffiths, T. L. (2021). Fixation patterns in simple choice reflect optimal information sampling. *PLoS Computational Biology*, *17*(3), e1008863.
Cavanagh, J. F., Wiecki, T. V., Cohen, M. X., Figueroa, C. M., Samanta, J., Sherman, S. J., & Frank, M. J. (2011). Subthalamic nucleus stimulation reverses mediofrontal influence over decision threshold. *Nature Neuroscience*, *14*(11), 1462–1467.
Dasgupta, I., & Gershman, S. J. (2021). Memory as a computational resource. *Trends in Cognitive Sciences*, *25*(3), 240–251.
Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135.
Doebel, S. (2020). Rethinking executive function and its development. *Perspectives on Psychological Science*, *15*(4), 942–956.
Frömer, R., & Shenhav, A. (2022). Filling the gaps: Cognitive control as a critical lens for understanding mechanisms of value-based decision-making. *Neuroscience & Biobehavioral Reviews*, *134*, 104483.
Gershman, S. (2021). *What makes us smart: The computational logic of human cognition*. Princeton University Press.
Grahek, I., Frömer, R., Prater Fahey, M., & Shenhav, A. (2023). Learning when effort matters: Neural dynamics underlying updating and adaptation to changes in performance efficacy. *Cerebral Cortex*, *33*(5), 2395–2411.
Griffiths, T. L., Callaway, F., Chang, M. B., Grant, E., Krueger, P. M., & Lieder, F. (2019). Doing more with less: Meta-reasoning and meta-learning in humans and machines. *Current Opinion in Behavioral Sciences*, *29*, 24–30.
Hunt, L. T., Daw, N. D., Kaanders, P., MacIver, M. A., Mugan, U., Procyk, E., … Kolling, N. (2021). Formalizing planning and information search in naturalistic decision-making. *Nature Neuroscience*, *24*(8), 1051–1064.

Hutcherson, C. A., Bushong, B., & Rangel, A. (2015). A neurocomputational model of altruistic choice and its implications. *Neuron*, *87*(2), 451–462.
Jang, A. I., Sharma, R., & Drugowitsch, J. (2021). Optimal policy for attention-modulated decisions explains human fixation behavior. *eLife*, *10*, e63436.
Lieder, F., Shenhav, A., Musslick, S., & Griffiths, T. L. (2018). Rational metareasoning and the plasticity of cognitive control. *PLoS Computational Biology*, *14*(4), e1006043.
Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*(4), 492.
Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*(1), 167–202.
Otto, A. R., Braem, S., Silvetti, M., & Vassena, E. (2022). Is the juice worth the squeeze? Learning the marginal value of mental effort over time. *Journal of Experimental Psychology: General*, *151*(10), 2324–2341.
Ratcliff, R., & Frank, M. J. (2012). Reinforcement-based decision making in corticostriatal circuits: Mutual constraints by neurocomputational and diffusion models. *Neural Computation*, *24*(5), 1186–1229.
Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, *20*(4), 260–281.
Shadlen, M. N., & Shohamy, D. (2016). Decision making and sequential sampling from memory. *Neuron*, *90*(5), 927–939.
Shenhav, A. (2017). The perils of losing control: Why self-control is not just another value-based decision. *Psychological Inquiry*, *28*(2–3), 148–152.
Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron*, *79*(2), 217–240.
Smith, S. M., & Krajbich, I. (2019). Gaze amplifies value in decision making. *Psychological Science*, *30*(1), 116–128.
Son, J. Y., Bhandari, A., & FeldmanHall, O. (2019). Crowdsourcing punishment: Individuals reference group preferences to inform their own punitive decisions. *Scientific Reports*, *9*(1), 1–15.
Ulrich, R., Schröter, H., Leuthold, H., & Birngruber, T. (2015). Automatic and controlled stimulus processing in conflict tasks: Superimposed diffusion processes and delta functions. *Cognitive Psychology*, *78*, 148–174.
Wang, J. X. (2021). Meta-learning in natural and artificial intelligence. *Current Opinion in Behavioral Sciences*, *38*, 90–95.
Yang, Q., Xing, J., Braem, S., & Pourtois, G. (2022). The selective use of punishments on congruent versus incongruent trials in the Stroop task. *Neurobiology of Learning and Memory*, *193*, 107654.

Deliberative control is more than just reactive: Insights from sequential sampling models

Hyuna Cho[a], Yi Yang Teoh[a], William A. Cunningham[a,b] and Cendri A. Hutcherson[b,c]

[a]Department of Psychology, University of Toronto, Toronto, ON, Canada; [b]Department of Marketing, Rotman School of Management, University of Toronto, Toronto, ON, Canada and [c]Department of Psychology, University of Toronto Scarborough, Scarborough, ON, Canada
hyuna.cho@mail.utoronto.ca
yang.teoh@mail.utoronto.ca
wil.cunningham@utoronto.ca
c.hutcherson@utoronto.ca; https://torontodecisionneurolab.com

doi:10.1017/S0140525X22003120, e116

Abstract

Activating relevant responses is a key function of automatic processes in De Neys's model; however, what determines the order or magnitude of such activation is ambiguous. Focusing on recently developed sequential sampling models of choice, we argue that proactive control shapes response generation but does not cleanly fit into De Neys's automatic-deliberative distinction, highlighting the need for further model development.

We applaud De Neys's work to define a set of domain-agnostic organizing principles that better clarify discussion on dual-process theories. This reformulation makes a welcome contribution to the field by proposing that (1) fast and intuitive response generation can activate multiple competing responses, leading to choice uncertainty, and (2) that this uncertainty drives subsequent activation of control-related deliberation. However, critical properties of these processes remain ambiguous in the current framework. Specifically, given the central role of fast and intuitive response generation processes, it is imperative to better specify how response options are generated, what determines their relative strength and time-course, and how these intuitions are compared to select a response. In this commentary, we draw on insights from the sequential sampling modeling literature to argue that even initial response generation and evaluation may not be exclusively driven by fast, automatic, and intuitive associative recall, but also modulated by controlled processes that operate rapidly from prior knowledge. In particular, we argue that deliberative control is deployed to prioritize information sampling and attribute evaluation, and thus response generation. We discuss how these forms of proactive control, in contrast to reactive control, pose a challenge to De Neys's current framework.

In De Neys's formulation, intuitive *responses* are the computational units that drive decisions. But these responses are themselves driven by the consideration of different cues or samples of information. Thus, the intuition generation process seems conceptually related to, if not synonymous with, the activation of relevant choice attributes in sequential sampling models. In these models, samples are drawn from noisy distributions of attribute values and accumulated as evidence for response options until the evidence passes a threshold for choice (Ratcliff & McKoon, 2008; Shadlen & Shohamy, 2016). The order in which attributes are considered can strongly influence decisions (Sullivan & Huettel, 2021; Sullivan, Hutcherson, Harris, & Rangel, 2015). The present dual-process framework appears to use a similar probabilistic sampling process, suggesting that insights from the growing literature on sequential sampling models could prove informative.

Recent work on sequential sampling models demonstrates that people strategically prioritize gathering more valuable information, which can change both the temporal dynamics and strength of response generation. For example, in altruistic choice under time pressure, selfish people prioritize gathering information about their own, rather than others', outcomes (Teoh, Yao, Cunningham, & Hutcherson, 2020). This systematically biases visual attention within the first few hundred milliseconds of choice presentation. In De Neys's terms, strategic allocation of attention changes the order of intuitive responses. Furthermore, this rapid reprioritization is context-sensitive: Changing the incentives of a social interaction (e.g., dictator vs. ultimatum game; Teoh & Hutcherson, 2022) or the framing of a risky gamble (e.g., gain vs. loss frame; Roberts, Teoh, & Hutcherson, 2022) change which information is processed first, in a goal-consistent manner. Thus, prior information shapes information search patterns *prior* to information sampling and response generation, appearing to operate *independently* of the uncertainty-triggered control in De Neys's model.

Similarly, prestimulus control-related signals can also change the order or strength of information recall, proactively shaping the response generation process. For example, time-varying sequential sampling models of food choice demonstrate that instructions to focus on health-related goals – a presumably deliberative process – results in faster activation of health-related information (Maier et al., 2020). In addition to changing the temporal dynamics of information retrieval, holding health-related goals increases how much weight people place on health relative to taste in their food choices (Hare, Malmaud, & Rangel, 2011; Tusche & Hutcherson, 2018). This suggests that retrieving and generating response options is not solely automatic. Instead, effortfully maintained goals can determine which information is most relevant, and can change the order in which response-relevant attributes are considered.

These results from both attention and memory sampling highlight an important distinction between *reactive control*, which are triggered by an event and strongly resembles the uncertainty-triggered deliberation of De Neys's model, and *proactive control*, which refers to regulatory processes that occur before encountering a stimulus (Braver, 2012; Braver, Gray, & Burgess, 2007). Importantly, as we have suggested above, our own and others' work suggests that this form of control can modulate when and what intuitions are activated even in the absence of conflict, and can alter the strength or order of information processing *before* rather than *after* intuitions are retrieved.

Better specifying how prestimulus control influences response generation may not only better link the current model to the self-regulation literature, but also extend it to more general models of information processing. The iterative reprocessing framework (Cunningham, Zelazo, Packer, & Van Bavel, 2007) is one such model which allows both stimulus-driven, bottom-up processes to inform goal-based, top-down processes, and vice versa. This echoes findings in attention (Asplund, Todd, Snyder, & Marois, 2010; Corbetta & Shulman, 2002) and memory (Burianová, Ciaramelli, Grady, & Moscovitch, 2012; Ciaramelli, Grady, & Moscovitch, 2008) which propose that there are distinct but related top-down and bottom-up processes which mutually inform each other. Under this framework, organizational, top-down processes are *always* informing what is considered most relevant by stimulus-driven processes. This top-down influence could become more effortful or directed with reflective control (Cunningham & Zelazo, 2007), but pre-existing knowledge plays an important causal role in determining the relevance of automatically retrieved information.

As uncertainty-triggered deliberative processes remain to be fully specified in De Neys's model, it is unclear whether proactive control processes should be considered a separate process, or whether it might use the same architecture. Regardless, considering when and how proactive deliberative processes are activated represents a fruitful area of inquiry. For example, dieters are often highly motivated to engage in healthy eating, yet may fail to spontaneously engage in proactive control (Cosme, Zeithamova, Stice, & Berkman, 2020). Although learning can automatize these priorities, as De Neys discusses, the effortful engagement of proactive control is not well incorporated into the current automatic-deliberative division. This case study thus highlights the need for a better articulation of how *both* intuitive and deliberative processes shape the initial response generation process, and points to the benefits of marrying dual-process models with the richness of recent computational models of information sampling and choice.

Acknowledgments. H.C. would like to thank the members of the Toronto Decision Neuroscience Lab and the 2020 enrollee cohort of graduate students at the University of Toronto, Department of Psychology for their continued support.

Financial support. Funding from the Canada Research Chairs program (to C.A.H.), the Natural Science and Engineering Research Council (to W.A.C.), and an Ontario Graduate Scholarship (to Y.Y.T.) is gratefully acknowledged. All views expressed in this article represent the views of the authors, and not of the funding bodies.

Competing interest. None.

References

Asplund, C. L., Todd, J. J., Snyder, A. P., & Marois, R. (2010). A central role for the lateral prefrontal cortex in goal-directed and stimulus-driven attention. *Nature Neuroscience*, *13*(4), 507–512. https://doi.org/10.1038/nn.2509

Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences*, *16*(2), 106–113. https://doi.org/10.1016/j.tics.2011.12.010

Braver, T. S., Gray, J. R., & Burgess, G. C. (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. *Variation in Working Memory*, *75*, 106. http://doi.org/10.1093/acprof:oso/9780195168648.003.0004

Burianová, H., Ciaramelli, E., Grady, C. L., & Moscovitch, M. (2012). Top-down and bottom-up attention-to-memory: Mapping functional connectivity in two distinct networks that underlie cued and uncued recognition memory. *NeuroImage*, *63*(3), 1343–1352. https://doi.org/10.1016/j.neuroimage.2012.07.057

Ciaramelli, E., Grady, C. L., & Moscovitch, M. (2008). Top-down and bottom-up attention to memory: A hypothesis (AtoM) on the role of the posterior parietal cortex in memory retrieval. *Neuropsychologia*, *46*(7), 1828–1851. https://doi.org/10.1016/j.neuropsychologia.2008.03.022

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201–215. https://doi.org/10.1038/nrn755

Cosme, D., Zeithamova, D., Stice, E., & Berkman, E. T. (2020). Multivariate neural signatures for health neuroscience: Assessing spontaneous regulation during food choice. *Social Cognitive and Affective Neuroscience*, *15*(10), 1120–1134. https://doi.org/10.1093/scan/nsaa002

Cunningham, W. A., & Zelazo, P. D. (2007). Attitudes and evaluations: A social cognitive neuroscience perspective. *Trends in Cognitive Sciences*, *11*(3), 97–104. https://doi.org/10.1016/j.tics.2006.12.005

Cunningham, W. A., Zelazo, P. D., Packer, D. J., & Van Bavel, J. J. (2007). The iterative reprocessing model: A multilevel framework for attitudes and evaluation. *Social Cognition*, *25*(5), 736–760. http://doi.org/10.1521/soco.2007.25.5.736

Hare, T. A., Malmaud, J., & Rangel, A. (2011). Focusing attention on the health aspects of foods changes value signals in vmPFC and improves dietary choice. *Journal of Neuroscience*, *31*(30), 11077–11087. https://doi.org/10.1523/JNEUROSCI.6383-10.2011

Maier, S. U., Raja Beharelle, A., Polanía, R., Ruff, C. C., & Hare, T. A. (2020). Dissociable mechanisms govern when and how strongly reward attributes affect decisions. *Nature Human Behaviour*, *4*(9), 949–963. https://doi.org/10.1038/s41562-020-0893-y

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922. https://doi.org/10.1162/neco.2008.12-06-420

Roberts, I. D., Teoh, Y. Y., & Hutcherson, C. A. (2022). Time to pay attention? Information search explains amplified framing effects under time pressure. *Psychological Science*, *33*(1), 90–104. https://doi.org/10.1177/09567976211026983

Shadlen, M. N., & Shohamy, D. (2016). Decision making and sequential sampling from memory. *Neuron*, *90*(5), 927–939. https://doi.org/10.1016/j.neuron.2016.04.036

Sullivan, N., Hutcherson, C., Harris, A., & Rangel, A. (2015). Dietary self-control is related to the speed with which attributes of healthfulness and tastiness are processed. *Psychological Science*, *26*(2), 122–134. https://doi.org/10.1177/0956797614559543

Sullivan, N. J., & Huettel, S. A. (2021). Healthful choices depend on the latency and rate of information accumulation. *Nature Human Behaviour*, *5*, 1698–1706. https://doi.org/10.1038/s41562-021-01154-0

Teoh, Y. Y., & Hutcherson, C. A. (2022). The games we play: Prosocial choices under time pressure reflect context-sensitive information priorities. *Psychological Science*, *33*(9), 1541–1556. https://doi.org/10.1177/09567976221094782

Teoh, Y. Y., Yao, Z., Cunningham, W. A., & Hutcherson, C. A. (2020). Attentional priorities drive effects of time pressure on altruistic choice. *Nature Communications*, *11*(1), 1–13. https://doi.org/10.1038/s41467-020-17326-x

Tusche, A., & Hutcherson, C. A. (2018). Cognitive regulation alters social and dietary choice by changing attribute representations in domain-general and domain-specific brain circuits. *eLife*, *7*, e31185. https://doi.org/10.7554/eLife.31185

What is intuiting and deliberating? A functional–cognitive perspective

Jan De Houwer, Yannick Boddez and Pieter Van Dessel

Department of Experimental Clinical and Health Psychology, Ghent University, Ghent, Belgium
jan.dehouwer@ugent.be
yannick.boddez@ugent.be
pieter.vandessel@ugent.be; www.liplab.be

doi:10.1017/S0140525X22003028, e117

Abstract

We applaud De Neys for drawing attention to the interaction between intuiting and deliberating without committing to single- or dual process models. It remains unclear, however, how he conceptualizes the distinction between intuiting and deliberating. We propose several levels at which the distinction can be made and discuss the merits of defining intuiting and deliberating as different types of behavior.

The idea of two distinct types of thinking has been highly influential within psychology and beyond. De Neys refers to these types of thinking as intuiting and deliberating and identifies core aspects of the interaction between both. In doing so, he provides a valuable contribution to the literature.

It remains unclear, however, how the distinction between intuiting and deliberating itself should be conceptualized. There are at least three levels of analysis at which the distinction can be made: (1) at the descriptive level in terms of subjective experience (i.e., the experience of intuiting and deliberating); (2) at the functional level in terms of the environmental conditions under which thinking occurs (e.g., whether it requires time or the absence of other tasks); and (3) at the mental level in terms of mental mechanisms and the mental representations on which they operate (e.g., associative or propositional representations).

Like others before him (e.g., Kahneman, 2011), De Neys draws the distinction in terms of speed and effort: Though intuiting is used to refer to fast and effortless thinking, deliberating refers to slow and effortful thinking. It is not entirely clear whether speed and effort are conceptualized at the descriptive level (i.e., a subjective experience) or at the functional level (e.g., actual time required; interference by other tasks) but De Neys does not seem to situate the distinction at the mental level. For instance, he argues that "both single- and dual-process theories focus on the interaction between intuition and deliberation" and that his "criticism and recommendations equally apply to single- and dual-process models" (target article, Introduction, para. 5). Assuming that the distinction between single- and dual-process models is situated at the mental level, these arguments suggest that the distinction between intuiting and deliberating needs to be made at another level than the mental one.

We definitely agree that there are many benefits of separating to-be-explained phenomena (such as intuiting and deliberating) from explanatory mental mechanisms (e.g., spreading of activation; propositional reasoning; see Hempel, 1970; Hughes, De Houwer, & Perugini, 2016). However, in his target article, De Neys does so in a manner that is not entirely coherent. Most importantly, he allows for the concept of low effort deliberation. If deliberation is by definition effortful (either descriptively or functionally), then how can it be effortless? In our opinion, the idea of low effort deliberation makes sense only if deliberation is situated at the mental level, for instance, when postulating a single-process theory in which all deliberating involves the manipulation of propositional representations (i.e., propositional reasoning). Hence, by allowing for the idea of low effort deliberation, De Neys seems to implicitly conceptualize deliberation at the mental level. We encourage him to be more explicit about how exactly he draws the distinction between intuiting and deliberating, most importantly, with regard to the level of analysis at which this distinction is situated.

In the remainder of this commentary, we discuss two ideas for clarifying the nature of intuiting and deliberating that, in our opinion, have not yet been given sufficient consideration in the literature. First, when delineating intuiting and deliberating, we see merit in taking seriously the descriptive level. In recent years, important progress has been made in studying a variety of subjective experiences such as the experience of confidence (e.g., Desender, Boldt, & Yeung, 2018), sense of agency (Marcel, 2003), conflict (e.g., Desender, Van Opstal, & Van den Bussche, 2014), making an effort (e.g., Naccache et al., 2005), and the urge to err (e.g. Questienne, van Dijck, & Gevers, 2018). We believe it would be interesting and feasible to study also the experience of intuiting and deliberating. This approach would draw attention away from the ontological and most likely unproductive debates about what is the "true" nature of intuiting and deliberating. It would also allow researchers to document the conditions under which people report intuiting and deliberating, as well as the possible differences in decisions produced under these conditions (i.e., to conduct functional research on intuiting and deliberating as descriptive phenomena). Finally, knowledge about these conditions and differences would help constrain theories about the mental mechanisms that produce the subjective experience of intuiting and deliberating.

Second, clarifying the nature of intuiting and deliberating not only requires specifying how they differ but also what they have in common. Both are typically thought of as instances of thinking but what is thinking? Here we see merit in conceptualizing thinking as a type of behavior (De Houwer, 2022; De Houwer, Barnes-Holmes, & Barnes-Holmes, 2018). Functional psychologists have successfully explored the benefits of this approach with regard to a variety of cognitive activities such as perceiving (e.g., Skinner, 1963), memorizing (e.g., Guinther & Dougher, 2014), and learning (De Houwer & Hughes, 2020). Conceiving of intuiting and deliberating as behavioral phenomena allows one to distinguish them at the descriptive level (i.e., as different subjective experiences; see De Houwer, 2022) or at the functional level (e.g., as relational responding in a slow or fast manner; see De Houwer et al., 2018; Hughes, Barnes-Holmes, & Vahey, 2012) without making a priori assumptions at the mental level (i.e., about the mental mechanisms that allow for thinking as behavior). From this behavioral perspective, the primary aim of research is to understand the environmental conditions that moderate these phenomena. For this research, inspiration can be found in the extensive literature on known moderators of behavior in general (e.g., Catania, 2013; Fisher, Piazza, & Roane, 2011). For instance, it is likely that switching between the behavior of intuiting and the behavior of deliberating is heavily dependent on antecedents (i.e., discriminative stimuli) and consequences (i.e., reinforcers and punishers). In line with the functional–

cognitive framework for research on psychology (De Houwer, 2011; Hughes et al., 2016), knowledge about the moderators of intuiting and deliberating not only has merit as such (i.e., it allows for prediction and control) but also facilitates the development of theories about the mental mechanisms that mediate these phenomena. In this way, combining descriptive and functional definitions with a behavioral perspective can provide a new impetus for both functional and cognitive research on intuiting and deliberating.

Financial support. The preparation of this paper was made possible by Ghent University grant BOF16/MET_V/002 to Jan De Houwer.

Competing interest. None.

References

Catania, A. C. (2013). *Learning* (5th ed.). Sloan.
De Houwer, J. (2011). Why the cognitive approach in psychology would profit from a functional approach and vice versa. *Perspectives on Psychological Science*, *6*, 202–209.
De Houwer, J. (2022). *On the merits and challenges of treating conscious and unconscious thoughts and feelings as behavior*. PsychArchives. https://doi.org/10.23668/psycharchives.5332
De Houwer, J., Barnes-Holmes, D., & Barnes-Holmes, Y. (2018). What is cognition? A functional–cognitive perspective. In S. C. Hayes & S. G. Hofmann (Eds.), *Process-based CBT* (pp. 119–135). New Harbinger.
De Houwer, J., & Hughes, S. (2020). *The psychology of learning: An introduction from a functional–cognitive perspective*. MIT Press.
Desender, K., Boldt, A., & Yeung, N. (2018). Subjective confidence predicts information seeking in decision making. *Psychological Science*, *29*,761–778.
Desender, K., Van Opstal, F., & Van den Bussche, E. (2014). Feeling the conflict. The crucial role of conflict experience in adaptation. *Psychological Science*, *25*(3), 675–683.
Fisher, W. W., Piazza, C. C., & Roane, H. S. (2011). *Handbook of applied behavior analysis*. Guilford Press.
Guinther, P. M., & Dougher, M. J. (2014). Partial contextual control of semantic false memories in the form of derived relational intrusions following training. *The Psychological Record*, *64*, 457–473.
Hempel, C. G. (1970). *Aspects of scientific explanation and other essays in the philosophy of science*. Free Press.
Hughes, S., Barnes-Holmes, D., & Vahey, N. (2012). Holding on to our functional roots when exploring new intellectual islands: A voyage through implicit cognition research. *Journal of Contextual Behavioral Science*, *1*, 17–38.
Hughes, S., De Houwer, J., & Perugini, M. (2016). The functional–cognitive framework for psychological research: Controversies and resolutions. *International Journal of Psychology*, *51*, 4–14.
Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
Marcel, A. J. (2003). The sense of agency: Awareness and ownership of action. In J. Roessler & N. Eilan (Eds.), *Agency and self-awareness* (pp. 48–93). Oxford University Press.
Naccache, L., Dehaene, S., Cohen, L., Habert, M.-O., Guichart-Gomez, E., Galanaud, D., & Willer, J.-C. (2005). Effortless control: Executive attention and conscious feeling of mental effort are dissociable. *Neuropsychologia*, *43*, 1318–1328.
Questienne, L., van Dijck, J. P., & Gevers, W. (2018). Introspection of subjective feelings is sensitive and specific. *Journal of Experimental Psychology. Human Perception and Performance*, *44*, 215–225.
Skinner, B. F. (1963). Behaviorism at fifty. *Science*, *140*, 951–958.

Categorizing judgments as likely to be selected by intuition or deliberation

Aliya R. Dewey

Department of Philosophy, University of Arizona, Tucson, AZ, USA
aliyardewey@gmail.com
https://aliyadewey.com

doi:10.1017/S0140525X22002886, e118

Abstract

De Neys argues against the exclusivity assumption: That many judgments are exclusively selected by intuition or deliberation. But this is an excessively strong formulation of the exclusivity assumption. We should aim to develop weaker, more plausible formulations that identify which judgments are likely to be selected by intuition or deliberation. This is necessary for empirical comparisons of intuition and deliberation.

De Neys observes that dual-process theorists often assume that certain responses under certain conditions are only possible for either intuition or deliberation. For example, he points out that it's often assumed that the incorrect response to the bat-and-ball task is the result of intuition, and the correct response is the result of deliberation. If true, this would be convenient: We could compare intuition versus deliberation just by comparing processing that results in incorrect versus correct responses to the bat-and-ball task, respectively. But De Neys offers two arguments for why this exclusivity assumption is false: one theoretical and one empirical.

For his theoretical argument, De Neys argues that the exclusivity assumption contradicts the only plausible explanation for switching between intuition and deliberation. He explains that switching occurs when intuition detects conflict between responses and causes deliberation to intervene and resolve the conflict by selecting one of the responses. However, this contradicts the exclusivity assumption: If some responses are generated by intuition and other responses are generated by deliberation, intuition won't be able to detect conflict between intuitive and deliberative responses. So, he concludes, both responses must be generated by intuition and re-generated by deliberation.

But we must be careful to distinguish between response generation and response selection. The switching model only contradicts an exclusivity assumption about response generation – as we just noted. However, his switching model is consistent with an exclusivity assumption about response selection: Even if intuition *generates* both responses and deliberation *re-generates* them, it's still possible that intuition exclusively *selects* one response and deliberation exclusively *selects* another response. So, an exclusivity assumption about response selection is theoretically coherent, but is it empirically plausible?

For his empirical argument, De Neys argues that the exclusivity assumption contradicts a growing body of evidence. He points to two-response paradigms as an example: Subjects must give a first response very quickly and then are given plenty of time to reconsider and give a second response. The paradigm is designed to prevent deliberation in the first stage, isolating an intuitive response, and then to allow deliberation in the second stage, permitting a deliberative response. If intuitive responses to bias tasks are incorrect, as many assume, then the first responses should almost always be incorrect. But the evidence shows that many subjects who give the correct response on the second time gave the correct response on the first time too. This indicates that correct responses can be both intuitive and deliberative – contra the exclusivity assumption. I agree with De Neys that this evidence suggests that we're often wrong about which responses are selected by intuition versus deliberation.

However, I think that it's critical to emphasize that we can't compare intuition and deliberation (see sect. 4.2) unless we find better ways of categorizing responses as intuitive or deliberative.

After all, we must classify responses as intuitive or deliberative in order to compare intuition and deliberation. For example, consider how Greene, Sommerville, Nystrom, Darley, and Cohen (2001) looked for the neural correlates of moral intuition versus deliberation. They had to start by categorizing moral judgments as intuitive and deliberative. This way, they could look for the neural correlates of intuitive and deliberative judgments. Then they could infer that the neural correlates of intuitive judgments were neural correlates for moral intuition itself and likewise for moral deliberation itself.

To be clear, I don't believe that Greene et al. (2001) correctly identified which moral judgments were intuitive and deliberative (see Kahane, 2012; Kahane et al., 2012). My point is only that they had to categorize moral judgments as intuitive and deliberative to identify the neural correlates of moral intuition and deliberation. Unless we're prepared to offer a better notion of exclusivity, though, it's unclear how else we could compare the neural basis for intuition or deliberation – or make any other comparison between them. So, I recommend that we should aim to develop weaker formulations of the exclusivity assumption: We should (a) only categorize responses as *more likely* to be selected by intuition or deliberation, (b) find more reliable ways of classifying responses as probably intuitive and probably deliberative, and (c) be careful to validate whether responses really are more likely to be selected by intuition or deliberation.

I believe that De Neys has made a valuable contribution here by calling attention to the exclusivity assumption and rejecting its strongest formulation. But the correct response, I think, is to calibrate our exclusivity assumptions more carefully. I've tried to do this in recent work, where I develop a weaker formulation of the exclusivity assumption that draws on switching models, like the one that De Neys offers here (Dewey, 2022). It claims that (a) conditions that impair metacognitive heuristics (e.g., that decrease the salience of the correct response) result in responses that are *most* likely to be intuitive and (b) conditions that improve metacognitive heuristics (e.g., that increase the salience of the correct response) result in responses that are *most* likely to be deliberative. Of course, I don't mean to be defending my account here: I'm just pointing to it as an example for how to formulate weaker exclusivity assumptions that avoid the issues that De Neys raises here.

Finally, this paper highlights a shift in the psychology of thinking and reasoning. Traditionally, single- and dual-process theorists mostly cared about how to compare intuition and deliberation. But these questions have fallen out of vogue after years of intractable debates between single- and dual-process theorists. Recently, the focus has shifted from comparing intuition and deliberation to the metacognitive mechanisms that switch between intuition and deliberation. But old questions about how to compare intuition and deliberation deserve answers too! De Neys does call for answers to these questions in section 4.2, but I'd urge a more specific call: To get started, we need better formulations of the exclusivity assumption. So, I encourage the reader to read this paper as a welcome challenge from De Neys to sharpen our exclusivity assumptions so that we can get clearer on how to reliably compare intuition and deliberation.

Acknowledgment. I thank Wim De Neys for his feedback on this commentary and for our past exchanges, which have inspired this commentary.

Financial support. This research was not supported by any funding.

Competing interest. None.

References

Dewey, A. R. (2022). Metacognitive control in single- vs. dual-process theory. *Thinking & Reasoning*, 1–36. https://doi.org/10.1080/13546783.2022.2047106

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science (New York, N.Y.)*, *293*(5537), 2105–2108. https://doi.org/10.1126/science.1062872

Kahane, G. (2012). On the wrong track: Process and content in moral psychology. *Mind & Language*, *27*(5), 519–545. https://doi.org/10.1111/mila.12001

Kahane, G., Wiech, K., Shackel, N., Farias, M., Savulescu, J., & Tracey, I. (2012). The neural basis of intuitive and counterintuitive moral judgment. *Social Cognitive and Affective Neuroscience*, *7*(4), 393–402. https://doi.org/10.1093/scan/nsr005

Advancing theorizing about fast-and-slow thinking: The interplay between fast and slow processing

Rosa Angela Fabio[a] and Tindara Caprì[b]

[a]Department of Economics, University of Messina, Messina, Italy and
[b]Department of Life and Health Sciences, Link Campus University, Rome, Italy
rafabio@unime.it
tindara.capri@unime.it

doi:10.1017/S0140525X22003156, e119

Abstract

We agree with the author's working model, but we suggest that (a) the classical distinction between fast and slow processes as separable processes can be softened, and (b) human performance might result from an interplay between fast and slow processing and these processes may be mediated by systems that evolve to satisfy the need for operation in a complex environment.

In the target article, De Neys argues that dual-process models are obsolete and empirically and conceptually problematic to explain human reasoning. The author claims that problem is the tendency to conceive fast-and-slow processes or intuition and deliberation as two separate processes producing unique responses. So, the core problem of dual-process models is to assume that fast-and-slow processes are based on exclusivity features. In contrast to these models, De Neys suggests the exclusivity and switch are two interconnected features of fast-and-slow processes. We agree with the author's idea, the classical distinction between fast and slow processes as separable and exclusive processes can be softened.

In literature, it has been demonstrated that the slow processes can gradually become flexible and context dependent (Fabio, Caprì, & Romano, 2019). The term context refers to those perceptual features of the task setting that are not formally required for successful task performance, yet which may influence performance with practice based on contingencies with task-relevant information. Several studies (D'Angelo, Jiménez, Milliken, & Lupiáñez, 2012, 2013, 2014; Fabio et al., 2019; Ruitenberg, Abrahamse, De Kleine, & Verwey, 2012a, 2012b, 2015; Ruitenberg, Abrahamse, & Verwey, 2013) have examined if the slow processes can become flexible through reliance on contextual features, indicating that, when a task requires the activation of both fast-and-slow processes, subjects can switch between both processes and context features facilitate this switching. Therefore, it is reasonable to assume that the fast-and-slow processes do not necessarily generate unique responses and they have not the exclusivity features, because the slow processes can become more flexible through the inclusion of context-specific features and subjects can operate a switch between these two processes. So, the idea that the exclusivity and switch are two interconnected features of fast-and-slow processes is in line with these researches.

As suggested by De Neys, and we agree with his idea, traditional dual-process models fail to explain a viable internal switch mechanism. De Neys proposes a more viable general model that can serve as theoretical groundwork to build future dual-process models. The author's working model focuses on four components: intuitive activation, uncertainty monitoring, deliberation, and feedback. In this model, the starting point are two intuitions that can generate a different response. This component represents an alternative and new point of view if it is compared to a dual-process model in which the starting point is one intuitive or deliberative process. Moreover, according to the working model intuitive responses occur through an automatization or learning process. During development, any response might initially activate exclusive deliberation but through experience and practice this response will become automatized. This point of view is in line with the theories of automatization, demonstrating that after much practice the subjects show significant improvement in performing a task that initially require deliberative processes (Caprì, Santoddì, & Fabio, 2020; Fabio, 2009, 2017; Fabio & Caprì, 2019; Shiffrin & Schneider, 1977).

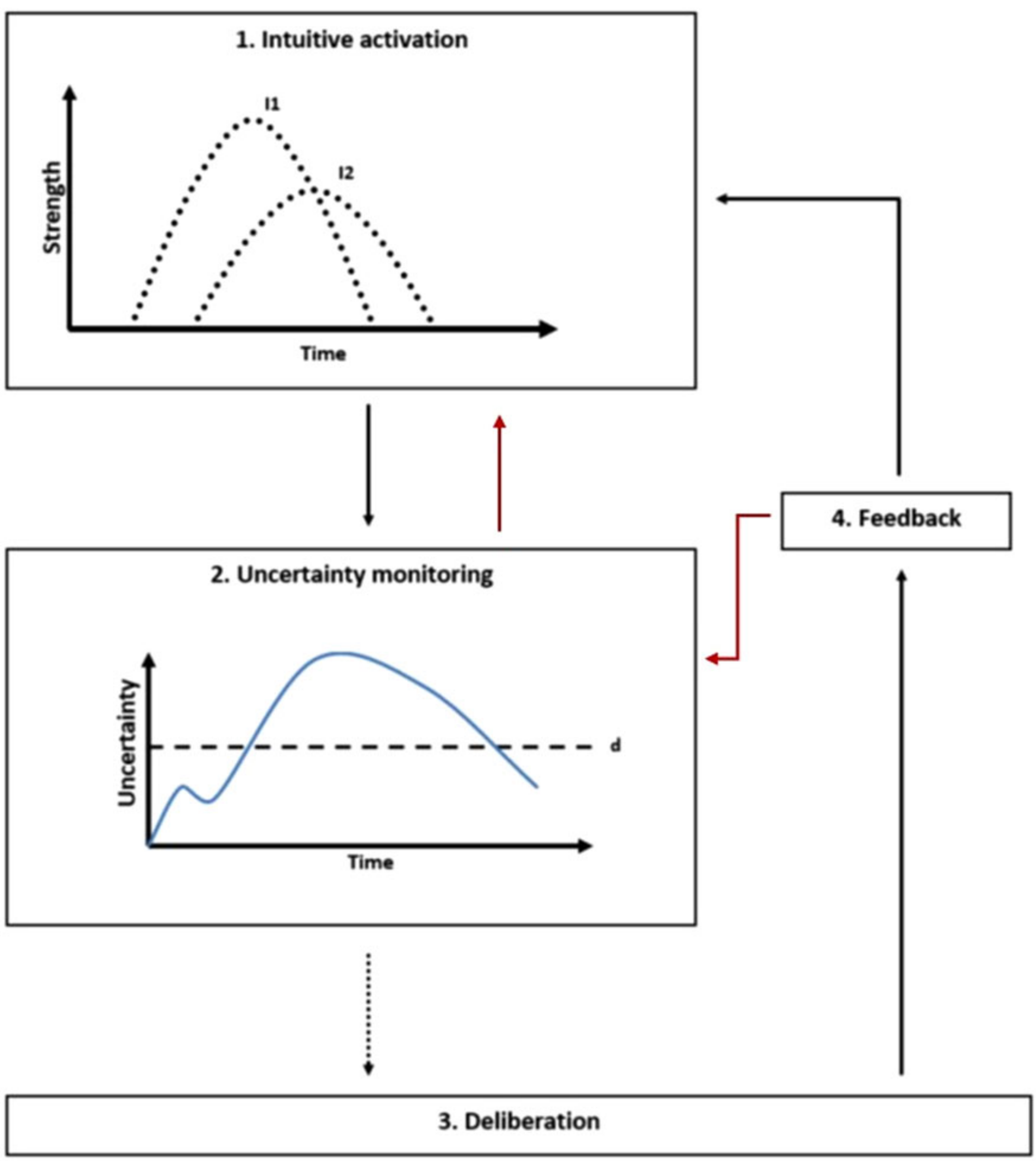

Figure 1 (Fabio and Caprì). A schematic illustration of the working model.

The second component of the De Neys's model is an uncertainty monitoring process, conceived as a mediator for access to the deliberative processes. According to the author, the uncertainty monitoring process calculates the strength difference between different activated intuitions, the more similar the activation strength and the higher the uncertainty will be. The argument on the role and functions of uncertainty monitoring process is very interesting. However, differently to De Neys's idea, we believe that it is not a problem if the uncertainty threshold is the same when there are two strong intuitions or two weak intuitions, this is correct because the uncertainty does not depend on the intensity of the two strengths. So, if both intuitions are weak or strong it is right to achieve that the uncertainty is strong.

The third component of the De Neys's model is the deliberation that comes from the strength of the different activated intuitions. Deliberation might generate a combination of response suppression, generation, justification, or additional processes, and not necessarily a decrease in uncertainty. Consequently, it could occur a feedback loop in which system 1 and system 2 interact. The feedback stage is the last component of the De Neys's model. We agree with the argument about the third and fourth components, but we propose a change of the schematic illustration of the working model's core components. In the De Neys's illustration the arrow of fourth component goes intuitive activation (first component), we think a circular architecture in which the arrow of the feedback (fourth component) should return to uncertainty stage, and if the uncertainty is decreased, the arrow of feedback goes toward intuition; whereas if it is increased, the arrow goes toward deliberation (Fig. 1).

Our suggestion on the change of direction of the arrow related to feedback stage is not a mere schematic suggestion, but it reflects a theoretical conceptualization of fast-and-slow processes as interconnected processes in which it is possible to switch between these and the uncertainty operate as mediator.

In conclusion, human performance might result from an interplay between fast and slow processing and these processes may be mediated by systems that evolve to satisfy the need for a decrease of uncertainty and operate in a complex environment.

Financial support. The authors declare no funding for the present commentary.

Competing interest. None.

References

Caprì, T., Santoddì, E., & Fabio, R. A. (2020). Multi-source interference task paradigm to enhance automatic and controlled processes in ADHD. *Research in Developmental Disabilities*, *97*, 1035–1042. https://doi.org/10.1016/j.ridd.2019.103542

D'Angelo, M. C., Jiménez, L., Milliken, B., & Lupiáñez, J. (2012). On the specificity of sequential congruency effects in implicit learning of motor and perceptual sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 69–84. https://doi.org/10.1037/a0028474

D'Angelo, M. C., Milliken, B., Jiménez, L., & Lupiáñez, J. (2013). Implementing flexibility in automaticity: Evidence from context-specific implicit sequence learning. *Consciousness and Cognition*, *22*(1), 64–81. https://doi.org/10.1016/j.concog.2012.11.002

D'Angelo, M. C., Milliken, B., Jiménez, L., & Lupiáñez, J. (2014). Re-examining the role of context in implicit sequence learning. *Consciousness and Cognition*, *27*, 172–193. https://doi.org/10.1016/j.concog.2014.05.005

Fabio, R. A. (2009). *Relationship between automatic and controlled processes of attention and Leading to complex thinking*. Nova.

Fabio, R. A. (2017). The study of automatic and controlled processes in ADHD: A reread and a new proposal. *Mediterranean Journal of Clinical Psychology*, *5*, 1–33.

Fabio, R. A., & Caprì, T. (2019). Automatic and controlled attentional capture by threatening stimuli. *Heliyon*, *5*, 17–52. https://doi.org/10.1016/j.heliyon.2019.e01752

Fabio, R. A., Caprì, T., & Romano, M. (2019). From controlled to automatic processes and back again: The role of contextual features. *Europe's Journal of Psychology*, *15* (4), 773–788. https://doi.org/10.5964/ejop.v15i4.1746

Ruitenberg, M. F. L., Abrahamse, E. L., De Kleine, E., & Verwey, W. B. (2012a). Context-dependent motor skill: Perceptual processing in memory-based sequence production. *Experimental Brain Research*, *222*, 31–40. https://doi.org/10.1007/s00221-012-3193-6

Ruitenberg, M. F. L., Abrahamse, E. L., & Verwey, W. B. (2013). Sequential motor skill in preadolescent children: The development of automaticity. *Journal of Experimental Child Psychology*, *115*, 607–623. https://doi.org/10.1016/j.jecp.2013.04.005

Ruitenberg, M. F. L., De Kleine, E., Van der Lubbe, R. H. J., Verwey, W. B., & Abrahamse, E. L. (2012b). Context-dependent motor skill and the role of practice. *Psychological Research Psychologische Forschung*, *76*, 812–820. https://doi.org/10.1007/s00426-011-0388-6

Ruitenberg, M. F. L., Verwey, W. B., & Abrahamse, E. L. (2015). What determines the impact of context on sequential action? *Human Movement Science*, *40*, 298–314. https://doi.org/10.1016/j.humov.2015.01.006

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, *84*, 1–66. https://doi.org/10.1037/0033-295X.84.2.127

Conflict paradigms cannot reveal competence

Roman Feiman

Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, RI, USA
roman_feiman@brown.edu
https://sites.brown.edu/bltlab/

doi:10.1017/S0140525X22002941, e120

Abstract

De Neys is right to criticize the *exclusivity assumption* in dual-process theories, but he misses the original sin underlying this assumption, which his working model continues to share. *Conflict paradigms*, in which experimenters measure how one cognitive process interferes (or does not interfere) with another, license few inferences about how the interfered-with process works on its own.

Imagine you want to study how people walk. If you're a dual-process theorist, you would use a *conflict paradigm*. You start by shackling a big weight to your participants' ankles and testing if they can walk a lap around a track. If they fall down or give up, you conclude they must not be very good at walking. If they make it, you conclude that with enough effort and motivation, people's walking capacity can overcome their tendency to fall. Either way, you conclude that walking is effortful, requiring focused attention, motivation, and some combination of talent and training.

De Neys is rightly unhappy with this picture. To advance dual-process theory, he proposes to figure out exactly how ankle weights affect walking. He reviews a wealth of evidence that complicates the picture: Some people (maybe Hafþór Björnsson) can still walk well even with the weight. If the weight is smaller, people walk better. If, instead of a weight, you attach a rope that pulls people toward the finish line (a so-called *non-conflict paradigm*), they actually get there faster. This shows that walking and added weight do not have to produce different outcomes (De Neys: *the alleged system 2 response does not seem to be out of reach of the intuitive system 1*).

Wouldn't it be better to remove the ankle weight altogether? De Neys concedes that people do seem to be pretty good at walking without it (*nobody will disagree that educated adults can intuitively solve a problem such as "Is 9 more than 1?"*), but he thinks that dual-process theories are not responsible for explaining this fact (*as any scientific theory, dual-process models make their assertions within a specific application context. For the dual-process model of logical reasoning, the application context concerns situations in which an intuitively cued problem solution conflicts with a logico-mathematical norm*). As a description of dual-process theories, this may be true. Still, it's fair to ask whether it should be.

Whether conflict paradigms are informative depends on what dual-process theories are meant to be theories of. If they aim to explain interference itself – how and under what circumstances it appears, disappears, hinders, or helps – then conflict paradigms are an excellent tool for eliciting the explanandum. But if dual-process theories are theories of reasoning, then studying interference can tell us roughly as much about reasoning as shackling strongmen on a gym track can tell us about walking. If, as is typically the case, the interference is designed to impede reasoning, then conflict paradigms will create a performance limitation that necessarily underestimates reasoning competence. Nevertheless, despite their limited "application context," dual-process theories make many claims about reasoning, tout court. For instance, De Neys describes how reasoning develops: *The working model postulates that intuitive responses primarily emerge through an automatization or learning process*. But his working model is based on evidence from different flavors of conflict and no-conflict paradigms, so the developmental claim is a non-sequitur. Evidence about how some other process does or does not interfere with reasoning cannot warrant any conclusion about how the interfered-with reasoning develops.

This is, in fact, a hard-won lesson from the history of developmental psychology. Jean Piaget (1950) famously studied children's ability to reason about number, volume, and other abstract concepts, and he frequently used conflict paradigms. For example, to investigate how children thought about number, Piaget showed them two identical rows of coins across from each other. When he asked children if the rows had the same number, they correctly said "yes." But Piaget worried that children were relying on a proxy to number, the equal lengths of the rows. To test this, he created a conflicting cue. He spread one row out so it looked longer and asked the same question again. Children as old as 6 years of age consistently switched to saying "no," the rows did not have the same number. Piaget concluded that 6-year-olds could not reason about number per se without conflating it with other properties, like length or area. Just like later dual-process theorists, Piaget presented his participants with a conflicting cue designed to tempt the wrong answer, showed that participants fell for it, and concluded that there was something wrong with their reasoning ability generally.

In the seven decades since, a vast body of work has shown that much younger children know much more about number than Piaget believed (see Carey, 2009; Carey & Barner, 2019). Summarizing this literature would take a book, but for present purposes it holds two critical lessons for dual-process theories. First, evidence that younger children have rich numerical understanding did not come from more or better variants of conflict paradigms. It came from new tasks that were designed to eliminate both the confounds that Piaget worried about and the conflicting cues he added, to make reasoning as easy as possible given the requisite competence. Second, this new understanding emerged without anyone figuring out exactly why children fail on Piaget's conflict paradigm. It turns out there are many different ways to make that task easier (e.g., McGarrigle & Donaldson, 1974; Mehler & Bever, 1967; Rose & Blank, 1974; Samuel & Bryant, 1984), but still no comprehensive theory of exactly what makes it hard. Understanding the interference proved unnecessary for understanding the interfered-with competence.

The science of children's thinking progressed not by drilling down on conflict paradigms, but by leaving them behind. It is well past time to let the science of adults' thinking do the same. The deep, difficult question about reasoning is, and has always been, the one De Neys and other dual-process theorists locate outside of the theory's scope. Discussing the conjunction fallacy, Kahneman (2011) notes in passing that it doesn't always arise. Everyone agrees that "Jane is a teacher" is more likely than "Jane is a teacher and walks to work." Kahneman even explains why: *In the absence of a competing intuition, logic prevails*. Right! Now, how does *that* work?

Acknowledgments. The author is grateful to David Barner, Gabor Brody, Rhea Feiman, and Tyler Knowlton for helpful discussion.

Financial support. This work was supported in part by a Research Fellowship from the Jacobs Foundation, and by NSF DRL Grant no. 2000661.

Competing interest. None.

References

Carey, S. (2009). *The origin of concepts*. Oxford University Press.
Carey, S., & Barner, D. (2019). Ontogenetic origins of human integer representations. *Trends in Cognitive Sciences*, *23*(10), 823–835.
Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.
McGarrigle, J., & Donaldson, M. (1974). Conservation accidents. *Cognition*, *3*(4), 341–350.
Mehler, J., & Bever, T. G. (1967). Cognitive capacity of very young children. *Science (New York, N.Y.)*, *158*(3797), 141–142.
Piaget, J. (1950). *The psychology of intelligence*. Routledge.
Rose, S. A., & Blank, M. (1974). The potency of context in children's cognition: An illustration through conservation. *Child Development*, *45*(2), 499–502.
Samuel, J., & Bryant, P. (1984). Asking only one question in the conservation experiment. *Journal of Child Psychology and Psychiatry*, *25*(2), 315–318.

Fast and slow language processing: A window into dual-process models of cognition

Fernanda Ferreira[a] and Falk Huettig[b]

[a]Department of Psychology, University of California, Davis, Davis, CA, USA and [b]Department of Language and Communication, Max Planck Institute for Psycholinguistics, Radboud University Nijmegen, Nijmegen, The Netherlands
fferreira@ucdavis.edu; https://ferreiralab.faculty.ucdavis.edu/
falk.huettig@mpi.nl; https://www.mpi.nl/people/huettig-falk

doi:10.1017/S0140525X22003041, e121

Abstract

Our understanding of dual-process models of cognition may benefit from a consideration of language processing, as language comprehension involves fast and slow processes analogous to those used for reasoning. More specifically, De Neys's criticisms of the exclusivity assumption and the fast-to-slow switch mechanism are consistent with findings from the literature on the construction and revision of linguistic interpretations.

Sometimes language processing can be hard. Just as many problems are easy to solve, many sentences are easy to interpret – for example, *the cat chased the dog*. Alternatively, just as some problems leave us stumped, some sentences defy our comprehension mechanisms – for example, the infamous *the horse raced past the barn fell*. For decades, psycholinguists have attempted to explain what makes sentences difficult to understand, with some models pointing to the costs of integrating information over long distances (Gibson, 1998), others focusing on the effects of the unexpectedness of each word as it is encountered (so-called surprisal-based models; Hale, 2016), and others emphasizing the consequences of ambiguity (Ferreira & Henderson, 1991). Here we concentrate on syntactic ambiguity because it highlights many of the issues associated with fast and slow processing. Specifically, in his target article, De Neys challenges researchers in decision making to reevaluate the exclusivity assumption and to specify how the switch mechanism that triggers the switch from fast to slow reasoning works. We believe consideration of these issues from the perspective of language processing could prove useful, as they have been at the center of theoretical debates in psycholinguistics.

During comprehension, the system that assigns syntactic structure, the parser, will often encounter a sequence that can be assigned more than one grammatical analysis. In those cases, given a range of linguistic biases, the parser may select an analysis that will require revision. Take the sequence *Mary believes Tom*. On the parser's first encounter with the postverbal noun phrase *Tom*, it will likely analyze the phrase as a direct object. But if the sentence continues with a verb such as *lied*, the parser has a problem: *lied* must be syntactically integrated but there is no grammatical place for it in the structure. The only solution is for the initial analysis to be revised so that *Tom* is not a direct object but rather the subject of a complement clause. Moreover, not only does the structure require revision, but also the meaning must be recomputed as well, because Mary does not in fact believe Tom. These processes can be viewed within the dual-processing framework De Neys discusses, with the initial analysis being the output of system 1 and the revised interpretation the output of system 2. The first response is fast and automatic, and the second requires a slower, more deliberate mode of processing in which the structure and the interpretation are systematically undone and rebuilt.

Much debate has centered around the question of what determines the initial analysis. For the purposes of this commentary, we set that question aside to focus on the two issues De Neys considers in the target article: the exclusivity assumption and the switch mechanism. Taking exclusivity first, psycholinguists know that often an initial, intuitive analysis will align with what a more deliberate process would deliver. Sentences sometimes resolve themselves in a way that is consistent with initial syntactic expectations (e.g., *Mary believes Tom implicitly*), and with knowledge and experience, many experienced language users will succeed in obtaining the correct interpretation of even the more challenging sentences right from the start, with no need for revision. In other cases, the initial system will deliver multiple interpretations of an ambiguous sequence, which means revision may involve a simple shift from one analysis to another. Findings from language comprehension, then, make clear that system 1 can deliver a correct analysis.

Turning now to the switch mechanism, much is known in psycholinguistics about what triggers the switch to a more deliberate, system 2 processing mode. One critical factor is a breakdown in coherence. In the case of so-called garden-path sentences such as *Mary believes Tom lied*, the trigger is syntactic collapse: The tree formed for the first three words cannot accommodate the verb *lied*. This breakdown in syntactic coherence shifts the parser into a repair mode in which it revisits its previous syntactic decisions, attempts new solutions, and tries to create a revised, integrated structure. In other cases, the trigger is a breakdown in semantic coherence. For example, given *Mary believes the rain…* (as in *Mary believes the rain will stop soon*), an initial analysis on which *the rain* is analyzed as a direct object can be revised when the more deliberative system detects the semantic anomaly of believing rain. This semantic incoherence will cause the parser to review its past syntactic decisions and attempt new choices that lead to a better semantic outcome. In reasoning, a switch from fast to slow processing may similarly be triggered by a breakdown in coherence, albeit at a conceptual rather than a linguistic level of representation.

Recent work on the influence of literacy can also be interpreted according to this dual-processing framework and is particularly relevant for thinking about exclusivity and the switch from system 1 to 2 modes that De Neys discusses. Literacy, for instance, uniquely predicts participants' ability to correctly accept and reject spoken sentences according to the prescriptive grammatical norms of their language (Favier & Huettig, 2021). In linguistics, such judgments are known to involve both systems 1 and 2 processes. Literacy also makes comprehension of challenging linguistic forms more automatic (as evidenced by enhanced prediction abilities; Favier, Meyer, & Huettig, 2021), providing one potential mechanism for how system 2 can, over time, turn into system 1 processing. A dual-systems approach to language processing thus has the potential to provide new mechanistic answers about the automatization of system 2 responses as well as the interplay between fast and slow systems.

In summary, our view is that a domain in which the exclusivity assumption and the switch mechanism highlighted by De Neys can be profitably scrutinized is language processing, a cognitive system that has not often been invoked in discussions of systems 1 and 2 processing and the coordination of their outputs. We believe that considering language processing through the lens of this dual-processing framework will help to illuminate the issues related to thinking that De Neys discusses in the target article.

Financial support. Preparation of this work was supported by the National Institutes of Health/NICHD under grant number 1R01HD100516 awarded to Fernanda Ferreira.

Competing interest. None.

References

Favier, S., & Huettig, F. (2021). Long-term written language experience affects grammaticality judgements and usage but not priming of spoken sentences. *Quarterly Journal of Experimental Psychology*, *74*(8), 1378–1395.

Favier, S., Meyer, A. S., & Huettig, F. (2021). Literacy can enhance syntactic prediction in spoken language processing. *Journal of Experimental Psychology: General*, *150*(10), 2167–2174.

Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, *30*(6), 725–745.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68* (1), 1–76.

Hale, J. (2016). Information-theoretical complexity metrics. *Language and Linguistics Compass*, *10*(9), 397–412.

Toward dual-process theory 3.0

Keith Frankish

Department of Philosophy, University of Sheffield, Sheffield, UK
k.frankish@gmail.com
www.keithfrankish.com

doi:10.1017/S0140525X22003144, e122

Abstract

This commentary is sympathetic to De Neys's revision of dual-process theory but argues for a modification to his position on exclusivity and proposes a bold further revision, envisaging a dual-process theory 3.0, in which system 1 not only initiates system 2 thinking but generates and sustains it as well.

Despite the huge amount of experimental work done under the dual-process banner, dual-process theory still lacks an agreed theoretical framework. One issue concerns implementation. Are the neural resources involved in supporting each type of processing discrete, or do they overlap? Another issue is coordination. How are the two types of processing related functionally, and how are their activities coordinated? Most dual-process theorists adopt a default-interventionist view, according to which system 1 generates default responses, and system 2 is activated only occasionally, generating a more considered response, which overrides the default one. (I use "system 1" as a label for the suite of intuitive processes, without assuming that they form a unified neural system; similarly for "system 2.") However, this prompts the further question of exactly how the switch between intuitive system 1 processing and deliberate system 2 processing is managed.

It is this question – the "switch issue" – that occupies De Neys. He proposes that switching is controlled by system 1: System 1 monitors its own responses, calculates a measure of their uncertainty, and initiates system 2 when the measure exceeds a certain threshold. System 1 also monitors the outputs of system 2 and terminates system 2 processing when a response with a suitably low uncertainty is generated. De Neys argues that this requires us to give up the assumption that system 2 responses are beyond the reach of system 1 (the "exclusivity assumption").

I like De Neys's revision of dual-process theory (a form of "dual-process theory 2.0"), but I am going to suggest that system 1 has an even bigger role in system 2 processing than De Neys recognizes. First, however, I want to make a comment about exclusivity.

De Neys argues that if switching is under system 1 control, then exclusivity cannot hold, because system 1 initiates a switch when it generates both the intuitive and the deliberate response and is uncertain which to select (target article, sect. 2.2). This is too strong, however. For as De Neys acknowledges, system 1 can also initiate a switch when it has generated just one response or no response at all (target article, sect. 4.4). So *some* switching could occur even if exclusivity held. However, I think we should deny exclusivity all the same. De Neys presents empirical evidence against it, and he may be right that switching is *often* triggered by conflict within system 1. Moreover, as De Neys notes, system 1 may include automatized versions of system 2 processes, and if it does, then exclusivity will not hold. The upshot is that while we should reject the strong exclusivity claim that *no* system 2 response can be generated by system 1, we should not endorse the strong *inclusivity* claim that *every* system 2 response can be generated by system 1.

On now to the larger issue. I agree with De Neys that system 1 plays a role in controlling system 2, but I think we need to go further – much further. System 1, I propose, does not only initiate and monitor system 2 processes; it also *generates* them. I have developed this idea in previous work (e.g., Frankish, 2009, 2018, 2021), so I shall merely sketch it here.

The core idea is that system 2 processing involves the conscious manipulation of culturally transmitted symbols – words, numerals, diagrams, and so on – either external or, more often, mentally imaged. The manipulations are generated by system 1, and they serve to break down the original problem into simpler subproblems which system 1 can solve. I have described the process as one of *deliberative mastication*. If all goes well, it culminates in a solution to the original problem.

As an example, take division. We can solve simple division problems intuitively, but we deal with more complex ones by executing a procedure for long division, writing down dividend and divisor in a certain format, solving the simpler problems the format highlights, writing down the answers to these problems, and so on, till we have our answer. This is, I suggest, an example – albeit an unusually explicit one – of slow, effortful, system 2 reasoning, and it is under continuous system 1 control. System 1 initiates the actions involved (writing and manipulating the numerals), receives relevant perceptual inputs, recognizes the subproblems posed, solves these subproblems, and so on – all the while monitoring to see if a solution to the overall problem has been reached.

All system 2 processes, I suggest, are similar, being constituted by activities that decompose a problem into intuitively solvable chunks, though these activities are usually internalized ones involving the manipulation of inner speech or other mental imagery rather than external symbols (for examples, see the works cited above).

This proposal explains why system 2 processing places heavy demands on attention and working memory (which are required for imagery manipulation) and why its processes are transparent (the images can be recalled and reported). Moreover, it offers an economical answer to the implementation question I mentioned at the start. In this view, the core cognitive resources driving system 2 processing are those of system 1, though further resources, including those of working memory, language, and perception, are employed as well. Thus, system 2 is not a separate neural system but a *virtual* system, realized in activities generated by system 1.

This view extends the approach De Neys proposes, and his speculations about how system 1 controls system 2 could be elaborated to reflect system 1's expanded role. At each stage in a system 2 process, system 1 will calculate what activity to generate next, receive perceptual or imagistic feedback, generate responses to the subproblem presented, and calculate whether and how to continue the process, using techniques of the sort De Neys describes, including uncertainty monitoring and calculation of opportunity costs.

In conclusion, De Neys's proposal not only advances theorizing about fast-and-slow thinking but also points to how it might be advanced still further, moving us toward a dual-process theory 3.0 in which system 1 not only initiates system 2 thinking but generates and sustains it as well.

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Competing interest. None.

References

Frankish, K. (2009). Systems and levels: Dual-system theories and the personal-subpersonal distinction. In J. S. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 89–107). Oxford University Press.
Frankish, K. (2018). Inner speech and outer thought. In P. Langland-Hassan & A. Vicente (Eds.), *Inner speech: New voices* (pp. 221–243). Oxford University Press.
Frankish, K. (2021). Technology and the human minds. In R. W. Clowes, K. Gärtner, & I. Hipólito (Eds.), *The mind-technology problem: Investigating minds, selves and 21st century artefacts* (pp. 65–82). Springer.

Dual-process moral judgment beyond fast and slow

Joshua D. Greene

Department of Psychology, Center for Brain Science, Harvard University, Cambridge, MA, USA
jgreene@wjh.harvard.edu
https://www.joshua-greene.net

doi:10.1017/S0140525X22003193, e123

Abstract

De Neys makes a compelling case that the sacrificial moral dilemmas do not elicit competing "fast and slow" processes. But are there even two processes? Or just two intuitions? There remains strong evidence, most notably from lesion studies, that sacrificial dilemmas engage distinct cognitive processes generating conflicting emotional and rational responses. The dual-process theory gets much right, but needs revision.

As a proponent of the dual-process theory of moral judgment (Greene, 2013; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001, 2004, 2008) I had, following Kahneman (2003, 2011), long thought of its dual processes as respectively "fast and slow." De Neys makes a compelling case that this is not so. One might conclude, then, that the dual-process theory of moral judgment should be retired. But that, I believe, would be a mistake. There remains strong evidence that moral dilemmas elicit competing responses that are supported by distinct cognitive systems, that one response is meaningfully characterized as more emotional, and that the other is meaningfully characterized as more rational. We may simply need to drop the idea that one response gets a head start in decision making.

For the uninitiated… we are considering sacrificial moral dilemmas such as the *footbridge* case in which one can save five lives by pushing someone in front of a speeding trolley (Thomson, 1985). According to the dual-process theory, the characteristically deontological response (that it's wrong to push) is supported by an intuitive negative emotional response to the harmful action, whereas the characteristically utilitarian judgment (that it's morally acceptable to push) is driven by more deliberative cost–benefit reasoning. What's now in question is whether the utilitarian judgment is in fact more deliberative and less intuitive.

There is accumulating evidence that the utilitarian response is not slower (Bago & De Neys, 2019; Baron & Gürçay, 2017; Cova et al., 2021; Gürçay & Baron, 2017; Koop, 2013; Rosas, 2019; Tinghög et al., 2016), despite a body of evidence indicating that it is (Greene, Morelli, Lowenberg, Nystrom, & Cohen, 2008; Suter & Hertwig, 2011; Trémolière, De Neys, & Bonnefon, 2012). De Neys argues that such dilemmas simply involve two competing intuitions, and he gives no reason to think that they are driven by distinct processes. And yet, if one looks beyond reaction times and cognitive load, evidence for distinct processes abounds.

I can't properly review this evidence here, but I can describe some highlights. Here I focus on studies of lesion patients, which have produced some of the most dramatic evidence supporting the dual-process theory. Koenigs et al. (2007) showed that patients with damage to the ventromedial prefrontal cortex are overwhelmingly more likely to make utilitarian judgments compared to healthy patients and brain-damaged controls. What's more, these patients have impaired emotional responses, as demonstrated by skin-conductance data. Similar results with VMPFC patients are reported by Ciaramelli, Muccioli, Làdavas, and Di Pellegrino (2007), Thomas, Croft, and Tranel (2011), and Moretto, Làdavas, Mattioli, and Di Pellegrino (2010). Demonstrating the opposite effect, McCormick, Rosenthal, Miller, and Maguire (2016) show that patients with hippocampal damage are overwhelmingly more likely to give deontological responses, and they provide parallel evidence using both skin-conductance data and patient self-reports that these responses are because of dominant emotional responses. Verfaellie, Hunsberger, and Keane (2021) report similar results. (But see Craver et al., 2016.) Finally, and most recently, van Honk et al. (2022) show that patients with damage to the basolateral amygdala (implicated in goal-directed decision making) is associated with increased deontological judgment. And here, too, the effects appear to be because of dominant emotional responses. (Note that the basolateral amygdala is distinct from the central-medial amygdala, which is associated with classic affective responses and is what psychologists typically think of as "the amygdala.")

Cushman (2013) has reconceptualized the dual-process theory as a contrast between model-based and model-free algorithms for learning and decision making (Balleine & O'Doherty, 2010; Sutton & Barto, 2018). (See also Crockett, 2013.) Model-based judgment is based on an explicit representation of cause–effect relationships between actions and outcomes, plus values attached to outcomes. Model-free judgment depends, instead, on accessing values attached directly to actions based on prior reinforcement. Cushman and colleagues have since provided compelling evidence that utilitarian judgments are model-based, while deontological judgments are driven by model-free responses (Miller & Cushman, 2013; Patil et al., 2021). Moreover, the model-based/model-free distinction specifically explains why patients with hippocampal damage and basolateral amygdala damage make fewer utilitarian judgments (McCormick et al., 2016; van Honk et al., 2022). As Cushman emphasizes, model-based judgment is not emotion free, as value must be attached to outcomes. But as the patient data indicate, not all emotion is equally emotional.

Putting all of this together, the following picture emerges: Deontological and utilitarian judgments are driven by different processes, as indicated by the contrasting effects of damage to different brain regions. And yet the behavioral data suggest that neither of these processes is reliably faster than the other. Should we say that both responses are intuitive, as De Neys suggests? Yes, in a sense. Both responses come to mind quickly, and further processing is needed to adjudicate between them. (See Shenhav & Greene, 2014, on how these responses may be integrated.) But there is an important sense in which the deontological response is more intuitive. It is based on a feeling that the action is wrong. And, in dilemmas like the *footbridge* case, this feeling is affected by whether the action involves pushing versus hitting a switch (Bago et al., 2022; Greene et al., 2009). This sensitivity to the physical mechanism of harm is unconscious (Cushman, Young, & Hauser, 2006) and not easy to rationally defend (Greene, 2014, 2017). By contrast, the model-based response is based on an explicit understanding of cost and benefits. This may not require much deliberation, at least when it's just five lives versus one, but it is recognizably rational. Indeed, such judgments are correlated with a range of judgments in non-moral contexts that are unequivocally rational (Patil et al., 2021).

All of this suggests that the dual-process theory's fundamental distinction between emotional and rational responses remains intact, but with the surprising twist, supported by De Neys's synthesis, that it's not about fast versus slow.

Competing interest. None.

References

Bago, B., & De Neys, W. (2019). The intuitive greater good: Testing the corrective dual process model of moral cognition. *Journal of Experimental Psychology: General*, *148* (10), 1782.
Bago, B., Kovacs, M., Protzko, J., Nagy, T., Kekecs, Z., Palfi, B., … Matibag, C. J. (2022). Situational factors shape moral judgements in the trolley dilemma in eastern, southern and western countries in a culturally diverse sample. *Nature Human Behaviour*, *6*(6), 880–895.
Balleine, B. W., & O'Doherty, J. P. (2010). Human and rodent homologies in action control: Corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology*, *35*(1), 48–69.
Baron, J., & Gürçay, B. (2017). A meta-analysis of response-time tests of the sequential two-systems model of moral judgment. *Memory & Cognition*, *45*(4), 566–575.
Ciaramelli, E., Muccioli, M., Làdavas, E., & Di Pellegrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience*, *2*(2), 84–92.
Cova, F., Strickland, B., Abatista, A., Allard, A., Andow, J., Attie, M., … Zhou, X. (2021). Estimating the reproducibility of experimental philosophy. *Review of Philosophy and Psychology*, *12*(1), 9–44.
Craver, C. F., Keven, N., Kwan, D., Kurczek, J., Duff, M. C., & Rosenbaum, R. S. (2016). Moral judgment in episodic amnesia. *Hippocampus*, *26*(8), 975–979.
Crockett, M. J. (2013). Models of morality. *Trends in Cognitive Sciences*, *17*(8), 363–366.
Cushman, F. (2013). Action, outcome, and value: A dual-system framework for morality. *Personality and Social Psychology Review*, *17*(3), 273–292.
Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science*, *17* (12), 1082–1089.
Greene, J. (2013). *Moral tribes: Emotion, reason, and the gap between us and them.* Penguin.
Greene, J. D. (2014). Beyond point-and-shoot morality: Why cognitive (neuro) science matters for ethics. *Ethics*, *124*(4), 695–726.
Greene, J. D. (2017). The rat-a-gorical imperative: Moral intuition and the limits of affective learning. *Cognition*, *167*, 66–77.
Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition*, *111*(3), 364–371.
Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107* (3), 1144–1154.
Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*(2), 389–400.
Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science (New York, N.Y.)*, *293*(5537), 2105–2108.
Gürçay, B., & Baron, J. (2017). Challenges for the sequential two-system model of moral judgement. *Thinking & Reasoning*, *23*(1), 49–80.
Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, *58*(9), 697.
Kahneman, D. (2011). *Thinking, fast and slow.* Macmillan.
Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, *446*(7138), 908–911.
Koop, G. J. (2013). An assessment of the temporal dynamics of moral decisions. *Judgment and Decision Making*, *8*(5), 527.
McCormick, C., Rosenthal, C. R., Miller, T. D., & Maguire, E. A. (2016). Hippocampal damage increases deontological responses during moral decision making. *Journal of Neuroscience*, *36*(48), 12157–12167.
Miller, R., & Cushman, F. (2013). Aversive for me, wrong for you: First-person behavioral aversions underlie the moral condemnation of harm. *Social and Personality Psychology Compass*, *7*(10), 707–718.
Moretto, G., Làdavas, E., Mattioli, F., & Di Pellegrino, G. (2010). A psychophysiological investigation of moral judgment after ventromedial prefrontal damage. *Journal of Cognitive Neuroscience*, *22*(8), 1888–1899.
Patil, I., Zucchelli, M. M., Kool, W., Campbell, S., Fornasier, F., Calò, M., … Cushman, F. (2021). Reasoning supports utilitarian resolutions to moral dilemmas across diverse measures. *Journal of Personality and Social Psychology*, *120*(2), 443.
Rosas, A., Bermúdez, J. P., & Aguilar-Pardo, D. (2019). Decision conflict drives reaction times and utilitarian responses in sacrificial dilemmas. *Judgment & Decision Making*, *14*(5), 555–564.
Shenhav, A., & Greene, J. D. (2014). Integrative moral judgment: Dissociating the roles of the amygdala and ventromedial prefrontal cortex. *Journal of Neuroscience*, *34*(13), 4741–4749.
Suter, R. S., & Hertwig, R. (2011). Time and moral judgment. *Cognition*, *119*(3), 454–458.
Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction.* MIT Press.
Thomas, B. C., Croft, K. E., & Tranel, D. (2011). Harming kin to save strangers: Further evidence for abnormally utilitarian moral judgments after ventromedial prefrontal damage. *Journal of Cognitive Neuroscience*, *23*(9), 2186–2196.
Thomson, J. J. (1985). The trolley problem. *Yale Law Journal*, *94*(6), 1395–1415.
Tinghög, G., Andersson, D., Bonn, C., Johannesson, M., Kirchler, M., Koppel, L., & Västfjäll, D. (2016). Intuition and moral decision-making – The effect of time pressure and cognitive load on moral judgment and altruistic behavior. *PLoS ONE*, *11*(10), e0164012.
Trémolière, B., De Neys, W., & Bonnefon, J. F. (2012). Mortality salience and morality: Thinking about death makes people less utilitarian. *Cognition*, *124*(3), 379–384.
van Honk, J., Terburg, D., Montoya, E. R., Grafman, J., Stein, D. J., & Morgan, B. (2022). Breakdown of utilitarian moral judgement after basolateral amygdala damage. *Proceedings of the National Academy of Sciences*, *119*(31), e2119072119.
Verfaellie, M., Hunsberger, R., & Keane, M. M. (2021). Episodic processes in moral decisions: Evidence from medial temporal lobe amnesia. *Hippocampus*, *31*(6), 569–579.

More than two intuitions

David J. Grüning[a,b] and Joachim I. Krueger[c]

[a]Psychology Department, Heidelberg University, Heidelberg, Germany; [b]Department of Survey Design and Methodology, GESIS – Leibniz Institute for the Social Sciences, Mannheim, Germany and [c]Cognitive, Linguistic & Psychological Sciences, Brown University, Providence, RI, USA
david.gruening@psychologie.uni-heidelberg.de
joachim_krueger@brown.edu

doi:10.1017/S0140525X22002965, e124

Abstract

We consider an underdeveloped feature of De Neys's model. Decisions with multiple intuitions per option are neither trivial to explain nor rare. These decision scenarios are crucial for an assessment of the model's generalizability and adequacy. Besides monitoring absolute differences in intuition strength, the mind might add the strengths of intuitions per choice option, leading to competing and testable hypotheses.

The first stage of De Neys's model, the processing of intuitions, requires elaboration. We here respond to two of the author's key assumptions. The first assumption is that "the uncertainty parameter might focus on the absolute difference" (target article, sect. 4.4, para. 1) between the strongest competing intuitions. The second assumption is that decision-making scenarios with more than two intuitions are "a-typical cases" (target article, sect. 4.4, para. 1). As to the former assumption, we show, by example, that there are several different ways in which decision makers might process more than two competing intuitions. As to the latter assumption, we argue that having multiple intuitions can be considered the norm rather than an anomaly.

For simplicity, we will only consider decisions with two choice options. Extending the author's (target article, sect. 3.2, para. 4) example, suppose John has to choose between a cupcake and an apple for dessert. Although John's first intuition (I1) favors the cupcake for its sweet taste, two other intuitions come to mind. The second intuition (I2) is the realization that an apple is tasty too, and the third intuition (I3) is that the apple is healthier than the cupcake. Like De Neys, we assume that these intuitions differ in strength. Although the cupcake is tastier (with a weight of 0.80) than the apple (0.60), the apple's healthiness is also noteworthy (0.50). We can now imagine two pathways for the intuitive response. In one pathway, the strongest intuition, I1 wins, and John decides to eat the cupcake. This is the outcome De Neys's model predicts from the monitoring of the absolute difference between the strongest competing intuitions. In the other pathway, the combined strengths of intuitions I2 and I3 override the strength of I1; John eats the apple because its acceptable tastiness and evident health benefit together trump the allure of the cupcake's sweetness (Anderson, 1981; Juslin, Karlsson, & Olsson, 2008).

Limited comparisons between the strengths of the strongest intuitions become less compelling as the number of intuitions favoring consumption of the apple grows, even if each additional intuition is weak. De Neys's process assumption neglects the possibility that separate but weak intuitions may amount to a strong incentive for choice. Here, the absolute difference between the two strongest intuitions has not changed as medium tastiness remains the primary intuition favoring the apple. By contrast, it is even conceivable that the absolute difference between cupcake and apple increases because the average intuition strength in favor of the apple decreases with each weak additional intuition in favor of it. Table 1 shows both possibilities of absolute difference and also the simple additive processing.

Table 1 (Grüning and Krueger). Different processing styles of multiple intuitions in competing options

Intuitions	Cupcake	Apple
Intuition 1	0.80	0.60
Intuition 2	–	0.50
Intuition 3	–	0.10
Intuition 4	–	0.10
Intuition 5	–	0.10
Difference between strongest competing intuitions	0.80	0.60
Difference between means of competing groups of intuitions	0.80	0.28
Simple addition of intuitions	0.80	1.40

This multiple-intuitions example is one of many non-trivial cases where individuals process multiple intuitions for competing options. At this early stage of model building, purely theoretical exploration and thought experiments are a productive beginning. Yet experiments and empirical research are needed to determine the conditions allowing the rise of multiple intuitions and to understand how people process them. A first step could be to test competing hypotheses about how decision makers consolidate multiple intuitions (e.g., absolute difference vs. addition) by presenting them with combinations of intuitions that lead to different decisions depending on the assumed processing.

We recognize De Neys's call for research on cases with more than two competing intuitions, but we disagree with his emphasis. Research on cue usage and information integration (e.g., Candolin, 2003; Grüning, Alves, Mata, & Fiedler, in press; Gunes, Piccardi, & Pantic, 2008; Plessner, Schweizer, Brand, & O'Hare, 2009) shows that multiple intuitions are the norm in decision making. Although novel situations may end in deliberation because intuitions are missing, familiar situations offer a rich range of different cues (e.g., color, availability, smell, and nutrition value) supporting the rise of multiple intuitions in the decision maker's mind. Accordingly, we regard the question of how decision makers process multiple competing intuitions to be pressing, especially for cases where individuals have learned new features of choice objects (e.g., that an apple also provides more energy than a cupcake and it requires less energy to be produced). Naturally, many familiar decision-making situations will be dominated by a few very strong intuitions per option. However, for these situations, too, our example above necessitates thinking about how more than two competing intuitions are consolidated. At the limit, we argue, the presence of many weak intuitions in favor of option B can overcome a single strong intuition favoring the alternative option A.

Conceptualizing decision makers' multiple-intuitions processing as an act of addition instead of monitoring absolute differences changes the predictions about whether thoughtful reflection will occur. The view that individuals compare their strongest intuitions by assessing the absolute difference yields a clear prediction regarding the onset of deliberation: If the two strongest competing intuitions are close enough, the resulting feeling of indifference triggers the need for deliberate thinking. Adding a large number of weak intuitions to one option (i.e., the apple) should not change the outcome. Assuming additive processing, however, decision makers with added weak intuitions in favor of the apple should come to a point of indifference about choosing the apple or the cupcake; exactly when the strong intuition about the cupcake's tastiness is matched by the composite of the apple's lower tastiness and the additional intuitions favoring it. Again, our main argument is not that addition is the more probable basic element behind intuition processing. This has to be tested. We suggest that thinking about how multiple intuitions are processed is central to predicting not only which intuitive decisions are made but also when deliberation occurs. How decision makers process their multiple intuitions is the fundamental predictive mechanism of De Neys's model and understanding it is a key challenge for his theory.

In conclusion, we welcome De Neys presenting an intriguing and novel model of decision making. Nevertheless, we think that one central aspect of the model, namely how decision makers process multiple intuitions, requires more attention. Theoretical and empirical advancements in understanding this process are possible and would move De Neys's model closer to a general theory of decision making.

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Competing interest. None.

References

Anderson, N. H. (1981). *Foundations of information integration theory*. Academic Press..

Candolin, U. (2003). The use of multiple cues in mate choice. *Biological Reviews*, *78*(4), 575–595. https://doi.org/10.1017/S1464793103006158

Grüning, D. J., Alves, H., Mata, A. O. P., & Fiedler, K. (in press). Reversing the cumulative redundancy bias to demonstrate metacognitive flexibility in cue utilization. *Journal of Experimental Social Psychology*.

Gunes, H., Piccardi, M., & Pantic, M. (2008). From the lab to the real world: Affect recognition using multiple cues and modalities. In J. Or (Ed.), *Affective computing: Focus on emotion expression, synthesis, and recognition* (pp. 185–218). InTechOpen.

Juslin, P., Karlsson, L., & Olsson, H. (2008). Information integration in multiple cue judgment: A division of labor hypothesis. *Cognition*, *106*(1), 259–298. https://doi.org/10.1016/j.cognition.2007.02.003

Plessner, H., Schweizer, G., Brand, R., & O'Hare, D. (2009). A multiple-cue learning approach as the basis for understanding and improving soccer referees' decision making. *Progress in Brain Research*, *174*, 151–158. https://doi.org/10.1016/S0079-6123(09)01313-2

Illusory intuitions: Challenging the claim of non-exclusivity

Simon J. Handley[a], Omid Ghasemi[b] and Michal Bialek[c]

[a]School of Psychological Sciences, Macquarie University, Macquarie Park, NSW, Australia; [b]School of Psychology, University of New South Wales, Sydney, NSW, Australia and [c]Institute of Psychology, University of Wroclaw, Wrocław, Poland

simon.handley@mq.edu.au
o.ghasemi@unsw.edu.au
michal.bialek3@uwr.edu.pl
https://researchers.mq.edu.au/en/persons/simon-handley
https://scholar.google.com/citations?user=LyL_5s8AAAAJ&hl=en
https://sites.google.com/view/michalbialek

doi:10.1017/S0140525X22003168, e125

Abstract

A person who arrives at correct solutions via false premises is right and wrong simultaneously. Similarly, a person who generates "logical intuitions" through superficial heuristics can likewise be right and wrong at the same time. However, heuristics aren't guaranteed to deliver the logical solution, so the claim that system 1 can routinely produce the alleged system 2 response is unfounded.

De Neys sets out two of the key challenges for dual-process theories as they have been traditionally conceived, the detection of response conflict and specification of the mechanism that controls the switch between system 1 (S1) and system 2 (S2) processing. In this commentary, we focus on first of these concerns and in particular the role of S1 in the generation of the "alleged S2 response"; the claim that reasoners routinely generate "logical intuitions." Evidence for logical intuitions has been claimed across a range of paradigms, where there is evidence that conflict between knowledge-based and rule-based responses can be detected automatically, without the engagement of deliberative reasoning. According to De Neys, the proposal that S1 can generate a logical response solves one of the fundamental challenges for the dual-process framework; how can conflict be detected without the prior engagement of S2 in calculating the normative response? The solution to this quandary is to posit non-exclusivity, the idea that the generation of a logical response is not the unique purview of S2.

For the past decade, our own work has similarly suggested that reasoners show intuitive sensitivity to the logical validity of simple deductive reasoning arguments. One of the most convincing pieces of evidence comes from belief–logic instructional paradigm, where, across multiple studies, we have shown that a conflict between the logical and belief status of a conclusion influences judgments of conclusion believability as much, if not more than conclusion believability influences logical judgments (Handley, Newstead, & Trippas, 2011; Howarth, Handley, & Walsh, 2016, 2019; Trippas, Thompson, & Handley, 2017). The impact of conflict on belief judgments indicates that the logical inference is drawn automatically and intuitively and hence interferes with fast belief judgments. These findings appear to support De Neys claim for non-exclusivity, that S1 can generate logical intuitions.

However, our most recent research, together with parallel findings from other labs, suggests that the picture is not quite so straightforward. These findings suggest that "logical intuitions" may have little to do with formal logic, but instead reflect sensitivity to superficial structural features (Ghasemi, Handley, Howarth, Newman, & Thompson, 2022a; Meyer-Grant et al., 2022). Converging evidence from research on the liking-logic task and the instructional paradigm show that logical intuition effects emerge because on valid arguments there is a match between the polarity of the premises and the conclusion, which is not present on invalid versions of the same arguments. Ghasemi et al. (2022a) tested this explanation by using invalid arguments in which such a match was present, and showed that "logical intuition" effects are equally as strong on these invalid argument forms. They concluded that reasoners are not intuitively sensitive to logical validity in a formal sense, instead they are picking up on structural cues that reflect the repetition of elements in the premises. Our most recent research further shows that training in logical principles improves discrimination between valid and invalid logical forms under logical instructions, but does not reduce the propositional matching effect under belief instructions, providing convincing evidence that logical intuitions arise because of sensitivity to non-logical features rather than logical validity per se (Ghasemi, Handley, & Stephens, 2022b).

De Neys argues that although the S2 response may be generated by S1 processes, the equivalence is situated at the response level. Hence, an equivalent response generated by the intuitive and deliberative systems does not imply that the response was generated by the same mechanism or has the same features. Perhaps S1 does in fact rely on heuristics that draw on surface features but these co-vary with the logical status of the conclusion, hence on average deliver an intuition that aligns with the S2 response? So, does it matter if the intuitive response is generated by heuristics? We argue that it does, because a heuristic is no guarantee to a logical conclusion. In fact, more than half a century of work shows that heuristics regularly lead to systematic errors in reasoning and judgment tasks (Tversky & Kahneman,

1983) and the alignment of the output of a heuristic mechanism with the logically correct response can often be a matter of chance or clever experimental design (Evans & Lynch, 1973; Handley & Evans, 2000). We have recently run a series of studies in which the output of a matching heuristic and the logical response were misaligned. In these circumstances, matching dominates S1 outputs, whereas logic dominates S2 (Ghasemi, Handley, & Howarth, 2023). What these studies illustrate is that there is no guarantee that a response based upon superficial problem features will align with the formal logical response. You might get it right for the wrong reason, but you are as likely to get it wrong for the wrong reason also.

Is there a way of reconciling our findings with De Neys model? We think that there is at least one resolution which draws upon alternative normative accounts of human thinking that do not rely on formal logic as a normative standard. Such accounts are framed within the new paradigm in the psychology of reasoning which edifies the usage of heuristics as an adaptive mechanism, sensitive to probabilistic logic or information gain (Oaksford & Chater, 2020). An intriguing possibility is that logical intuitions, while arising through the application of simple heuristics, nevertheless respect the probabilistic structure of the environment and hence deliver outputs that have a rational basis. Such outputs will often align with an S2 response that draws upon deductive logic and hence the non-exclusivity principle will often, but not always hold. Perhaps intuitive reasoners do indeed sometimes get it right, but for a different reason, not the wrong one.

Financial support. This research received no specific grant funding from any funding agency, commercial, or not-for-profit sectors.

Competing interest. None.

References

Evans, J. S. B., & Lynch, J. S. (1973). Matching bias in the selection task. *British Journal of Psychology*, *64*(3), 391–397.

Ghasemi, O., Handley, S., & Howarth, S. (2023). Illusory intuitive inferences: Matching heuristics explain logical intuitions. *Cognition, 235*. https://doi.org/10.1016/j.cognition.2023.105417

Ghasemi, O., Handley, S., Howarth, S., Newman, I. R., & Thompson, V. A. (2022a). Logical intuition is not really about logic. *Journal of Experimental Psychology: General*, *151*, 2009–2028.

Ghasemi, O., Handley, S., & Stephens, R. (2022b). Logical intuitions or matching heuristic? Examining the effect of deduction training on belief-based reasoning judgments. https://doi.org/10.31219/osf.io/gxd73

Handley, S. J., & Evans, J. S. B. (2000). Supposition and representation in human reasoning. *Thinking & Reasoning*, *6*(4), 273–311.

Handley, S. J., Newstead, S. E., & Trippas, D. (2011). Logic, beliefs, and instruction: A test of the default interventionist account of belief bias. *Journal of Experimental Psychology: Learning Memory and Cognition*, *37*, 28–43. https://doi.org/10.1037/a0021098

Howarth, S., Handley, S., & Walsh, C. (2019). The logic sense: Exploring the role of executive functioning in belief and logic-based judgments. *Thinking & Reasoning*, *25*(1), 416–448. https://doi.org/10.1080/13546783.2018.1523808

Howarth, S., Handley, S. J., & Walsh, C. (2016). The logic-bias effect: The role of effortful processing in the resolution of belief–logic conflict. *Memory and Cognition*, *44*, 330–349. https://doi.org/10.3758/s13421-015-0555-x

Meyer-Grant, C. G., Cruz, N., Singmann, H., Winiger, S., Goswami, S., Hayes, B. K., & Klauer, K. C. (2022). Are logical intuitions only make-believe? Reexamining the logic-liking effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://psycnet.apa.org/doi/10.1037/xlm0001152

Oaksford, M., & Chater, N. (2020). New paradigms in the psychology of reasoning. *Annual Review of Psychology*, *71*(1), 305–330.

Trippas, D., Thompson, V. A., & Handley, S. J. (2017). When fast logic meets slow belief: Evidence for a parallel-processing model of belief bias. *Memory and Cognition*, *45*(4), 539–552. https://doi.org/10.3758/s13421-016-0680-1

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*, 293–315. https://doi.org/10.1037/0033-295X.90.4.293

Unifying theories of reasoning and decision making

Brett K. Hayes[a], Rachel G. Stephens[b] and John C. Dunn[c]

[a]University of New South Wales, Sydney, Australia; [b]University of Adelaide, Adelaide, Australia and [c]University of Western Australia, Perth, Australia
b.hayes@unsw.edu.au
rachel.stephens@adelaide.edu.au
john.dunn@uwa.edu.au
https://www.unsw.edu.au/staff/brett-hayes

doi:10.1017/S0140525X22002916, e126

Abstract

De Neys offers a welcome departure from the dual-process accounts that have dominated theorizing about reasoning. However, we see little justification for retaining the distinction between intuition and deliberation. Instead, reasoning can be treated as a case of multiple-cue decision making. Reasoning phenomena can then be explained by decision-making models that supply the processing details missing from De Neys's framework.

This provocative target article questions several key assumptions of popular dual-process models of reasoning and outlines a novel cognitive architecture for explaining the relationship between intuition and deliberation. A central argument is that a reasoner can have multiple, competing "intuitions" about the correct solution to a problem, with the activation strength of each intuition varying over time. We see this as a potentially valuable step in theory building in the field of human reasoning, that brings the field closer to other productive areas of research in human judgment and decision making.

We question, however, whether there is any need to retain the distinction between intuitive and deliberative processing. If we allow for multiple intuitions, some of which align with normative principles of probability or logic, there seems little need for an additional deliberative system. This is evident in the target article, where De Neys struggles to define a unique role for "deliberation." One suggestion is that deliberation involves the application of an algorithm or execution of a set of rules when solving a problem. But given that such rules can become automated with experience (Logan & Klapp, 1991), this seems like a weak definition. Another suggestion by De Neys removes deliberation from the decision-making process altogether – relegating it the role of rationalizing or justifying decisions that have already been made.

As an alternative approach, we suggest that the notion of multiple "intuitions" should be re-framed in more general terms as attention to multiple cues that define alternative decision options. In this approach, reasoning in tasks like ratio bias, moral judgment, or verbal syllogisms, can be captured by the same general cognitive architecture used to explain other decisions involving multiple cues or features. To illustrate the basic idea, consider planning to purchase a new car. This is likely to involve consideration of multiple cues (e.g., electric vs. petrol power source, price, manufacturer's reputation). As in the most interesting reasoning problems, these cues will often be in conflict (e.g., electric cars are more environmentally sustainable but are often more expensive). A theoretical model of decision making in such cases needs to explain how the various cues are weighted when comparing options and how trade-offs between cues take place. In this framework, decision-making cues may vary in complexity, salience, and familiarity. However, there is no need to assume discrete types of processing (e.g., intuition vs. deliberation, system 1 vs. system 2) for dealing with different cues.

A key implication is that models of multiple-cue decision making (e.g., Busemeyer & Townsend, 1993) can be applied to understand reasoning phenomena. We believe that this has many advantages. For one thing, the processing assumptions of these decision-making models have been laid out in far more detail, and been subject to more extensive empirical testing, than the architecture sketched by De Neys. For example, following the structure of popular "evidence accumulation" models of decision making (Busemeyer & Townsend, 1993; Ratcliff, Smith, Brown, & McKoon, 2016), reasoning could be thought of as a process of the dynamic accumulation of evidence relevant to each decision option (e.g., options based on absolute number vs. ratio in ratio bias problems; utilitarian vs. deontological responses in moral judgments; judging whether a verbal argument is valid or invalid). A decision is made when the evidence for a given option reaches a threshold. Unlike De Neys's approach, such models provide a principled account of how and why the "activation strength" associated with each cue changes over time (cf. Ratcliff et al., 2016). They also explain how the accumulation of evidence for each cue interacts with other components of the decision-making process such as how one sets a decision threshold and how one encodes the relevant cues.

Together these model components have the prospect of explaining many key reasoning phenomena. For example, the fact that arguments with believable conclusions are more likely to be judged as valid regardless of logical structure (Dube, Rotello, & Heit, 2010), may be explained by assuming that believable arguments have a higher "start-point" for evidence accumulation than unbelievable arguments. Hence, they require less evidence to reach threshold for a "valid" response. Higher rates of endorsement of arguments based on their believability rather than validity under time pressure (Evans & Curtis-Holmes, 2005; Hayes, Stephens, Ngo, & Dunn, 2018) can be explained by adjustment of the relevant decision thresholds. Evidence accumulation models are also well-equipped to explain the inconsistency we often see in individual reasoning patterns, such as shifts between utilitarian and deontological options across different moral judgments (e.g., Cushman, Young, & Hauser, 2006) or shifts between a focus on the visual appearance of text as opposed to logical structure or argument plausibility in verbal reasoning (Hayes et al., 2022). Such shifts can be explained as context-driven changes in the rate of evidence accumulation for rival decision options.

To date, evidence accumulation models have most often been applied to simple perceptual decisions. However, there is good evidence that they "scale-up" to capturing the processes involved in complex decisions that more closely resemble those involved in reasoning tasks (e.g., Hawkins, Hayes, & Heit, 2016; Krajbich, Bartling, Hare, & Fehr, 2015; Palada et al., 2016).

In sum, the approach suggested by De Neys is a welcome departure from the dual-process accounts that have dominated recent theorizing about human reasoning. Retaining a hard distinction between intuitive and deliberative processes (regardless of whether this distinction is viewed as "qualitative" or "quantitative"), however, does little to advance our understanding. Instead, we suggest that reasoning in classic conflict tasks be treated as a special case of multiple-cue decision making. Doing so will allow us to apply powerful theoretical models that supply much of the processing detail missing from the architecture proposed by De Neys.

Financial support. This work was supported by Australian Research Discovery Grant, DP190102160 to B.K.H. and J.C.D.

Competing interest. None.

References

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*(3), 432–459. https://doi.org/10.1037/0033-295X.100.3.432

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science*, *17* (12), 1082–1089. https://doi.org/10.1111/j.1467-9280.2006.01834.x

Dube, C., Rotello, C. M., & Heit, E. (2010). Assessing the belief bias effect with ROCs: It's a response bias effect. *Psychological Review*, *117*(3), 831–863. https://psycnet.apa.org/doi/10.1037/a0019634

Evans, J. St. B. T., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning*, *11*, 382–389. http://dx.doi.org/10.1080/13546780542000005

Hawkins, G. E., Hayes, B. K., & Heit, E. (2016). A dynamic model of reasoning and memory. *Journal of Experimental Psychology: General*, *145*, 155–180. doi: 10.1037/xge0000113

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

Hayes, B. K., Stephens, R. G., Ngo, J., & Dunn, J. C. (2018). The dimensionality of reasoning: Evidence for a single process account of inductive and deductive inference. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *44*, (9), 1333–1351. https://doi.org/10.1037/xlm0000527

Krajbich, I., Bartling, B., Hare, T., & Fehr, E. (2015). Rethinking fast and slow based on a critique of reaction-time reverse inference. *Nature Communications*, *6*(1), 1–9. doi: 10.1038/ncomms8455

Logan, G. D., & Klapp, S. T. (1991). Automatizing alphabet arithmetic: I. Is extended practice necessary to produce automaticity? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(2), 179–195. https://psycnet.apa.org/doi/10.1037/0278-7393.17.2.179

Palada, H., Neal, A., Vuckovic, A., Martin, R., Samuels, K., & Heathcote, A. (2016). Evidence accumulation in a complex task: Making choices about concurrent multiattribute stimuli under time pressure. *Journal of Experimental Psychology: Applied*, *22*(1), 1–23. https://psycnet.apa.org/doi/10.1037/xap0000074

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, *20*(4), 260–281. https://doi.org/10.1016/j.tics.2016.01.007

Switching between system 1 and system 2: The nature of competing intuitions and the role of disfluency

Maryna Kołeczek and Maciej Sekerdej

Institute of Psychology, Jagiellonian University, Krakow, Poland
m.koleczek@student.uj.edu.pl; maciej.sekerdej@uj.edu.pl

doi:10.1017/S0140525X22003016, e127

Abstract

This commentary identifies two problems concerning the switch mechanism: The model explains too few instances of switching, and the switching mechanism itself seems fallible. The improvements we suggest are to clarify the nature of the competing intuitions as the initial intuition and its negation or alternative ways to solve the problem, and to incorporate cognitive disfluency into the switching mechanism.

Although we agree that the proposed model is innovative and coherent, and that it undoubtedly contributes to the field, its further development may require addressing two problems regarding the switching mechanism. The first problem is that the model explains switching only in a particular type of situation: When two or more competing intuitions are available for system 1, or there are no intuitions at all. The model, however, does not explain the spontaneous activation of system 2 in situations when the "alleged system 1 response" is automatic, but the "alleged system 2 response" cannot be available for system 1 without system 2 activation because it *must* be first produced by the deliberation or calculation of system 2. For example, in the commonly known cognitive reflective task (CRT) bat and ball task, the answer "$0.10" can be available to system 1 because $1.10 − $1.00 is a very easy equation that gets calculated automatically. However, to find the correct answer, more complex calculations are required: finding x if (a) $x + y = \$1.10$ and (b) $y = \$1.00 + x$. It is unlikely that this response can be automatically available to system 1.

De Neys's answer to this problem is the non-exclusivity assumption and the idea that system 1 can produce the correct answer itself. However, if system 1 can conduct such complex operations as the calculations presented above, then why do we even need system 2? De Neys further proposes that the correct intuition may be available to system 1, not thanks to its ability to conduct the calculations, but because of previous exposure to this riddle (or similar riddles) and learning the correct answer. However, by applying such reasoning, the model still does not explain switching when one encounters a completely new problem that requires calculating the correct answer first. It does not even explain switching in the case of a well-known problem presented with different numbers (even if the schema of a riddle is the same, but the numbers are changed, the response still needs to be calculated from scratch). Therefore, non-exclusivity does not fully resolve the problem of switching paradoxes.

Moreover, the proposed switching mechanism does not explain various manipulations that trigger system 2. For example, system 2 deliberation may be turned on by priming (Gervais & Norenzayan, 2012), presenting questions in a way that makes them difficult to read (Alter, Oppenheimer, Epley, & Eyre, 2007; Song & Schwarz, 2008), asking participants to frown during the study (Alter et al., 2007), explicitly asking participants to deliberate on the questions before answering, and other interventions (see Horstmann, Hausmann, & Ryf [2009] for an overview). None of these effects can be explained by means of the proposed model.

The second problem is that the switching mechanism seems not to work in a considerable number of situations. First, a similar activation of competing intuitions should trigger system 2, but this does not always happen. For example, in logical riddles (e.g., syllogisms, base-rate problems) or moral dilemmas, we simultaneously present several alternative answers from which the participant may choose. Therefore, all alternatives should be equally strongly activated and trigger system 2. However, an explicit presentation of alternative responses does not make people more reflective or may even enhance more intuitive processing (Sirota & Juanchich, 2018).

Furthermore, the proposed feedback loop seems fallible. If system 2 deliberation leads to choosing an answer, then it should decrease the relative activation of the rejected intuition. For example, when one solves base-rate problems or assesses probability or randomness (e.g., what is more probable, six heads in a row or head–tail–tail–head–tail–head), deliberation lets them use the probability distributions to find the right answer. However, even though deliberation and formal knowledge allow giving the correct answer with high certainty, the "homunculus" keeps jumping and shouting the intuitive response (Gould, 1992, p. 469; see also Kahneman [2011] and Thompson [2009] for discussions on the subjective feeling that the intuitive answer is correct). This suggests that the relative activation of the initial, intuitive response is still very high.

The solution to these problems is either to clarify the nature of competing intuitions or to propose a more all-encompassing switching trigger. Regarding the first possibility, the competing intuitions should not be pictured as alternative responses (p vs. q) but rather as either an intuitive response and its negation (p vs. not-p) or as two alternative ways of solving the problem (e.g., the simple equation $1.10 −$1.00 vs. applying formal algebra). With competing intuitions defined this way, the non-exclusivity assumption holds: System 1 is still able to generate the alternative intuition. It is impossible for system 1 to have access to certain responses (e.g., based on advanced calculations), but it may have intuitions that the initial response is incorrect (i.e., not-p intuition) or that there are different ways to approach the problem. To put it differently, it is impossible for system 1 to do complex calculations, however, people can still automatise the reaction of suspiciousness in response to logical riddles or the belief that it is better to rely on formal algebra than on a gut feeling.

Nevertheless, even after clarifying the nature of intuitions, the model still does not explain why manipulations such as priming, frowning, explicit instructions, or difficult-to-read font trigger system 2. Therefore, we suggest the incorporation of a more universal switching mechanism into the model. It is not a new idea to identify processing disfluency as the ultimate intuitive trigger of system 2 (e.g., Alter et al., 2007; Thompson, 2009). The general idea behind disfluency is that system 2 is activated if processing within system 1 does not go smoothly. Including a disfluency-triggered switch does not rule out the possibility of the switching being caused by uncertainty, as disfluency should lead to uncertainty (Gill, Swann, & Silvera, 1998). However, this modification will allow the model to explain a wider range of phenomena: Disfluency may be caused by difficult-to-read font, the instruction to think twice before answering, a lack of faith in the intuitive answer, and so on.

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Competing interest. None.

References

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, *136*(4), 569–576. doi:10.1037/0096-3445.136.4.569

Gervais, W. M., & Norenzayan, A. (2012). Analytic thinking promotes religious disbelief. *Science*, *336*(6080), 493–496. doi:10.1126/science.1215647
Gill, M. J., Swann Jr, W. B., & Silvera, D. H. (1998). On the genesis of confidence. *Journal of Personality and Social Psychology*, *75*(5), 1101–1114. doi:10.1037/0022-3514.75.5.1101
Gould, S. J. (1992). *Bully for brontosaurus: Reflections in natural history*. WW Norton.
Horstmann, N., Hausmann, D., & Ryf, S. (2009). Methods for inducing intuitive and deliberate processing modes. In A. Glöckner & C. Witteman (Eds.), *Foundations for tracing intuition* (pp. 227–245). Psychology Press.
Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
Sirota, M., & Juanchich, M. (2018). Effect of response format on cognitive reflection: Validating a two-and four-option multiple choice question version of the cognitive reflection test. *Behavior research methods*, *50*(6), 2511–2522. doi:10.3758/s13428-018-1029-4
Song, H., & Schwarz, N. (2008). Fluency and the detection of misleading questions: Low processing fluency attenuates the Moses illusion. *Social Cognition*, *26*(6), 791–799. doi:10.1521/soco.2008.26.6.791
Thompson, V. A. (2009). Dual-process theories: A metacognitive perspective. In J. S. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 171–195). Oxford University Press. doi:10.1093/acprof:oso/9780199230167.003.0008

Why is system 1/system 2 switching affectively loaded?

Charlie Kurth[a,b]

[a]Helsinki Collegium for Advanced Studies, University of Helsinki, Helsinki, Finland and [b]Department of Philosophy, Western Michigan University, Kalamazoo, MI, USA
charles.kurth@wmich.edu
www.charliekurth.com

doi:10.1017/S0140525X22002989, e128

Abstract

Why are only some occasions of system 1 to system 2 switching affectively loaded? This commentary not only draws attention to this neglected phenomenon, but also shows how research in philosophy and the social and cognitive sciences sheds light on it, doing so in ways that may help answer some of the open questions that De Neys's paper highlights.

We can begin with a question that goes unasked in De Neys's otherwise wide-ranging and insightful essay: Why is it that switching from system 1 to system 2 cognition is so often – though not always – an *affectively loaded* experience? Here I take the affective dimension of switching to be familiar. It appears, for instance, as the feeling of unease that one experiences when an initially routine and mindless task suddenly becomes more difficult and cognitively demanding. But notice that a similar feeling is typically absent when, for instance, cognitive resources are ramped up more gradually. So why do we see this difference? This commentary will not only draw attention to this neglected aspect of "system switching," but also show how research in philosophy and the social and cognitive sciences sheds light on it. The result will be greater clarity on some of the open questions that De Neys raises at the end of his paper.

Like De Neys, philosophers have recognized that dual-process theorists owe us an explanation of system switching. But these philosophers also add substance by highlighting a distinctive family of metacognitive emotions as underlying many of these system 1 to system 2 transitions. These emotions are *metacognitive* in the sense that they function to regulate one's first-order cognitive processing; they are *emotions* in the sense that they are automatically engaged, motivationally laden feelings. More specifically, on these philosophical accounts, metacognitive emotions are viewed as system 1 forms of cognition that use heuristics to map occasions of positive/negative value to distinctive feelings, and then use these feelings to generate an "affective alarm" – a warning that (re)directs attention and engages a distinctive suite of system 2 processing (Arango-Muñoz, 2011; de Sousa, 2008; Kurth, 2015, 2018a). So, returning to the earlier example, those feelings of unease are, on this metacognitive account, to be understood as automatically engaged responses to problematic changes in one's circumstances: Responses that function to warn of potential trouble and prompt higher cognitive processing to help one address the issue at hand.

Importantly, these theoretical proposals in philosophy are supported by empirical work from the social and cognitive sciences. For instance, we have research highlighting the role that *feelings of familiarity* play in the tip-of-the-tongue phenomenon (e.g., Schwartz & Metcalfe, 2014). This work identifies these feelings as mechanisms that function to engage and sustain conscious mnemonic effort after heuristic-based monitoring systems have identified an instance of failed memory with partial recall (e.g., occasions where you cannot remember a person's name, but sense that it begins with a "P"). Similar findings point to the role that *anxiety* plays in prompting deliberation in the face of difficult moral and political policy issues like affirmative action, immigration, and climate change (Fernando et al., 2016; MacKuen et al., 2010; Valentino et al., 2008). Here these feelings of worry are seen as working to flag decisions like these as being particularly difficult, thereby engaging reflection and information gathering efforts to help one work through the issue.

Crucially for our purposes, in both cases the empirical findings implicate the metacognitive role of emotion in these system 1 to system 2 transitions: We see affect acting as a system 1 *alarm* – one that both shifts our attention to the problem at hand and engages system 2 resources to help us address it. Moreover, notice as well that these two examples characterize their target phenomena as involving distinctive triggers, felt experiences, and forms of higher cognitive engagement – features that suggest we have a family of distinct metacognitive emotions here, and not just a single, all-purpose mechanism (Arango-Muñoz, 2011; Kurth, 2018b; Thompson, 2009).

Here's why all this matters. First, the above examples provide us with concrete, empirically grounded models of how central instances of system 1 to system 2 switching may operate, thus responding to De Neys's call for "further fleshed out, fine-tuned, and developed" accounts these processes (target article, sect. 4, para. 1). Second, by understanding these switching mechanisms *as emotions*, these models shed new light on some of the "deliberation issues" that De Nays flags in section 4.3. For instance, given the familiar role that emotions play in *directing attention* and *sustaining effort*, we get the makings of explanations for questions about, respectively, the prioritization of deliberative effort across tasks and the amount of effort expended on a given occasion. Finally, if emotions have the alarm function suggested above, then we also get an answer to the question we started with: The reason why only some cases of system 1 to system 2 switching are affect-engaging is that only some of them are ones that our emotions have deemed to be alarm-worthy.

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Competing interest. None.

References

Arango-Muñoz, S. (2011). Two levels of metacognition. *Philosophia*, *39*, 71–82.

de Sousa, R. (2008). Epistemic feelings. In G. Brun, U. Doğuoğlu, & D. Kuenzle (Eds.), *Epistemology and emotions* (pp. 185–204). Ashgate.

Fernando, A. G., Sivakumaran, B., & Suganthi, L. (2016). Message involvement and attitude towards green advertisements. *Marketing Intelligence and Planning*, *34*, 863–882.

Kurth, C. (2015). Moral anxiety and moral agency. In M. Timmons (Ed.), *Oxford studies in normative ethics* (Vol. 5, pp. 171–195). Oxford University Press.

Kurth, C. (2018a). Emotion, deliberation, and the skill model of virtuous agency. *Mind & Language*, *33*, 299–317.

Kurth, C. (2018b). *The anxious mind*. MIT Press.

MacKuen, M., Wolak, J., Keele, L., & Marcus, G. (2010). Civil engagements. *American Journal of Political Science*, *54*, 440–458.

Schwartz, B., & Metcalfe, J. (2014). Tip-of-the-tongue (TOT) states: Mechanisms and metacognitive control. In B. Schwartz & A. Brown (Eds.), *Tip-of-the-tongue states and related phenomena* (Vol. 39, pp. 737–749). Cambridge University Press.

Thomson, V. (2009). Dual-process theories: A metacognitive perspective. In J. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 171–195). Oxford University Press.

Valentino, N., Hutchings, V., Banks, A., & Davis, A. (2008). Is a worried citizen a good citizen? *Political Psychology*, *29*, 247–273.

Correction, uncertainty, and anchoring effects

Chang-Yuan Lee[a] and Carey K. Morewedge[b]

[a]Rotman School of Management, University of Toronto, Toronto, ON, Canada and [b]Questrom School of Business, Boston University, Boston, MA, USA
changyuan.lee@utoronto.ca; morewedg@bu.edu
https://www.bu.edu/questrom/profile/carey-morewedge/

doi:10.1017/S0140525X2200303X, e129

Abstract

We compare the predictions of two important proposals made by De Neys to findings in the anchoring effect literature. Evidence for an anchoring-and-adjustment heuristic supports his proposal that system 1 and system 2 are non-exclusive. The relationship between psychophysical noise and anchoring effects, however, challenges his proposal that epistemic uncertainty determines the involvement of system 2 corrective processes in judgment.

As a case study, we compare two of De Neys's important proposals to findings from the literature on anchoring effects. Evidence for an anchoring-and-adjustment heuristic supports De Neys's proposal that systems 1 and 2 are non-exclusive (Epley & Gilovich, 2001; Simmons, LeBoeuf, & Nelson, 2010; Tversky & Kahneman, 1974). The increase in anchoring effects with quantifiable measures of uncertainty (Lee & Morewedge, 2022), however, challenges his proposal that uncertainty monitoring drives the involvement of system 2 correction processes in judgment.

As background, an anchoring effect occurs when considering an initial value (i.e., an anchor) biases subsequent estimates of a stimulus (i.e., the target). Inaccurate estimates of the target are more likely to fall between the anchor and the correct answer than beyond the correct answer (Tversky & Kahneman, 1974). When people are asked to estimate the duration of Mars' orbit, for instance, the number that typically first comes to mind is the duration of Earth's orbit (i.e., 365 days). This anchor influences estimates of Mars' orbit. People are more likely to underestimate the duration of Mars' orbit than to overestimate its duration (Epley & Gilovich, 2001).

An influential anchoring-and-adjustment heuristic theory of anchoring effects suggests that people make estimates by correcting from intuitive (but wrong) anchors until they reach the first plausible value of the target of their estimate. As the first value in the range of plausible values is usually incorrect, adjustment from anchors tend to be insufficient (Epley & Gilovich, 2001; Simmons et al., 2010; Tversky & Kahneman, 1974). Two forms of evidence from tests of the anchoring-and-adjustment heuristic support De Neys's proposal that systems 1 and 2 are non-exclusive. First, few participants in anchoring studies give anchors as their final responses (e.g., <2.8% in study 1B, Simmons et al., 2010). Even participants constrained by cognitive load or intoxication, for instance, would be unlikely to guess the duration of Mars' orbit to be the same as Earth's orbit. They would guess the duration of Mars' orbit to be closer to 365 days than thinkers who are unconstrained (Epley & Gilovich, 2001, 2006). Second, motor movements associated with the rejection and acceptance of answers influence the degree to which people correct from anchors (e.g., head nodding and shaking; Epley & Gilovich, 2001). These results suggest system 2 correction occurs even under constraint and system 1 can influence when system 2 correction ends.

Anchoring paradigms are also useful for examining De Neys's uncertainty monitoring proposal. Anchoring is a bias where the degree of uncertainty within the judge can be quantified. Uncertainty can be expressed as the width of the plausible range of values of the target of judgment (Jacowitz & Kahneman, 1995); the distance between the lowest and highest plausible value. This range varies with factors like the expertise of the judge (Smith, Windschitl, & Bruchmann, 2013; Wilson, Houston, Etling, & Brekke, 1996) and with correlates of uncertainty like psychophysical noise. Because of the increasing psychophysical noise associated with numbers as they increase in magnitude (Feigenson, Dehaene, & Spelke, 2004), the plausible range (uncertainty) of values for estimates of larger numbers is wider than for smaller numbers (Lee & Morewedge, 2022; Quattrone et al., 1984). People estimate the calories in a small serving of McDonald's French fries be anywhere between 141.48 and 223.46, whereas they estimate the calories in a large serving to be anywhere between 266.98 and 423.36. In other words, they perceive the plausible range of calories in a large serving of French fries to be wider than the plausible range of calories in a small serving of French fries (widths of 156.38 and 81.98 calories, respectively). The same pattern holds for novel unfamiliar stimuli like smaller and larger dot-arrays.

Challenging De Neys's proposal that uncertainty monitoring determines the activation and engagement of system 2 adjustment processes, epistemic uncertainty does not increase the probability or extremity of system 2 correction from anchors (i.e., system 1 intuitions). De Neys's proposed mechanism for the intervention of system 2, uncertainty monitoring, implies that correction from anchors should be more likely and more extreme when uncertainty about the value of the target is greater: When the plausible range of values for a target stimulus is wider. However, research participants do not exhibit more correction from anchors when plausible ranges of a target are wider. Correction from anchors is as likely and proportionally similar for targets with wider and narrower plausible ranges. When estimating the number of dots in 35-dot and 273-dot arrays, for instance, people first exposed to a low anchor (a 10-dot array) tend to give answers similar in relative distance from the lowest plausible values of the target. When one examines the raw size of anchoring effects (i.e., absolute values), people exhibit larger anchoring effects when the plausible range of stimulus values are wider than narrower (i.e., when uncertainty is greater). This pattern holds whether people are estimating the number of dots in larger or smaller dot-arrays, the prices of larger and smaller servings of donuts, the weight of larger or smaller dog breeds, the prices of higher or lower rated hotels, or the number of calories in larger and smaller servings of McDonald's French fries (Lee & Morewedge, 2022). These findings suggest epistemic uncertainty bounds the extent to which anchors influence judgment. It does not determine the extent of system 2 adjustment from anchors.

More generally, our comparison illustrates the value of anchoring paradigms for tackling the exciting questions De Neys raises about the relationship between intuitive and corrective mental processes.

Acknowledgments. We thank Nicholas Epley, Dilip Soman, and Ryan Webb for helpful feedback.

Competing interest. None.

References

Epley, N., & Gilovich, T. (2001). Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors. *Psychological Science*, *12*(5), 391–396.

Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological Science*, *17*(4), 311–318.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314.

Jacowitz, K. E., & Kahneman, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin*, *21*(11), 1161–1166.

Lee, C. Y., & Morewedge, C. K. (2022). Noise increases anchoring effects. *Psychological Science*, *33*(1), 60–75.

Quattrone, G. A., Lawrence, C. P., Finkel, S. E., & Andrus, D. C. (1984). Explorations in anchoring: The effects of prior range, anchor extremity, and suggestive hints (unpublished manuscript). Stanford University.

Simmons, J. P., LeBoeuf, R. A., & Nelson, L. D. (2010). The effect of accuracy motivation on anchoring and adjustment: Do people adjust from provided anchors?. *Journal of Personality and Social Psychology*, *99*(6), 917–932.

Smith, A. R., Windschitl, P. D., & Bruchmann, K. (2013). Knowledge matters: Anchoring effects are moderated by knowledge level. *European Journal of Social Psychology*, *43*(1), 97–108.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, *185* (4157), 1124–1131.

Wilson, T. D., Houston, C. E., Etling, K. M., & Brekke, N. (1996). A new look at anchoring effects: Basic anchoring and its antecedents. *Journal of Experimental Psychology: General*, *125*(4), 387–402.

A view from mindreading on fast-and-slow thinking

Jason Low[a], Stephen A. Butterfill[b] and John Michael[c]

[a]School of Psychology, Victoria University of Wellington, Wellington, New Zealand; [b]Department of Philosophy, University of Warwick, Coventry, UK and [c]Department of Philosophy, University of Milan, Milano, Italy

Jason.Low@vuw.ac.nz
s.butterfill@warwick.ac.uk
john.michael@unimi.it
https://www.researchgate.net/profile/Jason-Low-4
https://www.butterfill.com/
https://www.unimi.it/en/ugov/person/john-michael

doi:10.1017/S0140525X22002898, e130

Abstract

De Neys's incisive critique of empirical and theoretical research on the exclusivity feature underscores the depth of the challenge of explaining the interplay of fast and slow processes. We argue that a closer look at research on mindreading reveals abundant evidence for the exclusivity feature – as well as methodological and theoretical perspectives that could inform research on fast and slow thinking.

De Neys opposes the "exclusivity feature," on which fast and slow processes are "exclusively tied" to particular responses. De Neys explains that "there is no solid empirical ground for the exclusivity assumption" – this is the "fundamental problem" of the target article (sect. 1.1.3, para. 3). However, with respect to empirical evidence, De Neys mentions mindreading only in passing. Will a closer look at mindreading give him reason to reconsider the exclusivity assumption?

Methodologically, the studies De Neys relies on mostly involve observing direct, explicit choices, as is typically the case in research on reasoning. In mindreading research, by contrast, the norm is to observe both indirect implicit and direct explicit behaviours generated by a single scenario. These include anticipatory looking and verbal responses (Clements & Perner, 1994), early mediolateral motor activity and purposive action (Zani, Butterfill, & Low, 2020), response times and choices (Edwards & Low, 2017), or curvature and initiation time of computer-mouse movements (van der Wel, Sebanz, & Knoblich, 2014). In Clements and Perner's seminal study, 3-year-olds correctly looked in anticipation of the belief-based action of an agent even though they gave incorrect explicit verbal predictions about where the agent will go to search for the object. The case for accepting that certain eye movements can index a fast mindreading process that is largely unchanging over development is strengthened by evidence that anticipatory looking in infants (Meristo et al., 2012) and younger and older adults (Grainger, Henry, Naughtin, Comino, & Dux, 2018) show a similar pattern. Slow mindreading as indexed by verbal deliberations is scaffolded by culture, language, and building of schemas and causal representations (Christensen & Michael, 2016), and exhibits distinctive developmental trajectories.

None of this directly undermines De Neys's critique of the exclusivity feature. But a fruitful strand of developmental research relies on the method of signature limits (Carey, 2009). A signature limit of a process is a pattern of responses that the process generates which are incorrect or suboptimal (hence "limit") and which no other process under consideration would generate (hence "signature"). Butterfill and Apperly (2013) argued on theoretical grounds that some fast processes for tracking others' mental states are likely to generate incorrect predictions about beliefs involving mistakes about numerical identity. And in support of this, Low and Watts (2013) found that although 3-year-olds, 4-year-olds, and adults show correct looking behaviour in an object-location false-belief task, the same participants showed incorrect looking behaviour in an object-identity false-belief task. The switch from processing a location false-belief task to a numerical-identity false-belief task did not influence the usual age-related improvements in participants' explicit verbal judgements, as predicted. This is not just a hint that there is more than one process: Seeing the same signature limit in adults as in infants (Edwards & Low, 2019; Fizke, Butterfill, van de Loo, Reindl, & Rakoczy, 2017; Woo & Spelke, 2021), we infer that the fast process (and the conditions in which it occurs and the outputs it generates) does not completely overlap with the slow process (though not everyone would agree; Thompson, 2014). You cannot reject the exclusivity feature and use the method of signature limits. The view from mindreading therefore indicates that the exclusivity assumption is solidly grounded after all.

Given that the empirical basis for rejecting the exclusivity assumption is tenuous – at least in the context of mindreading research – it is important to evaluate the theoretical considerations offered by De Neys. He argues that, given the plausibility of automatization, any conclusion arrived at by a slow process could, in principle at least, also be arrived at by a fast process. However, this theoretical argument is less challenging than it first appears. Automatization tells us that any conclusion arrived at by a slow process could be arrived at by *some* fast process but not *which* fast processes could arrive at that conclusion.

Here we face a problem. A model of the interplay of fast and slow processes is needed, as De Neys argues. But De Neys's own elegant model is unavailable because it "forces us to get rid of exclusivity" (target article, sect. 2.2, para. 1). Further, developmental evidence speaks against it. On De Neys's model, the slow process should only be triggered if fast processes generate conflicting responses, leading to uncertainty. But consider children's responses to a mindreading context set up by Ruffman, Garnham, Import, and Connolly (2001). The children watched Ed acquire a false belief. They were then invited to place bets on which of two slides Ed would come down. Their bets revealed they felt no uncertainty (younger children went all in on the wrong slide). But Ruffman et al. also measured children's anticipatory looking as Ed was about to emerge, and this measure indicated a correct prediction. We take the betting to index a slow process and the looking to index a fast process. In this case we seem to have neither conflict among fast processes nor uncertainty (although of course we cannot entirely rule this out).

Is there an alternative to De Ney's model? The key is to understand what other than conflict in fast processes might trigger (and halt) slow processes. One candidate is low cognitive fluency. In Ruffman et al.'s (2001) study, asking children to choose in which of two locations to place their bets interrupts their processing and so triggers deliberation; as they reason through the problem (Ed will go where his chocolate is), they regain cognitive fluency. Because this does not require that slow processes concerning a question are driven by fast processes generating responses to the same question, this proposal leaves room for discretion whereby individuals are free to make explicit judgements which conflict with implicit responses. Just as the developmental evidence indicates.

In sum, widening De Neys's view to consider mindreading highlights the potential of more diverse methods than commonly employed in research on reasoning, and points towards empirical and theoretical obstacles to the proposed advance. Taking a step back, though, we find ourselves on common ground with De Neys: His critique shows both that more evidence is needed and that the interplay of fast and slow processes is truly a deep problem.

Competing interest. None.

References

Butterfill, S. A., & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind and Language*, *28*(5), 606–637. https://doi.org/10.1111/mila.12036

Carey, S. (2009). *The origin of concepts*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195367638.001.0001

Christensen, W., & Michael, J. (2016). From two systems to a multi-systems architecture for mindreading. *New Ideas in Psychology*, *40*(A), 48–64. https://doi.org/10.1016/j.newideapsych.2015.01.003

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, *9*(4), 377–395. https://psycnet.apa.org/doi/10.1016/0885-2014(94)90012-4

Edwards, K., & Low, J. (2017). Reaction time profiles of adults' action prediction reveal two mindreading systems. *Cognition*, *160*, 1–16. https://doi.org/10.1016/j.cognition.2016.12.004

Edwards, K., & Low, J. (2019). Level 2 perspective-taking distinguishes automatic and non-automatic belief-tracking. *Cognition*, *193*, 104017. https://doi.org/10.1016/j.cognition.2019.104017

Fizke, E., Butterfill, S. A., van de Loo, L., Reindl, E., & Rakoczy, H. (2017). Are there signature limits in early theory of mind? *Journal of Experimental Child Psychology*, *162*, 209–224. https://doi.org/10.1016/j.jecp.2017.05.005

Grainger, S. A., Henry, J. D., Naughtin, C. K., Comino, M. S., & Dux, P. E. (2018). Implicit false belief tracking is preserved in late adulthood. *Quarterly Journal of Experimental Psychology*, *71*(9), 1980–1987. https://doi.org/10.1177/1747021817734690

Low, J., & Watts, J. (2013). Attributing false beliefs about object identity reveals a signature blind spot in humans' efficient mind-reading system. *Psychological Science*, *24*(3), 305–311. https://doi.org/10.1177/0956797612451469

Meristo, M., Morgan, G., Geraci, A., Lozzi, L., Hjelmquist, E., Surian, L., & Siegal, M. (2012). Belief attribution in deaf and hearing infants. *Developmental Science*, *15*(5), 633–640. https://doi.org/10.1111/j.1467-7687.2012.01155.x

Ruffman, T., Garnham, W., Import, A., & Connolly, D. (2001). Does eye gaze indicate knowledge of false belief: Charting transitions in knowledge. *Journal of Experimental Child Psychology*, *80*(3), 201–224. https://doi.org/10.1006/jecp.2001.2633

Thompson, J. R. (2014). Signature limits in mindreading systems. *Cognitive Science*, *38*(7), 1432–1455. https://doi.org/10.1111/cogs.12117

van der Wel, R. P., Sebanz, N., & Knoblich, G. (2014). Do people automatically track others' beliefs? Evidence from a continuous measure. *Cognition*, *130*(1), 128–133. https://doi.org/10.1016/j.cognition.2013.10.004

Woo, B., & Spelke, E. (2021). Limits to early mental state reasoning: Fourteen-to 15-month-old infants appreciate whether others can see objects, but not others' experiences of objects. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*, 1914–1920. https://escholarship.org/uc/item/42n9x4n3

Zani, G., Butterfill, S. A., & Low, J. (2020). Mindreading in the balance: Adults' mediolateral leaning and anticipatory looking foretell others' action preparation in a false-belief interactive task. *Royal Society Open Science*, *7*(1), 191167. https://doi.org/10.1098/rsos.191167

Automatic threat processing shows evidence of exclusivity

David S. March[a], Michael A. Olson[b] and Lowell Gaertner[b]

[a]Department of Psychology, Florida State University, Tallahassee, FL, USA and [b]Department of Psychology, University of Tennessee, Knoxville, TN, USA
march@psy.fsu.edu
olson@utk.edu
gaertner@utk.edu
https://psy.fsu.edu/faculty/marchd/march.dp.php
https://psychology.utk.edu/faculty/gaertner.php

doi:10.1017/S0140525X22002928, e131

Abstract

De Neys argues against assigning exclusive capacities to automatic versus controlled processes. The dual implicit process model provides a theoretical rationale for the exclusivity of automatic threat processing, and corresponding data provide empirical evidence of such exclusivity. De Neys's dismissal of exclusivity is premature and based on a limited sampling of psychological research.

De Neys argues that assigning exclusive capacities to automatic (i.e., intuitive, system 1) versus controlled (i.e., deliberate, system 2) processes is unsupportable in current dual-process frameworks and unsupported by evidence. Dismissing such exclusivity, however, is premature and based on a limited sampling of psychological research. In particular, the dual implicit process model (DIPM; March, Gaertner, & Olson, 2018a, 2018b) details how automatic threat processing is fundamentally distinct from automatic valence processing and deliberate processing. According to the DIPM, a neural architecture that facilitates survival evolved to preferentially process immediate survival threats relative to other negatively and positively valenced stimuli. Such preferential processing manifests as faster and stronger perceptual, physiological, and behavioral reactions to physically threatening stimuli. Because of the necessarily fast time course of those reactions, their functional utility could not be supported by deliberate (system 2) processing.

March, Gaertner, and Olson (2017) provided initial evidence of the exclusivity of automatic threat processing based on reactions to four categories of stimulus images: threatening (e.g., snarling predators, gunmen), negative (e.g., feces, wounded animals), positive (e.g., puppies, babies), and neutral (e.g., doorknobs, cups). Three studies presented those stimuli in visual search, eye-tracking, and startle-eyeblink paradigms. Consistent with the exclusivity of automatic threat processing, threatening stimuli (relative to the other stimuli) were detected faster, more frequent targets of initial eye-gaze, and elicited stronger startle-eyeblinks (with responses occurring between 200 and 1,000 ms). March, Gaertner, and Olson (2022) provided even stronger evidence of exclusivity by suboptimally presenting those stimuli below conscious perception at 15–21 ms in three additional studies. Despite participants being unable to describe what was presented (based on two pilot studies), threatening stimuli (relative to the other stimuli) elicited stronger skin-conductance and startle-eyeblinks and more negative downstream evaluations. Automatic threat processing (but neither automatic valence processing nor deliberate processing) evoked functional responses to stimuli below conscious perception. It would be a strange argument indeed to suggest that participants deliberately reasoned skin-conductance and startle-eyeblink to vary uniquely with images of survival threats that they were unable to describe.

The DIPM provides a theoretical rationale for the exclusivity of automatic threat processing and is empirically supported by evidence of such exclusivity. The DIPM, however, is just one example and there are others. In the arena of implicit social cognition, research indicates that automatic processes can commence immediately upon perception of a relevant object, render decisional and behavioral outputs within milliseconds, and return to baseline within a second or so, well before one might wager a guess about the price of a ball (Bargh & Ferguson, 2000; Fazio, 2007). In evaluative priming studies, a prime presented for 150 ms can facilitate categorization of a valence-congruent target, but its spreading activation effect dissipates within a second (Hermans, De Houwer, & Eelen, 2001). At least in this context, system 1 culminates well before any deliberative decision making can occur, which might offer some insight into the "unequivocal threshold" problem posed by De Neys (target article). In contrast, to even understand the problem posed to a participant in a ratio bias task or a cognitive reflective task (CRT) problem takes several seconds. By then, system 2 is likely to be already up and running. Thus, the decision processes involved in the sorts of tasks De Neys focuses on are likely to miss the very early effects of system 1. By broadening the scope of dual-process models and research paradigms considered, De Neys would have realized that exclusivity is theoretically and empirically supported.

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Competing interest. None.

References

Bargh, J. A., & Ferguson, M. J. (2000). Beyond behaviorism: On the automaticity of higher mental processes. *Psychological Bulletin*, *126*, 925–945.

Fazio, R. H. (2007). Attitudes as object–evaluation associations of varying strength. *Social Cognition*, *25*, 603–637.

Hermans, D., De Houwer, J., & Eelen, P. (2001). A time course analysis of the affective priming effect. *Cognition & Emotion*, *15*, 143–165.

March, D. S., Gaertner, L., & Olson, M. A. (2017). In Harm's way: On preferential response to threatening stimuli. *Personality and Social Psychology Bulletin*, *43*, 1519–1529.

March, D. S., Gaertner, L., & Olson, M. A. (2018a). On the prioritized processing of threat in a dual implicit process model of evaluation. *Psychological Inquiry*, *29*, 1–13.

March, D. S., Gaertner, L., & Olson, M. A. (2018b). Clarifying the explanatory scope of the dual implicit process model. *Psychological Inquiry*, *29*, 38–43. doi: doi.org/10.1080/1047840X.2018.1435622

March, D. S., Gaertner, L., & Olson, M. A. (2022). On the automatic nature of threat: Physiological and evaluative reactions to survival-threats outside conscious perception. *Affective Science*, *3*, 135–144. doi: doi.org/10.1007/s42761-021-00090-6

Hoist by its own petard: The ironic and fatal flaws of dual-process theory

David E. Melnikoff[a] and John A. Bargh[b]

[a]Department of Psychology, Northeastern University, Boston, MA, USA and
[b]Department of Psychology, Yale University, New Haven, CT, USA
davidemelnikoff@gmail.com; https://www.davidmelnikoff.com/
john.bargh@yale.edu; https://acmelab.yale.edu/

doi:10.1017/S0140525X22003077, e132

Abstract

By stipulating the existence of a system 1 and a system 2, dual-process theories raise questions about how these systems function. De Neys identifies several such questions for which no plausible answers have ever been offered. What makes the nature of systems 1 and 2 so difficult to ascertain? The answer is simple: The systems do not exist.

Dual-process theories of human reasoning have yet to provide plausible answers to basic questions about the nature of system

1 and system 2 processing. Can system 1 reason logically? How do people switch from system 1 to system 2? As De Neys convincingly argues, existing answers to these questions fail under logical and empirical scrutiny.

There is an irony to this. The questions that confound dual-process theories are the very questions that these theories introduce in the first place. By positing that systems 1 and 2 exist in some meaningful sense, dual-process theories saddle themselves with the challenge of explaining how these systems (or types of processes) operate and interact. Having long failed to meet this challenge with even a single coherent hypothesis of something as basic as how system 2 is activated, it is worth asking if dual-process theories are wrong at the most fundamental level: Maybe system 1 and system 2 simply do not exist. Maybe what De Neys presents as important puzzles in need of solving are just red herrings symptomatic of a flawed theoretical foundation. We suspect that this is the case, and therefore recommend that instead of developing dual-process theories further, researchers abandon dual-process theories altogether (Melnikoff & Bargh, 2018).

What would it look like to abandon dual-process theories? Instead of asking questions about system 1 and system 2, researchers would ask questions like: How much time and effort is required to perform well on a given type of reasoning problem? Under what conditions do different components of reasoning occur spontaneously versus intentionally? How do different types of working memory load (e.g., visual vs. verbal) interact with different aspects of the reasoning process? By eschewing commitment to the existence of systems 1 and 2, questions like these avoid the obstacles that thwart dual-process theories.

Consider the central question of the target article: How do people switch between effortless ("intuitive") and effortful ("deliberative") reasoning? As De Neys shows, the answers offered by dual-process theories fail to meet even the minimum threshold of logical coherence. For instance, multiple dual-process theories claim that the activation of system 2 depends on system 2 already being activated. Such paradoxes vanish, however, when the concepts of system 1 and system 2 are abandoned. To illustrate, consider the learned value of control (LVOC) model of Lieder, Shenhav, Musslick, and Griffiths (2018). This model says that people use reinforcement learning to estimate the value of exerting different amounts of effort in particular situations. For example, people may learn from prior experience that when presented with certain reasoning problems in particular contexts, they tend to obtain better outcomes when they ignore their initial hunch and invest effort in further deliberation. This learned value can be used to decide how much effort to invest when similar situations are encountered in the future.

There are no paradoxes here. The LVOC model is a perfectly coherent account of how people modulate the effort they invest in reasoning – one of many (e.g., Abrahamse, Braem, Notebaert, & Verguts, 2016; Restle, 1962; Shenhav, Botvinick, & Cohen, 2013) – and the fact that it makes no reference to systems 1 and 2 is no coincidence. Just imagine if it did. Reframed in dual-process terms, the LVOC might posit that system 2 learns the expected value of deliberating, and then uses this information to decide when to ignore system 1 processing in favor of system 2 processing. Now we have a problem: The new theory implies that the activation of system 2 is a precondition for itself, introducing an infinite regress. An alternative reframing of the LVOC might posit that whenever system 1 processing is ignored in favor of system 2 processing, system 1 updates the value of deliberation and uses this updated value to decide when to deliberate in the future. But this is paradoxical. If system 1 processing is ignored, system 1 cannot, by definition, update the value of *anything*.

The point here is that by embracing the system 1/system 2 framework, nothing is gained but confusion. The conceptual commitments of the framework only make it harder to generate coherent answers to the very questions that dual-process theorists care about.

Of course, none of this would matter if there were evidence showing that systems 1 and 2 do in fact exist. Although fictional theoretical constructs can be discarded as soon as they prove unhelpful, actual features of the mind cannot be ignored simply because they introduce conceptual conundrums. So, are systems 1 and 2 real?

There is no reason to think so. Fundamentally, system 1/system 2 is a proposed dimension along which psychological processes vary: At one end of the spectrum, processes are fast, effortless, and spontaneous, and at the other end, processes are slow, effortful, and intentional. So, the reality of the system 1/system 2 distinction hinges on whether it is true that the features of speed, effort, and intentionality are in some sense reducible to a single, more basic dimension.

Evidence supporting this idea does not exist. There has been no attempt to establish that speed, effort, and intentionality are intercorrelated to any meaningful degree, let alone to the point that they might be reducible to a single underlying dimension (Bargh, 1994; Keren & Schul, 2009; Kruglanski & Gigerenzer, 2011; Melnikoff & Bargh, 2018; Moors & De Houwer, 2006). On the contrary, countless examples of "misalignments" between processing features have been documented, such as fast processes that are intentional (Melnikoff & Bailey, 2018) and inefficient (Gilbert & Hixon, 1991), and spontaneous processes that are effortful (Kim, Kim, & Chun, 2005). Such phenomena suggest that an attempt to validate the system 1/system 2 distinction, were it ever undertaken, is unlikely to succeed.

In short, there is no evidence that systems 1 and 2 exist in any meaningful sense, and head-scratching paradoxes vanish if we assume they do not. Therefore, we should abandon dual-process theory rather than embark on a doomed mission to save it.

Financial support. The authors have no funding sources contributing to this article.

Competing interest. None.

References

Abrahamse, E., Braem, S., Notebaert, W., & Verguts, T. (2016). Grounding cognitive control in associative learning. *Psychological Bulletin*, *142*(7), 693–728.

Bargh, J. A. (1994). The four horsemen of automaticity: Intention, awareness, efficiency, and control as separate issues. In R. S. Wyer, Jr., & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed., Vol. 1, pp. 1–40). Erlbaum.

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology*, *60*(4), 509–517.

Keren, G., & Schul, Y. (2009). Two is not always better than one: A critical evaluation of two-system theories. *Perspectives on Psychological Science*, *4*, 593–550.

Kim, S. Y., Kim, M. S., & Chun, M. M. (2005). Concurrent working memory load can reduce distraction. *Proceedings of the National Academy of Sciences*, *102*(45), 16524–16529.

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

Lieder, F., Shenhav, A., Musslick, S., & Griffiths, T. L. (2018). Rational metareasoning and the plasticity of cognitive control. *PLoS Computational Biology*, *14*(4), e1006043.

https://doi.org/10.1017/S0140525X2200142X Published online by Cambridge University Press

Melnikoff, D. E., & Bailey, A. H. (2018). Preferences for moral vs. immoral traits in others are conditional. *Proceedings of the National Academy of Sciences*, *115*(4), E592–E600.
Melnikoff, D. E., & Bargh, J. A. (2018). The mythical number two. *Trends in Cognitive Sciences*, *22*, 280–293.
Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, *132*(2), 297–326.
Restle, F. (1962). The selection of strategies in cue learning. *Psychological Review*, *69*(4), 329–343.
Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron*, *79*(2), 217–240.

Not feeling right about uncertainty monitoring

Ian R. Newman and Valerie A. Thompson

Department of Psychology and Health Studies, University of Saskatchewan, Saskatoon, SK, Canada
ian.newman@usask.ca; valerie.thompson@usask.ca

doi:10.1017/S0140525X22003089, e133

Abstract

De Neys proposed a "switch" model to address what he argued to be lacuna in dual-process theory, in which he theorized about the processes that initiate and terminate analytic thinking. We will argue that the author neglected to acknowledge the abundant literature on metacognitive functions, specifically, the meta-reasoning framework developed by Ackerman and Thompson (2017), that addresses just those questions.

The meta-reasoning framework (Ackerman & Thompson, 2017) differentiates between object-level processes that are responsible for thinking, reasoning, and deciding, and the meta-level process that monitors and controls them. Part of this theory addresses the processes that initiate and terminate analytic thinking, the so-called switch function proposed by De Neys. In the meta-reasoning model, the *feeling of rightness* (Thompson, 2009; Thompson, Prowse Turner, & Pennycook, 2011, 2013) has been proposed as one way that intuitive responses can be monitored and analytic processes initiated.

Thompson argued that system 1 processing is accompanied by a metacognitive experience, a feeling of rightness or certainty that the response(s) generated by system 1 are correct (Thompson, 2009). The relative strength of one's feeling of rightness determines the probability one will switch to and engage system 2; when the feeling is strong, the response is often accepted with little system 2 analysis. Conversely, when the feeling of rightness is weak, system 2 is often engaged. In other words, the feeling of rightness is a cue to either accept the outcome of system 1 processing or "switch" to system 2 processing. Thompson also proposed potential determinants of the strength of a feeling of rightness, such as fluency of processing (speed and ease of response generation) and, importantly, the presence of conflicting responses (Thompson & Johnson, 2014; Thompson et al., 2011).

Although De Neys briefly acknowledged that his working model could be integrated with the feeling of rightness (sect. 4.4), he neglected a broad and well-established model to propose an explanation for a narrower range of phenomena. The meta-reasoning framework (Ackerman & Thompson, 2017) is a multifaceted framework that encompasses a variety of monitoring and control processes, in addition to a "switch" mechanism. Thus, De Neys's proposal fleshed out the specific case of monitoring for conflict in the broader meta-reasoning framework. We also note the similarity between monitoring feelings of rightness and "uncertainty monitoring," where one can think about "uncertainty" as the inverse of confidence or feeling of rightness.

De Neys also argued that leading switch accounts presuppose exclusivity, where certain types of responses (e.g., a normatively correct response) are exclusively generated by one system or the other. Thompson and Newman (2020) noted that the exclusivity assumption has been long abandoned in dual-process theorizing (Evans, 2019; Evans & Stanovich, 2013; Stanovich, 2018; Thompson, 2011) and have obtained evidence in support of this in our own lab (Newman, Gibb, & Thompson, 2017). Most importantly, the feeling of rightness does not depend on the normative qualities of the response generated and, therefore, does not rely on an exclusivity assumption to be instantiated. Essentially, the claim that leading switch accounts presuppose exclusivity is not consistent with current metacognitive models.

We agree with De Neys that many of the processes that monitor our cognitions are likely "system 1" in character: They should not demand working memory resources and their origins are likely not subject to introspection (Koriat, 2007). However, we think there are several important differences between the uncertainty monitoring mechanisms proposed by De Neys and the types of monitoring mechanisms that are common in the metacognitive literature. First, although their origins may not be subject to introspection, their outputs are thought to enter awareness. We have hypothesized that feelings of rightness are felt subjectively (Thompson et al., 2011), meaning one is consciously aware, at least to some degree, that they have these feelings. Alternatively, it is not clear that uncertainty monitoring has the same properties.

This difference is subtle but crucial. From a metacognitive perspective, one can be aware of when they feel their response is completely right, completely wrong, or somewhere in between. From an uncertainty monitoring perspective, it seems that one is only aware when one feels a sufficient degree of uncertainty – either from a weakly generated response or two (or more) conflicting responses relatively similar in strength. Therefore, there is no mechanism by which an individual could be aware of when they feel extremely right or wrong, nor for how such feelings could cue the engagement of system 2. Finally, we note that although most monitoring is likely to be implicit, there is also a role for explicit beliefs about oneself, the task, the types of behaviours that constitute good reasoning, and so on (Ackerman, 2019; Jost, Kruglanski, & Nelson, 1998; Koriat, 2007). Thus, not all monitoring processes can be lumped under the "system 1" rubric.

Of most relevance, De Neys's proposed model cannot explain the phenomenon of being highly certain that something is wrong (i.e., *feeling of wrongness*), such as when a firefighter has a sudden but strong feeling that something is amiss and they must get out of the building immediately (Klein, Calderwood, & Clinton-Cirocco, 2010). Similarly, I could ask you: *Is Edinburgh the capital of Botswana?* You may not know the actual capital of Botswana (it is Gaborone), but you likely would have a remarkably strong feeling that it is *definitely not* Edinburgh. Similarly, one may have a strong *feeling of error* about a response one has given (Ackerman & Thompson, 2017; Fernandez Cruz, Arango-Muñoz, & Volz, 2016). Both dimensions of the spectrum are important because a strong feeling about a response (whether it be rightness or wrongness or error; Ackerman & Thompson, 2017) is a cue that action is necessary (Thompson, 2017).

De Neys provided a detailed analysis of how reasoners monitor for conflict. However, conflict monitoring is just one part of the role that metacognitive monitoring plays. Moreover, the proposal failed to acknowledge the more comprehensive framework developed by Ackerman and Thompson. It is ultimately an oversimplification of an existing framework that substitutes uncertainty for feeling of rightness, uncertainty monitoring for metacognitive monitoring, and an uncertainty criterion for the diminishing criterion model of confidence (Ackerman, 2014). We agree with much of the criticism and questions raised by De Neys here (both directly and indirectly) and acknowledge that they are important, but note that frameworks, such as the meta-reasoning framework of Ackerman and Thompson (2017), already exist to answer them.

Financial support. This work is supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) to Valerie A. Thompson.

Competing interest. None.

References

Ackerman, R. (2014). The diminishing criterion model for metacognitive regulation of time investment. *Journal of Experimental Psychology: General*, *143*(3), 1349–1368. https://doi.org/10.1037/a0035098

Ackerman, R. (2019). Heuristic cues for meta-reasoning judgments: Review and methodology. *Psihologijske Teme*, *28*(1), 1–20. https://doi.org/10.31820/pt.28.1.1

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

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*(4), 383–415. https://doi.org/10.1080/13546783.2019.1623071

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

Fernandez Cruz, A. L., Arango-Muñoz, S., & Volz, K. G. (2016). Oops, scratch that! Monitoring one's own errors during mental calculation. *Cognition*, *146*, 110–120. https://doi.org/10.1016/j.cognition.2015.09.00

Jost, J. T., Kruglanski, A. W., & Nelson, T. O. (1998). Social metacognition: An expansionist review. *Personality and Social Psychology Review*, *2*(2), 137–154. https://doi.org/10.1207/s15327957pspr0202_6

Klein, G., Calderwood, R., & Clinton-Cirocco, A. (2010). Rapid decision making on the fire ground: The original study plus a postscript. *Journal of Cognitive Engineering and Decision Making*, *4*(3), 186–209. https://doi.org/10.1518/155534310X12844000801203

Koriat, A. (2007). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 289–325). Cambridge University Press.

Newman, I. R., Gibb, M., & Thompson, V. A. (2017). Rule-based reasoning is fast and belief-based reasoning can be slow: Challenging current explanations of belief-bias and base-rate neglect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(7), 1154–1170. https://doi.org/10.1037/xlm0000372

Stanovich, K. E. (2018). Miserliness in human cognition: The interaction of detection, override and mindware. *Thinking & Reasoning*, *24*(4), 423–444. https://doi.org/10.1080/13546783.2018.1459314

Thompson, V. A. (2009). Dual-process theories: A metacognitive perspective. In J. S. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 171–195). Oxford University Press.

Thompson, V. A. (2011). Normativism versus mechanism. *Behavioral and Brain Sciences*, *34*(5), 272–273. https://doi.org/10.1017/S0140525X11000574

Thompson, V. A. (2017). Certainty and action. In N. Galbraith, E. Lucas, & D. E. Over (Eds.), *The thinking mind: A festschrift for Ken Manktelow* (pp. 66–82). Routledge.

Thompson, V. A., & Johnson, S. C. (2014). Conflict, metacognition, and analytic thinking. *Thinking & Reasoning*, *20*(2), 215–244. https://doi.org/10.1080/13546783.2013.869763

Thompson, V. A., & Newman, I. R. (2020). Working memory, autonomy, and dual process theories: A roadmap. In S. Elqayam, I. Douven, J. S. B. Evans, & N. Cruz (Eds.), *Logic and uncertainty in the human mind: A tribute to David E. Over* (pp. 227–242). Routledge.

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, V. A., Prowse Turner, J. A. P., Pennycook, G., Ball, L. J., Brack, H., Ophir, Y., & Ackerman, R. (2013). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition*, *128*(2), 237–251. https://doi.org/10.1016/j.cognition.2012.09.012

Could Bayesian cognitive science undermine dual-process theories of reasoning?

Mike Oaksford

Department of Psychological Sciences, Birkbeck College, University of London, London, UK
mike.oaksford@bbk.ac.uk
https://www.bbk.ac.uk/our-staff/profile/8009448/mike-oaksford

doi:10.1017/S0140525X22003132, e134

Abstract

Computational-level models proposed in recent Bayesian cognitive science predict both the "biased" and correct responses on many tasks. So, rather than possessing two reasoning systems, people can generate both possible responses within a single system. Consequently, although an account of why people make one response rather than another is required, dual processes of reasoning may not be.

Wim De Neys makes a compelling case that recent evidence showing that system 1 can make both incorrect or biased and correct responses raises problems for the switching mechanism that moves between system 1 and system 2. In this commentary, I argue that recent work in the new paradigm in human reasoning (Oaksford & Chater, 2020) or Bayesian cognitive science (Chater & Oaksford, 2008), more generally, shows that the so-called biased response can be correct, given the right background beliefs or in the right environment. Consequently, rather than requiring two reasoning systems, the evidence Wim cites may instead suggest that people consider more than one possible correct response.

Is it surprising that system 1 can compute the correct response? Other animals, which likely can only possess a putative system 1, are capable of rational decision making (Monteiro, Vasconcelos, & Kacelnik, 2013; Oaksford & Hall, 2016; Stanovich, 2013). Moreover, the unconscious inferences underpinning perception and action are widely believed to be the product of the same rational Bayesian inferences (Clark, 2013; Friston, 2010) that underpin new paradigm approaches to human verbal reasoning (Oaksford & Chater, 1994, 2012, 2020; Oaksford & Hall, 2016). Within a single model (reasoning system?), these approaches can predict both the "biased" and correct responses. For example, optimal data selection predicts not only so-called confirmation bias in Wason's selection task, but also, depending on the model's parameters, the reflective, falsification response (Oaksford & Chater, 1994; see also, Coenen, Nelson, & Gureckis, 2019). These different possibilities can be unconsciously simulated by varying these parameters.

The possibility that becomes the focus of attention in working memory (WM), and hence which response is made first, will depend on which is best supported by environmental cues or prior knowledge.

This pattern, whereby both the "biased" and correct response can arise from the same computational-level model of the reasoning process, is common across Bayesian cognitive science. A further example is Oaksford and Hall's (2016) model of the base-rate neglect task, on which Wim comments approvingly. This model is related to models of categorisation, where categories are causally related to their features (cues) (Rehder, 2017). Both responses arise from sampling a posterior distribution when the base rates of being female in a sample are updated by the cues to femininity in the description of a person randomly drawn from that sample. Whether the prior (respond male) is washed out (respond female) depends on the perceived strength of the cues in the description of the person sampled. So, both responses can be considered correct depending on other background knowledge.

Further examples abound. In deductive reasoning, a similar variation in endorsing conditional inferences is predicted by the same probabilistic factors as in data selection (Oaksford & Chater, 2007, Fig. 5.5; Vance & Oaksford, 2021). In computational-level theories of the conjunction fallacy (Tentori, Crupi, & Russo, 2013) and argumentation (Hahn & Oaksford, 2007), responses may be based on the probability of the conclusion (Pr(C)) or the Bayesian confirmation-theoretic relation between premises and conclusion (e.g., $\Pr(C|P) - \Pr(C)$, the likelihood ratio, etc.). These can lead to conflicting possible responses regarding the strength of the argument and to endorsing the conjunction fallacy. In argumentation, the same Bayesian model explains when an informal argument fallacy, for example, ad hominem or circular reasoning, is fallacious and when it is not. "Biased" responses may also arise from how the brain estimates probabilities by sampling (e.g., Dasgupta, Schulz, & Gershman, 2017; Zhu, Sanborn, & Chater, 2020). Small samples may be combined with priors to produce initially "biased" responses that move towards the correct response, given more sampling time.

For many of these tasks, it is doubtful that people can explicitly calculate the appropriate responses without formal training and pencil and paper. Although, given more time or a second chance to respond, they may produce the alternative possibilities produced by their unitary reasoning system. Even for tasks where explicit computation is possible, like the bat and ball task, the "biased" response is a necessary step in computing the reflective response. This task involves solving the simultaneous equations (a) $x + y = \$1.10$, (b) $x - y = \$1$, for y (cost of the ball). The first step involves taking (b) from (a): $y - y = \$1.10 - \$1$. The next step requires an understanding that $y - -y = 2y$, which is beyond many UG psychology students in the UK. Yet they may realise that the difference, $\$1.10 - \$1$, is on the way to the solution. Getting this far may also lead to their maths tutor awarding them more than 50% of the marks in a classroom test. But giving this answer may leave a feeling of unrightness because the process was not completed. Given a second chance, people respond with a figure less than 10 cents, indicating an understanding that $y - -y$ is greater than y (Bago, Raoelison, & De Neys, 2019). So, the intuitive response arises as part of computing the correct solution suggesting that heuristics are unnecessary. The two possible responses emerge from the same rational cognitive process.

In summary, so-called biases may often be a function of the same processes that lead to the reflective, rational response (see also, Kruglanski & Gigerenzer, 2011). The response depends on how prior knowledge or cues in the task materials set the parameters of the computational models. Because the environment can change, and cues are not always present, people may unconsciously simulate more than one possibility. That people do so and record the results may be the core insight of mental models theory (Oaksford, 2022). The bat and ball task shows that algebraic tasks, usually requiring pencil and paper, can be automatised and at least partially solved unconsciously. Dual-process theories and the new paradigm in reasoning were once in lockstep (Elqayam & Over, 2013). However, the specific computational-level theories developed within the new paradigm that predicts both the "biased" and the correct response on many tasks may be better interpreted as undermining the basis for this distinction on which dual-process theory depends.

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Competing interest. None.

References

Bago, B., Raoelison, M., & De Neys, W. (2019). Second-guess: Testing the specificity of error detection in the bat-and-ball problem. *Acta Psychologica*, *193*, 214–228. https://doi.org/10.1016/j.actpsy.2019.01.008

Chater, N., & Oaksford, M. (Eds.) (2008). *The probabilistic mind: Prospects for Bayesian cognitive science*. Oxford University Press.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral & Brain Sciences*, *36*, 181–253.

Coenen, A., Nelson, J. D., & Gureckis, T. M. (2019). Asking the right questions about the psychology of human inquiry: Nine open challenges. *Psychonomic Bulletin & Review*, *26*, 1548–1587. https://doi-org.ezproxy.lib.bbk.ac.uk/10.3758/s13423-018-1470-5

Dasgupta, I., Schulz, E., & Gershman, S. J. (2017). Where do hypotheses come from? *Cognitive Psychology*, *96*, 1–25.

Elqayam, S., & Over, D. E. (2013). New paradigm psychology of reasoning: An introduction to the special issue edited by Elqayam, Bonnefon, and over. *Thinking & Reasoning*, *19*, 249–265. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1080/13546783.2013.841591

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*, 127–138. https://doi.org/10.1038/nrn2787

Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A Bayesian approach to reasoning fallacies. *Psychological Review*, *114*, 704–732.

Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberate judgments are based on common principles. *Psychological Review*, *118*, 97–109. doi-org.ezproxy.lib.bbk.ac.uk/10.1037/a0020762

Monteiro, T., Vasconcelos, M., & Kacelnik, A. (2013). Starlings uphold principles of economic rationality for delay and probability of reward. *Proceedings of the Royal Society B*, *280*, 20122386. doi:10.1098/rspb.2012.2386

Oaksford, M. (2022). Mental models, computational explanation, and Bayesian cognitive science: Commentary on Knauff and Gazzo Castañeda (2022). *Thinking & Reasoning*. doi.org/10.1080/13546783.2021.2022531

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, *101*, 608–631.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press.

Oaksford, M., & Chater, N. (2012). Dual processes, probabilities, and cognitive architecture. *Mind & Society*, *11*, 15–26.

Oaksford, M., & Chater, N. (2020). New paradigms in the psychology of reasoning. *Annual Review of Psychology*, *71*, 305–330.

Oaksford, M., & Hall, S. (2016). On the source of human irrationality. *Trends in Cognitive Science*, *20*, 336–344.

Rehder, B. (2017). Concepts as causal models: Categorization. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 347–375). Oxford University Press.

Stanovich, K. E. (2013). Why humans are (sometimes) less rational than other animals: Cognitive complexity and the axioms of rational choice. *Thinking & Reasoning*, *19*, 1–26.

Tentori, K., Crupi, V., & Russo, S. (2013). On the determinants of the conjunction fallacy: Probability versus inductive confirmation. *Journal of Experimental Psychology: General*, *142*, 235–255.

Vance, J., & Oaksford, M. (2021). Explaining the implicit negations effect in conditional inference: Experience, probabilities, and contrast sets. *Journal of Experimental Psychology: General*, *150*, 354–384. https://doi.org/10.1037/xge0000954

Zhu, J.-Q., Sanborn, A. N., & Chater, N. (2020). The Bayesian sampler: Generic Bayesian inference causes incoherence in human probability judgments. *Psychological Review*, *127*, 719–748. https://doi.org/10.1037/rev0000190

Using the study of reasoning to address the age of unreason

Magda Osman

Centre for Science and Policy, University of Cambridge, Cambridge, Cambridgeshire, UK
m.osman@jbs.ac.uk
https://magdaosman.com

doi:10.1017/S0140525X22002953, e135

Abstract

If we accept that societally, politically, and even culturally enlightenment face some serious challenges, can we use this rethinking of theories of reasoning to address them? The aim here is to make a case for building on the work presented by De Neys as an opportunity to advance an applied reasoning research programme.

Since my critical review in 2004 (Osman, 2004), and valuable critiques of others (Keren & Schul, 2009; Melnikoff & Bargh, 2018), the same question keeps getting asked, can we be sure that there are two qualitatively distinct reasoning processes? De Neys's recent answer to this is no, and because of this, De Neys shows how to handle the additional conceptual difficulty in explaining switching between the two processes.

De Neys's way out is to characterise the basics in an agnostic way that anyone other than a dual-process purist, be they a single-system advocate, Bayesian, or other, might be happy. One key feature of his work is that the regulation of effort spent evaluating representations and inferences depends largely on internal (e.g., uncertainty, confidence) as well as external pressures (e.g., social interactions) to justify one's reasoning (De Neys, 2020). Dynamic-value-effort-based decision-making models have made similar proposals to explain moral behaviour (e.g., Osman & Wiegmann, 2017).

Where do we go from here?

What De Neys is proposing is as a new theoretical apparatus that diplomatically handles old internal factions. Can we use this as an opportunity to also rethink the study of reasoning on two other grounds: (1) What we do about normative standards? (2) How to promote the applied science of reasoning?

A feature unique to both reasoning and decision making is that they have at their disposal ways of benchmarking thought against normative standards, both a blessing and a curse. The research paradigms informed by how we ought to structure our thinking, and train us to do so better, is the success story. But, at the same time, we haven't gotten past the fact that we may be unfairly deferring to impossible benchmarks to assess the quality and success of an inferential process.

Maybe progress can be made if there is a more concerted interdisciplinary ambition like the one 100 years ago. In the 1890s the metaphysics club (for details see Kuklick, 2001) formed by Charles Sanders Peirce, William James, and John Dewey combined the interests of philosophy, mathematics, psychology, and linguistics. In their unified conception of language and thought viable inferences from impractical ones are sorted based on their communicative pragmatic value socially, politically, and culturally, as well as internal coherence. Just as De Neys's alludes to, deliberation as we come to understand it in the current study of reasoning, is not merely epiphenomenal. Its function is to take us beyond a first-pass inference to a defensible explanation that is persuasive to oneself and others; a position argued by others (Mercier, 2021). The reasoning field is already integrating insights from the psychology of persuasion, causal cognition, and linguistic pragmatism, but the next leap is to use this to agree on the normative approach to benchmark thought. We have the ingredients, but we need to agree on how to mix them.

Why is all this important?

As a field, we can capitalise on the popularised public face of reasoning, understood to be both fast and slow. But to do so, we might need to dedicate efforts to promoting the applied science of reasoning. Why? Because there is a sense that we are at a point in our history where enlightenment is taking a bruising. Equivalences are drawn between facts and feelings. The study of reasoning is crucial to addressing this, and other worrying patterns that emerging. The study of reasoning informs our understanding of how we develop sound arguments, how we identify sound arguments from bad, and how we reason from evidence. This is not only of scientific value, this is a given, the field is of value because the insights are essential in their applications to helping improve education, medicine, law, forensics, journalism, and public policy, to name but a few.

References

De Neys, W. (2020). Rational rationalization and system 2. *Behavioral and Brain Sciences*, *43*, e34.

Keren, G., & Schul, Y. (2009). Two is not always better than one: A critical evaluation of two-system theories. *Perspectives on Psychological Science*, *4*(6), 533–550.

Kuklick, B. (2001). Review of the metaphysical club: A story of ideas in America by Louis Menand. *Transactions of the Charles S. Peirce Society*, *37*(4), 635–638.

Melnikoff, D. E., & Bargh, J. A. (2018). The mythical number two. *Trends in Cognitive Sciences*, *22*(4), 280–293.

Mercier, H. (2021). How good are we at evaluating communicated information?. *Royal Institute of Philosophy Supplements*, *89*, 257–272.

Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin & Review*, *11*(6), 988–1010.

Osman, M., & Wiegmann, A. (2017). Explaining moral behavior: A minimal moral model. *Experimental Psychology*, *64*(2), 68–81.

Switching: Cultural fluency sustains and cultural disfluency disrupts thinking fast

Daphna Oyserman

Psychology, University of Southern California, Los Angeles, CA, USA
oyserman@usc.edu
https://dornsife.usc.edu/daphna-oyserman/

doi:10.1017/S0140525X22003119, e136

Abstract

Culture-as-situated cognition theory provides insight into the system 1 monitoring algorithm. Culture provides people with an organizing framework, facilitating predictions, focusing attention, and providing experiential signals of certainty and uncertainty as system 1 inputs. When culture-based signals convey that something is amiss, system 2 reasoning is triggered and engaged when resources allow; otherwise, system 1 reasoning dominates.

People can reason fast or slow. Slow reasoning requires more attentional resources and fast reasoning supports rapid movement through non-problematic sequences. De Neys's fast-and-slow systems perspective highlights a problem – the monitoring system that triggers switching between fast and slow thinking can only function if it resides within the fast system. Hence switching must depend on a rapid, non-resource intensive mechanism which can somehow detect when the slow system is or is not needed. De Neys focuses on the possibility that the mechanism entails a certainty algorithm that continuously compares certainty regarding fast system ideas to some certainty threshold. Metaphorically, De Neys predicts an internal thermostat that is turned on and off by a certainty threshold. It turns on when possible responses are below threshold certainty and off again when relative certainty surpasses threshold. Monitoring focuses on the quality of the proposed responses.

If these certainty inputs are internal as De Neys proposes, then looking at wedding photographs or reading an obituary should not affect cognitive reflective task (CRT) scores (a classic system 2 task), but it does (Mourey, Lam, & Oyserman, 2015). The researchers had participants view wedding photographs with the ostensible task of rating their quality and attractiveness, then had them respond to the CRT. Half of participants were randomized to a culturally fluent condition in which the photographs included a groom in a black suit, a bride in a white gown, a white fondant-iced tiered wedding cake, and a wedding party. The other half of participants were randomized to a culturally disfluent condition in which photos included a bride and groom and a cake, but the cake was decorated with cogs, and the clothing included purple and green. Participants who saw the culturally disfluent version scored higher on the CRT. The shift to slow, system 2 reasoning was triggered not by the quality of proposed responses as De Neys would propose but because the disfluent wedding photographs provided a situational signal that something is wrong. The implication is that system 2 is triggered by signals of a problematic state of affairs (outside input) and not simply by relative certainty about proposed CRT responses (internal input).

When familiar tasks are going well, vigilance is not needed and reliance on established routines and general intuitions is sufficient. When things go wrong, or tasks are unfamiliar, higher vigilance and effort are useful (Schwarz, 1990, 2001). One driver of these experiences is culture. Culture provides an organizational framework for how things will proceed, what matters, and how to make sense of experiences (Oyserman, 2011, 2017; Oyserman & Uskul, 2008/2015). People automatically use their culture-based expertise to make predictions, which typically sufficiently match what people observe that they experience a prediction-observation fit, yielding an experience of cultural fluency, a benign signal that things are as they ought to be (Lin, Arieli, & Oyserman, 2019).

From a culture-as-situated cognition perspective, inputs must come from features of the situation which themselves are cultural constructs. After all, thinking is for doing and doing is contextualized. People are not solving problems outside of contexts, they are solving them inside of contexts. Features of these contexts are of vital concern. An internally focused system that is not sensitive to contextual cues about certainty or uncertainty is evolutionarily implausible. From this perspective, what constitutes experienced certainty and uncertainty cannot be separated from the context in which thinking occurs. Hence, the switching mechanism must take into account what thinking feels like in the moment. The literature on the relationship between reasoning and culture provides a concretizing example as shown above.

Culture is a set of structures and institutions, values, traditions, and ways of engaging with the social and nonsocial world that are transmitted across generations in a certain time and place. Culture is thus temporally and geographically situated and multilevel. It is situated because it takes place in a certain time and place and dynamically changes as it is transmitted over time and across places. It is multilevel because its influence can be observed in societal-level constructs such as structures and institutions, group-level constructs such as traditions and ways of engaging in the world, and individual-level cultural mindsets – sets of mental representations containing culture-congruent mental content (knowledge about the self and the world), cognitive procedures, and goals (Oyserman, 2011). Considering culture highlights two paradoxes: Accessible cultural mindset and experiencing cultural fluency and cultural disfluency affects thinking (Oyserman, Sorensen, Reber, & Chen, 2009; Oyserman & Yan, 2019). Each increases how confident people are in their inferences and this confidence can result mismatches between reasoning and the task at hand. Moreover, culture shapes uncertainty avoidance, the extent to which a given level of uncertainty is likely to be subjectively experienced as a problem signal (Lu, 2023).

To get through the day, people routinely process enormous quantities of information. From an ecological perspective, people should be sensitive to cues about danger, shifting attention, and ratcheting up to system 2 reasoning in the face of danger signals – cues that something is not right in the situation. If fast reasoning is the default, the implication is that switching into and out of slow reasoning is a function of experiences that trigger uncertainty, suspicion, or other emotions relevant to danger. Culture-as-situated-cognition theory predicts that culture provides people with an implicit map for how everyday situations will unfold, what ambiguous situations likely mean, and why things happen (Oyserman, 2017). People use these feelings as informational inputs (Schwarz, 2001).

A culture-based perspective highlights that the metaphorical system 1 thermostat that turns on and off system 2 reasoning must receive inputs from subjective experiences drawn from the situation. First, the literature on the downstream consequences of cultural fluency and disfluency for reasoning suggests that people may switch to slow reasoning whenever they experience situations that are culturally disfluent. Disfluency provides a problem signal which should trigger system 2. Fluency provides an "all clear" signal that should support remaining in or returning to system 1. A metaphorical thermostat that shut out these cues would not be evolutionarily viable (Oyserman, Novin, Flinkenflögel, & Krabbendam, 2014).

Acknowledgments. The author thanks Norbert Schwarz and Gulnaz Kiper who read an earlier draft and provided helpful comments.

Competing interest. None.

References

Lin, Y., Arieli, S., & Oyserman, D. (2019). Cultural fluency means all is okay, cultural disfluency implies otherwise. *Journal of Experimental Social Psychology*, *84*, 103822.
Lu, J. G. (2023). Two large-scale global studies on COVID-19 vaccine hesitancy over time: Culture, uncertainty avoidance, and vaccine side-effect concerns. *Journal of Personality and Social Psychology*, *124*(4), 683–706. https://doi.org/10.1037/pspa0000320
Mourey, J. A., Lam, B. C., & Oyserman, D. (2015). Consequences of cultural fluency. *Social Cognition*, *33*(4), 308–344.
Oyserman, D. (2011). Culture as situated cognition: Cultural mindsets, cultural fluency, and meaning making. *European Review of Social Psychology*, *22*(1), 164–214. https://doi.org/10.1080/10463283.2011.627187
Oyserman, D. (2017). Culture three ways: Culture and subcultures within countries. *Annual Review of Psychology*, *68*, 435–463.
Oyserman, D., Novin, S., Flinkenflögel, N., & Krabbendam, L. (2014). Integrating culture-as-situated-cognition and neuroscience prediction models. *Culture and Brain*, *2*(1), 1–26.
Oyserman, D., Sorensen, N., Reber, R., & Chen, S. X. (2009). Connecting and separating mind-sets: Culture as situated cognition. *Journal of Personality and Social Psychology*, *97*(2), 217–235. https://doi.org/10.1037/a0015850
Oyserman, D., & Uskul, A. K. (2008/2015). Individualism and collectivism: Societal-level processes with implications for individual-level and society-level outcomes. In F. van de Vijver, D. van Hemert, & Y. Poortinga (Eds.), *Multilevel analysis of individuals and cultures* (pp. 145–173). Ebook: 9780203888032. Psychology Press.
Oyserman, D., & Yan, V. X. (2019). Making meaning: A culture-as-situated-cognition approach to the consequences of cultural fluency and disfluency. In D. Cohen & S. Kitayama (Eds.), *Handbook of cultural psychology* (pp. 536–565). Guilford Press.
Schwarz, N. (1990). Feelings as information: Informational and motivational functions of affective states. In E. T. Higgins & R. M. Sorrentino (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (Vol. 2, pp. 527–561). Guilford Press.
Schwarz, N. (2001). Feelings as information: Implications for affective influences on information processing. In L. L. Martin & G. L. Clore (Eds.), *Theories of mood and cognition: A user's guidebook* (pp. 159–176). Erlbaum.

Deliberation is (probably) triggered and sustained by multiple mechanisms

Gordon Pennycook[a,b]

[a]Hill/Levene Schools of Business, University of Regina, Regina, SK, Canada and [b]Department of Psychology, University of Regina, Regina, SK, Canada
grpennycook@gmail.com
gordonpennycook.com

doi:10.1017/S0140525X22003004, e137

Abstract

De Neys proposes that deliberation is triggered and sustained by uncertainty. I argue that there are cases where deliberation occurs with low uncertainty – such as when problems are excessively complicated and the reasoner decides against engaging in deliberation – and that there are likely multiple factors that lead to (or undermine) deliberation. Nonetheless, De Neys is correct to surface these issues.

De Neys argues – lucidly, in my view – that dual-process theories have been (largely) ill-focused. Indeed, it seems that the core aspect of any dual-process model is not simply how it describes the two types of processes (or "systems"), but rather how describes the *interaction* between the processes. To this end, De Neys proposes that *uncertainty* is the key explanatory factor that determines when "type 2" deliberation is triggered (i.e., the "switch feature").

Although many popular dual-process models have ignored the issue, as noted by De Neys, there is nonetheless a growing body of work that has focused on understanding what triggers deliberation. If intuitive ("type 1") processing is autonomous (i.e., it is triggered automatically from some stimulus or thought process), as myself and others have argued (e.g., Pennycook, 2017; Thompson, 2013), then deliberative processes must be triggered by some underlying cognitive factor or factors. That is, when explaining the progeny of the process, intuitive processes can be explained by simple stimulus–response pairings. However, one needs to posit additional factors that would then lead to subsequent deliberative processes.

In my past work, I have focused on the potential role of response conflict in triggering deliberation (Pennycook, 2023; Pennycook, Fugelsang, & Koehler, 2015) – that is, cases where the system detects a conflict between intuitive outputs lead to subsequent deliberation. The three-stage dual-process model separates the initial "intuition" stage (where processes are initiated autonomously) from the subsequent "metacognition" stage (where conflicts between outputs of stage 1 are monitored). The presence or absence of conflict then determines the extent of deliberation in the final "reason" stage. This matches well with the sort of tasks that are common in the literature because they typically involve a salient (but incorrect) intuition that conflicts with some other relevant (and more normatively accurate) factor. However, as De Neys notes, a model that focuses solely on intuitive conflict does not help explain why deliberation occurs in the absence of intuitions.

To solve this problem, De Neys argues that general uncertainty casts a wider net and can explain both cases where conflict detection leads to deliberation but also cases where there are no apparent intuitions present. As conflict between intuitions leads to uncertainty, so to does a lack of clear answer. There is a lot to like about this proposal; however, I am hesitant to adopt uncertainty as the key explanatory factor.

One possibility is that "uncertainty" is merely correlated with the underlying causal feature (or features) that trigger deliberation. Indeed, there may be cases when uncertainty is high but deliberation is low. For example, when an individual is facing a problem that is prohibitively complicated they surely have a feeling of high uncertainty – nonetheless, the individual may decide to *not* engage in deliberation and to simply not bother attempting a solution (and, indeed, people do tend to prefer tasks that require less cognitive effort; Chong, Bonnelle, & Husain, 2016; Shenhav et al., 2017). Hence, in such a scenario uncertainty would not lead to deliberation.

De Neys has offered a strong framework for understanding how type 1 outputs contribute to the engagement of subsequent deliberation. He even mentions a similar case where deliberation does not lead to an answer and the individual decides to stop deliberating. This is accommodating by noting that opportunity cost could be factored into the uncertainty parameter. However, one may question at that point whether deliberation is prompted by several separate mechanisms and that uncertainty is just a reasonable proxy for many of them.

This issue is related to recent work in the cognitive control literature that has investigated what might lead individuals to engage in effortful processing across situations. For example, the expected-value-of-control model posits that people weigh the costs and benefits of exerting mental effort (Shenhav, Botvinick,

& Cohen, 2013; Shenhav, Prater Fahey, & Grahek, 2021). Of course, it may be that this calculation only occurs after deliberative processes have been initiated by some other process and therefore that costs/benefits play a role in how much deliberation occurs, not whether it is triggered in the first place. Nonetheless, if one takes the view that dual processes can be considered as opposite poles of a single continuum (as opposed to two fundamentally separate processes; De Neys, 2021), then the expected value of control is a factor that may play a role in whether *substantive* deliberation occurs. This is noted by De Neys but the implications for the focus on uncertainty are not addressed.

Relatedly, a great deal of research has focused on apparent individual differences in the willingness to engage in deliberation (see Pennycook [2023] for a review). If one were to focus solely on uncertainty as the factor that triggers deliberation, this would presume that individual differences in uncertainty-related processes (e.g., sensitivity to uncertainty) would be central. However, individual differences in deliberation tend to focus on whether people are willing to engage in cognitive effort (e.g., in the "Need for Cognition" scale; Petty, Brinol, Loersch, & McCaslin, 2009) or in the willingness to question one's intuitions and prior beliefs (e.g., in the "Actively Open-minded Thinking" scale; Pennycook, Cheyne, Koehler, & Fugelsang, 2020). Here too, the key factors that explain whether people engage in a meaningful level of deliberation may not be explained very well by appealing to uncertainty.

To conclude, I agree wholeheartedly with De Neys's argument that the underlying assumptions of many dual-process theories are poorly conceived – particularly as it related to the exclusivity feature. However, I am less convinced that deliberation can be explained by a single underlying causal mechanism. It seems likely to me that there are multiple separate mechanisms that are relevant for understanding why deliberation occurs. In any case, even if I am uncertain about using uncertainty as a key mechanism in dual-process models, I am nonetheless convinced that De Neys's proposal is certainly a step forward.

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Competing interest. The author has co-authored research with the author of the target article in the last 5 years. There are no other conflicts of interests to declare.

References

Chong, T. T.-J., Bonnelle, V., & Husain, M. (2016). Chapter 4 – Quantifying motivation with effort-based decision-making paradigms in health and disease. In B. Studer & S. Knecht (Eds.), *Progress in brain research* (Vol. 229, pp. 71–100). Elsevier. https://doi.org/10.1016/bs.pbr.2016.05.002

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

Pennycook, G. (2017). A perspective on the theoretical foundation of dual-process models. In W. De Neys (Ed.), *Dual process theory 2.0* (pp. 5–39). Routledge. https://doi.org/10.4324/9781315204550

Pennycook, G. (2023). A framework for understanding reasoning errors: From fake news to climate change and beyond. *Advances in Experimental Social Psychology*, *67*, 131–208. https://doi.org/10.1016/bs.aesp.2022.11.003

Pennycook, G., Cheyne, J. A., Koehler, D. J., & Fugelsang, J. A. (2020). On the belief that beliefs should change according to evidence: Implications for conspiratorial, moral, paranormal, political, religious, and science beliefs. *Judgment and Decision Making*, *15*, 476–498. https://doi.org/10.31234/OSF.IO/A7K96

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

Petty, R. E., Brinol, P., Loersch, C., & McCaslin, M. J. (2009). The need for cognition. In M. R. L. & R. H. Hoyle (Eds.), *Handbook of individual differences in social behavior* (pp. 318–329). Guilford Press.

Shenhav, A., Botvinick, M., & Cohen, J. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron*, *79*(2), 217–240. https://doi.org/10.1016/j.neuron.2013.07.007

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience*, *40*(1), 99–124. https://doi.org/10.1146/annurev-neuro-072116-031526

Shenhav, A., Prater Fahey, M., & Grahek, I. (2021). Decomposing the motivation to exert mental effort. *Current Directions in Psychological Science*, *30*(4), 307–314. https://doi.org/10.1177/09637214211009510

Thompson, V. A. (2013). Why it matters: The implications of autonomous processes for dual process theories – Commentary on Evans & Stanovich (2013). *Perspectives on Psychological Science*, *8*, 253–256. https://doi.org/10.1177/1745691613483476

How research on persuasion can inform dual-process models of judgment

Richard E. Petty[a], Duane T. Wegener[a] and Pablo Briñol[b]

[a]Department of Psychology, The Ohio State University, Columbus, OH, USA and [b]Department of Psychology, Universidad Autónoma de Madrid, Madrid, Spain
petty.1@osu.edu
wegener.1@osu.edu
pablo.brinnol@uam.es
https://richardepetty.com
https://psychology.osu.edu/people/wegener.1
https://pablobrinol.com/

doi:10.1017/S0140525X2200293X, e138

Abstract

De Neys makes some useful points regarding dual-process models, but his critique ignores highly relevant theories of judgment from the persuasion literature. These persuasion models predate and often circumvent many of the criticisms he makes of the dual-process approaches he covers. Furthermore, the persuasion models anticipated some of the correctives to dual-process models that he proposes.

De Neys aims to provide a broad critique of prevailing dual-process and system (DP/S) models of judgment in "key fields," as well as introduce a more viable approach (see Petty & Briñol, 2008, on dual-process vs. system frameworks). However, his critique fails to consider theories from the persuasion literature such as the heuristic-systematic (HSM; Chaiken, Liberman, & Eagly, 1989) and elaboration likelihood (ELM; Petty & Cacioppo, 1986) models that are clearly relevant and more highly cited than several of the covered DP/S approaches. Critically, the relevant persuasion models often agree with and predate the core points De Neys makes, and have already addressed some of the key challenges he poses. De Neys emphasizes how his new model is superior to prevailing DP/S models, but ironically his new model is better largely because it mimics features of the earlier persuasion models that were ignored. We illustrate our points largely using the ELM because we are intimately familiar with it, but also because there are numerous ELM studies that support our points (Petty & Briñol, 2012).

The first critique De Neys's offers of DP/S models is that they rely on *exclusivity* – the notion that fast (relatively low thought) and slow (relatively high thought) systems should yield different judgments. In contrast, the author proposes that high- and low-thought processes can: (1) "cue the same response" and (2) might not have "the same features." These two ideas are fundamental to the ELM which explains how and why high- and low-thought processes can result in the same outcome under some circumstances but different outcomes under others. For example, is it better for persuasion to give people 3 or 9 message arguments? The ELM holds that it depends on whether the arguments are cogent or specious and whether people are engaged in relatively high or low amounts of thinking. When thinking is high, people evaluate the merits of the arguments, but when thinking is low, they are more likely to rely on simply counting the arguments using the heuristic – the more the better. Thus, when the arguments are strong, 9 arguments produce more persuasion than 3 regardless of the amount of thinking because processing for merit and counting produce the same outcome. However, when the arguments are weak, the high- and low-thought processes lead to different outcomes. Under low thinking, 9 weak arguments are still more persuasive than 3 because of the quantity heuristic. But, under high thinking, 9 weak arguments are less persuasive than 3 because they produce more negative thoughts (Petty & Cacioppo, 1984).

Regarding the second point, the ELM explains that even though the persuasion outcome is the same under high and low thinking when the arguments are strong (i.e., $9 > 3$), the "features" of enhanced persuasion under 9 arguments can differ because the processes that led to that superiority are different. Specifically, the evaluations induced by 9 arguments over 3 under high thinking are more likely to persist over time, resist change, and guide behavior than the very same evaluations induced via a lower thought heuristic process (Haugtvedt & Petty, 1992). Thus, although we agree with the author's insight, this notion has been evident in the ELM for a long time (for parallels in ELM-guided work on numerical anchoring, see Blankenship, Wegener, Petty, Detweiler-Bedell, & Macy, 2008).

Another critique of DP/S models De Neys offers is that they do not explain how and when people might *switch* between low- and high-deliberation modes. De Neys also postulates that people always switch from low to high deliberation. In contrast, the ELM holds that people can start their processing at high elaboration. For instance, when a person initially views a particular judgment as important enough to think about carefully, there is no need to start with or generate a low-deliberation response first that then has to be corrected (Petty & Cacioppo, 1990). That is, low deliberation is not assumed to be the default mode. Rather, many variables determine whether an initial judgment results from high or low deliberation or whether people shift from one mode to another (Carpenter, 2015).

To explain when people move from a low- to a high-deliberation mode, De Neys proposes that it stems from uncertainty about the correct output (i.e., when low- and high-deliberation modes produce different outcomes). When uncertainty reaches a particular threshold, people shift to high thinking and this deliberation ceases when uncertainty drops below that threshold. Although De Neys's *certainty threshold* notion is quite reasonable, we note that it parallels the earlier *sufficiency principle* from the HSM (Chaiken et al., 1989). Furthermore, according to the persuasion models, in addition to uncertainty (e.g., stemming from ambivalence; Petty, Briñol, & Johnson, 2012), many other variables have been shown to motivate and/or enable enhanced deliberation (e.g., personal relevance of the judgment, responsibility for the judgmental outcome, etc.; Petty & Wegener, 1998).

Another critique is that DP/S models largely hold that the flawed (biased) outcomes occur when the output of low thinking is not corrected by high thinking. In contrast, De Neys proposes that deliberation "does not magically imply that the resulting response will be correct" (target article, sect. 3.3, para. 2). Yet again, persuasion models had already proposed that the amount of thinking and the extent of bias in that thinking are orthogonal (Petty & Cacioppo, 1990; for an example about stereotyping, see Wegener, Clark, & Petty, 2006). Thus, high thinking can sometimes lead to an even more flawed (biased) judgment than low thinking when, for example, a prime biases an initial (fast) judgment that then guides and contaminates the subsequent thinking (Petty, 2001).

In sum, although De Neys makes some reasonable points, a number of those points parallel principles previously proposed and documented in research examining relevant persuasion theories. By ignoring those frameworks, including their applications beyond the persuasion context (e.g., Petty & Briñol, 2014) and especially in judgment and decision-making domains where the criticized DP/S approaches have dominated (e.g., see Wegener, Petty, Blankenship, & Detweiler-Bedell, 2010), De Neys missed an opportunity to provide a more complete and integrative critique of DP/S models.

Competing interest. None.

References

Blankenship, K. L., Wegener, D. T., Petty, R. E., Detweiler-Bedell, B., & Macy, C. L. (2008). Elaboration and consequences of anchored estimates: An attitudinal perspective on numerical anchoring. *Journal of Experimental Social Psychology*, *44*(6), 1465–1476. doi: 10.1016/j.jesp.2008.07.005

Carpenter, C. J. (2015). A meta-analysis of the ELM's argument quality × processing type predictions. *Human Communication Research*, *41*(4), 501–534. doi: 10.1111/hcre.12054

Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic information processing within and beyond the persuasion context. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 212–252). Guilford Press.

Haugtvedt, C. P., & Petty, R. E. (1992). Personality and persuasion: Need for cognition moderates the persistence and resistance of attitude changes. *Journal of Personality and Social Psychology*, *63*(2), 308–319. doi: 10.1037/0022-3514.63.2.308

Petty, R. E. (2001). Subtle influences on judgments and behaviors: Who is most susceptible? In J. Forgas & K. D. Williams (Eds.), *Social influence: Direct and indirect processes* (pp. 129–146). Psychology Press.

Petty, R. E., & Briñol, P. (2008). Persuasion: From single to multiple to meta-cognitive processes. *Perspectives on Psychological Science*, *3*(2), 137–147. doi: 10.1111/j.1745-6916.2008.00071.x

Petty, R. E., & Briñol, P. (2012). The elaboration likelihood model. In P. A. M. Van Lange, A. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (Vol. 1, pp. 224–245). Sage. doi: 10.4135/9781446249215.n12

Petty, R. E., & Briñol, P. (2014). The elaboration likelihood and meta-cognitive models of attitudes: Implications for prejudice, the self, and beyond. In J. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 172–187). Guilford Press.

Petty, R. E., Briñol, P., & Johnson, I. (2012). Implicit ambivalence. In B. Gawronski & F. Strack (Eds.), *Cognitive consistency: A fundamental principle in social cognition* (pp. 178–201). Guilford Press.

Petty, R. E., & Cacioppo, J. T. (1984). The effects of involvement on response to argument quantity and quality: Central and peripheral routes to persuasion. *Journal of Personality and Social Psychology*, *46*(1), 69–81. doi: 10.1037/0022-3514.46.1.69

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology*, *19*, 123–205. doi: 10.1016/S0065-2601(08)60214-2

Petty, R. E., & Cacioppo, J. T. (1990). Involvement and persuasion: Tradition versus integration. *Psychological Bulletin*, *107*(3), 367–374. doi: 10.1037/0033-2909.107.3.367

Petty, R. E., & Wegener, D. T. (1998). Attitude change: Multiple roles for persuasion variables. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 1, pp. 323–390). McGraw-Hill.

Wegener, D. T., Clark, J. K., & Petty, R. E. (2006). Not all stereotyping is created equal: Differential consequences of thoughtful versus non-thoughtful stereotyping. *Journal of Personality and Social Psychology*, *90*(1), 42–59. doi: 10.1037/0022-3514.90.1.42

Wegener, D. T., Petty, R. E., Blankenship, K. L., & Detweiler-Bedell, B. (2010). Elaboration and numerical anchoring: Implications of attitude theories for consumer judgment and decision making. *Journal of Consumer Psychology*, *20*(1), 5–16. doi: 10.1016/j.jcps.2009.12.003

The dual-system approach is a useful heuristic but does not accurately describe behavior

Jeffrey W. Sherman and Samuel A. W. Klein

Department of Psychology, University of California, Davis, CA, USA
jsherman@ucdavis.edu
sawklein@ucdavis.edu
https://socolab.faculty.ucdavis.edu/

doi:10.1017/S0140525X22003053, e139

Abstract

We argue that the dual-system approach and, particularly, the *default-interventionist* framework favored by De Neys unnecessarily constrains process models, limiting their range of application. In turn, the accommodations De Neys makes for these constraints raise questions of parsimony and falsifiability. We conclude that the extent to which processes possess features of system 1 versus system 2 must be tested empirically.

De Neys has described an elegant dual-process model to overcome conceptual shortcomings among other models. At the same time, the model is constrained to fit a systems approach and a *default-interventionist* framework, which significantly limits its range of application. We question the necessity and value of these constraints and key components of the model designed to accommodate those constraints.

De Neys restricts his model to accounting for behavior that can be described in a default-interventionist framework, in which system 2 processes are engaged only when system 1 fails to offer an adequate response. However, not all dual-process models share the default-interventionist structure (e.g., Gilbert, 1999; Sherman, Klauer, & Allen, 2010). In fact, many models assume that system 2 is the default. For example, Jacoby's work on recognition memory specifies that familiarity (system 1) only drives responses when recollection (system 2) fails (Jacoby, 1991). In Payne's work on implicit stereotyping, people rely on automatically activated stereotypic associations (system 1) only when judges are unable to determine whether they are looking at a gun or a tool (system 2; Payne, 2001). Ferreira, Garcia-Marques, Sherman, and Sherman (2006) extended the same logic to standard judgment and decision-making errors (e.g., base-rate; conjunction; ratio-bias effect; and law of large number problems). Importantly, direct modeling comparisons in these domains show that system 2 default models better account for these judgments than a default-interventionist model. As well, none of the tasks in these examples inherently demands the prioritization of system 2 (a condition De Neys identifies as irrelevant to his discussion of dual systems).

It is the default interventionism requirement that necessitates a switching mechanism, which we find problematic in a number of ways. Most basically, we are skeptical that a serial model is more efficient than a parallel model. Certainly, it is an unusual claim among general theories of information processing. In any case, De Neys solves this problem by positing that there may be system 1 versions of system 2 processes that do operate in parallel to system 1. However, this accommodation further requires that conflicting responses and their detection must also reside in system 1. These claims are undermined by considerable behavioral and neuroscience evidence that conflict monitoring requires attention and effort, presumably indicating a system 2 process. As well, conflict monitoring is associated with activation in the dorsal anterior cingulate cortex (dACC), a brain region involved in higher-level function. The dACC is associated with attention to a problem and effort to address it with intentional action (e.g., Carter & van Veen, 2007).

Of course, De Neys can evade these problems by simply positing that any conflict detection that appears to involve system 2, in fact, involves a system 1 routinization of system 2 (and, presumably, is generated in a site different than dACC). But doing so raises concerns about parsimony and unfalsifiability. If there is always the possibility of unmeasured system 1 operations, then it is not clear how the model could possibly be falsified.

Adherence to the requirements of a dual-process or system approach also unnecessarily constrains the model and its assumptions. We certainly concur with De Neys that systems 1 and 2 cannot be expected to yield unique responses. However, process exclusivity – the notion that, at any given time, processes must belong solely to system 1 or 2 – also is problematic. For example, driving may become quite efficient (system 1 feature) but continue to require intention (system 2 feature). The ability to inhibit racial bias is compromised by old age and alcohol (suggesting system 2), yet frequently operates effectively on implicit measures of bias (suggesting system 1; Calanchini & Sherman, 2013). Thus, the same process may possess features of either system and those features (e.g., intention; awareness; controllability; efficiency) rarely all coincide (Gawronski, Sherman, & Trope, 2014).

More broadly, these issues highlight the problematic dual-process tendency to conflate *operating principles* and *operating conditions*. Whereas operating principles refer to the qualitative nature of a process (i.e., what the process does – detect; suppress), operating conditions refer to the conditions under which the process operates (e.g., with or without intention or cognitive resources). In dual-process models, it is common to assume that certain processes (e.g., response inhibition) must possess certain features (e.g., resource-dependence). Such assumptions are often necessary to maintain the claim of two distinct process types or systems. However, whether a process possesses features ascribed to system 1 and/or 2 is an empirical question that should be tested directly (Sherman, Krieglmeyer, & Calanchini, 2014).

In our own research, we have adopted this approach via the use of multinomial modeling techniques (Sherman et al., 2010). We found De Neys's model especially interesting in that, in many ways, it aligns with a model we have applied extensively (Sherman et al., 2008). Briefly, the Quad model proposes that, when an automatized response (implicit bias) conflicts with an intended response (respond favorably), a third process acts as arbiter to decide the winner. Obviously, this bears similarity to De Neys's portrayal of conflict detection and resolution, which we found highly valuable. However, we make no assumptions about the system 1 versus 2 features of these processes. Rather, we measure the processes independently and directly examine how they respond to interventions. For example, we know that both the intended response and conflict arbiter processes are relatively inefficient because they are undermined by short response deadlines. We believe this is the way forward for describing and testing process models.

If dual processes or systems cannot be distinguished by exclusive outcomes, processes, or features of processes, one must ask what is the point, particularly if they necessitate the sorts of work arounds De Neys must build to make it all work. It is more productive to simply identify the processes involved in some operation and the conditions under which they operate with no constraint of fitting into distinct process types or systems. The dual-process approach is effective as a heuristic for thinking about human behavior, but rarely describes that behavior accurately.

Financial support. This work was supported by the National Science Foundation (BCS 2215236).

Competing interest. None.

References

Calanchini, J., & Sherman, J. W. (2013). Implicit attitudes reflect associative, non-associative, and non-attitudinal processes. *Social and Personality Psychology Compass*, *7*, 654–667.

Carter, C. S., & van Veen, V. (2007). Anterior cingulate cortex and conflict detection: An update of theory and data. *Cognitive, Affective, and Behavioral Neuroscience*, *4*, 367–379.

Ferreira, M. B., Garcia-Marques, L., Sherman, S. J., & Sherman, J. W. (2006). Automatic and controlled components of judgment and decision making. *Journal of Personality and Social Psychology*, *91*, 797–813.

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). Guilford Press.

Gilbert, D. T. (1999). What the mind's not. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 3–11). Guilford Press.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513–341.

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, *81*, 181–192.

Sherman, J. W., Gawronski, B., Gonsalkorale, K., Hugenberg, K., Allen, T. J., & Groom, C. J. (2008). The self-regulation of automatic associations and behavioral impulses. *Psychological Review*, *115*, 314–335.

Sherman, J. W., Klauer, K. C., & Allen, T. J. (2010). Mathematical modeling of implicit social cognition: The machine in the ghost. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 156–175). Guilford Press.

Sherman, J. W., Krieglmeyer, R., & Calanchini, J. (2014). Process models require process measures. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual process theories of the social mind* (pp. 121–138). Guilford Press.

The distinction between long-term knowledge and short-term control processes is valid and useful

Richard M. Shiffrin[a], Walter Schneider[b] and Gordon D. Logan[c]

[a]Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA; [b]Department of Psychology, University of Pittsburgh, Pittsburgh, PA, USA and [c]Department of Psychology, Vanderbilt University, Nashville, TN, USA
shiffrin@indiana.edu
wws@ptt.edu
gordon.logan@vanderbilt.edu

doi:10.1017/S0140525X2200317X, e140

Abstract

The binary distinction De Neys questions has been put forward many times since the beginnings of psychology, in slightly different forms and under different names. It has proved enormously useful and has received detailed empirical support and careful modeling. At heart the distinction is that between knowledge in long-term memory and control processes in short-term memory.

De Neys makes a case for the lack of support and specificity of the binary conceptual distinction between fast and slow thinking. It is certainly the case that any binary distinction applied to the complexities of human cognition (including perception, memory, and decision making) could not possibly be more than a crude approximation to reality. Yet the binary distinction he questions has been put forward many times since the beginnings of psychology, in slightly different forms and under different names. It has proved enormously useful; its constant resurrection testifies to its utility; in some of its forms it has received detailed empirical support and careful modeling. One can focus on any one of these binary instantiations and find much to criticize, but there is a fundamental basis for human cognition that is being captured and a look at the history of these concepts shows many similarities and a great deal of support.

The conceptual distinction is closely related to those between automatic and controlled processing, between short-term memory (including working memory) and long-term memory, between automatic and attentive processing, between working memory and semantic memory, between the use of rules versus expertise, between the use of algorithms versus memory, between beginners versus experts in motor tasks, games, and sports, between fast and slow thinking, between intuitive versus deliberate thinking and decision making, and more along these lines.

One of the first empirical presentations of the ideas was published by Bryan and Harter in *Psychological Review* in 1899. They examined the development of automaticity in the receiving of telegraphy, arguing for stages in which a kind of chunking in memory took place, so that perception of the dots and dashes being sent would occur at larger and larger scales, starting for example with letters, and later with words and then phrases or sentences. That led to a number of additional explorations in the 20 years following. The basic idea was that performance at first goes step by step, dot by dot, dash by dash, letter by letter, but as learning proceeds the incoming dots and dashes are perceived in larger and larger groups, and long-term memory and knowledge can thereby greatly improve speed of receiving telegraphy. Seventy-five years later, LaBerge and Samuels (1974) applied these ideas to the development of automaticity in reading.

Related distinctions proved critically important in theorizing how memory operates, as exemplified in the chapter by Atkinson and Shiffrin titled "*Human memory: A proposed system and its control processes*" (1968). A key distinction was between a relatively permanent long-term memory containing knowledge and a short-term memory, also called working memory, in which control processes controlled the operations of cognition, including access to long-term memory and knowledge.

The distinction between learned behavior stored in long-term memory and control processes in short-term memory received what surely is it most thorough and complete empirical exploration by Schneider and Shiffrin (1977) and Shiffrin and Schneider

(1977) in the form of a contrast between automatic and controlled processing (later termed a distinction between automatic and attentive processing; Shiffrin, 1988). They used visual and memory search to show how step-by-step controlled processing is used initially for both forms of search, and used throughout for both forms of search when training was inconsistent (termed varied mapping), but gradually became automatized in various ways as consistent training (termed consistent mapping) would cause learning to take place; for example, a target may come to call attention to itself automatically.

Another thorough and careful empirical and theoretical investigation of these ideas was carried out by Gordon Logan and colleagues, for example as laid out by Logan in Psychological Review in 1988. In various articles about that time Logan and colleagues investigated the automatization of multi-step processes like counting dots or verifying alphabet arithmetic equations, showing that with consistent practice, the multi-step algorithm is replaced by rapid retrieval of previously encountered solutions. Slow thinking is replaced by automatic retrieval from long-term memory.

The above two examples take the form of a distinction between active use of attention and automatic processing. In these examples, as in all other binary divisions of cognition, the boundary between the two forms of processing is imprecise. For example, in the absence of automatic processing and learned attention to targets, visual objects tend to be examined one at a time; processing time rises as the number of objects increases because on average the searched-for target is found halfway through the sequence of comparisons. As consistent training proceeds, a target comes to draw attention to itself, so that the target is found in the first step, rather than at a random point in the sequence of comparisons. However that automatic attention process may be slower than a single comparison carried out by a controlled process; when only one object is presented a controlled comparison can be faster than automatic attention attraction. We note that Vim De Neys faults dual-processing approaches because of lack of evidence of "exclusivity." In much of the automaticity literature the view is both processes operate in parallel and interact.

There is a growing literature of biological separation of automatic and controlled processing. Strokes that damage structures such as the right parietal cortex can severely compromise control processes, but spare automatic processing, as seen in neglect (Mesulam, 2000). Schneider and Chein (2003) review much of this literature. Chein and Schneider (2012) highlight the way that learning alters the activity of neural networks as automatic processing develops – see Figure 1.

The distinctions we have been discussing have also played an important role in applications in society. To take just a few examples they have driven research and practice in reading education in children (LaBerge & Samuels, 1974; Samuels & Flor, 1997), in medical decision making (Evans, Birdwell, & Wolfe, 2013), in aging (Fisk, Rogers, Cooper, & Gilbert, 1997; Hasher & Zacks, 1979), and in clinical science (Huijbregts et al., 2003).

It would take a book rather than this commentary to trace all these closely related binary distinctions, because they have appeared and been used throughout the history of psychological research, albeit under various names. Research on them

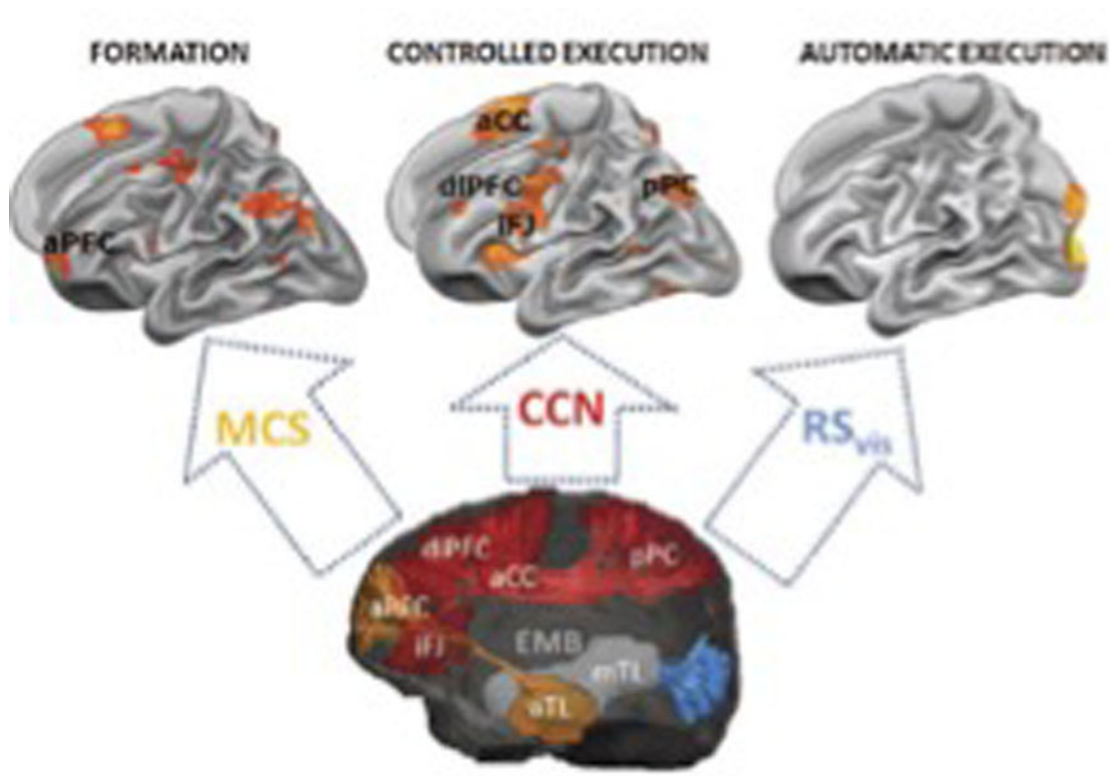

Figure 1 (Shiffrin et al.). How learning in the form of development of automaticity alters activity in neural networks, seen in Figure 2 of Chein and Schneider (2012). Functional MRI reveals changes in brain activation as learning proceeds in a simple visual-discrimination task. Initial performance is associated with increased activity in the anterior prefrontal cortex (aPFC), of the metacognitive system (MCS). After the first few trials, activity declines in the aPFC and increases in the interconnected regions of the cognitive control network (CCN) – the dorsolateral prefrontal cortex (dlPFC), the anterior cingulate cortex (aCC), the posterior parietal cortex (pPC), and the inferior frontal junction (iFJ) – to support controlled execution of the task. After considerable practice, automatic processes develop and activity declines generally.

https://doi.org/10.1017/S0140525X2200142X Published online by Cambridge University Press

demonstrates great utility in classifying human cognition in these ways, as demonstrated by a great deal of careful empirical research, theorizing, and quantitative modeling in certain of these domains. Notwithstanding the admitted imprecision of these binary conceptual divisions of cognition, and the differences between them, we believe there is a fundamental importance and utility to the distinction between control processes carried out in short-term working memory, and automatic learned processes stored in long-term memory as knowledge. The message conveyed by De Neys to this extent misses the "big picture" and is misleading.

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Competing interest. None.

References

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89–195). Academic Press.
Bryan, W. L., & Harter, N. (1899). Studies on the telegraphic language: The acquisition of a hierarchy of habits. *Psychological Review*, *6*(4), 345–375.
Chein, J., & Schneider, W. (2012). The brain's learning and control architecture. *Current Directions in Psychological Science*, *21*(2), 78–84.
Evans, K. K., Birdwell, R. L., & Wolfe, J. M. (2013). If you don't find It often, you often don't find It: Why some cancers are missed in breast cancer screening. *PLoS ONE*, *8*(5), e64366.
Fisk, A. D., Rogers, W. A., Cooper, B. P., & Gilbert, D. K. (1997). Automatic category search and its transfer: Aging, type of search, and level of learning. *Journal of Gerontology: Psychological Sciences*, *52B*, 91–102.
Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, *108*(3), 356–388.
Huijbregts, S. C. J., De Sonneville, L. M. J., Van Spronsen, F. J., Berends, I. E., Licht, R., Verkerk, P. H., & Sergeant, J. A. (2003). Motor function under lower and higher controlled processing demands in early and continuously treated phenylketonuria. *Neuropsychology*, *17*(3), 369–379.
LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*, 293–323. https://doi.org/10.1016/0010-0285(74)90015-2
Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*(4), 492–527.
Mesulam, M.-M. (2000). Attentional networks, confusional states, and neglect syndromes. In M.-M. Mesulam (Ed.), *Principles of behavioral and cognitive neurology* (2nd ed., pp. 174–256). Oxford University Press.
Samuels, S. J., & Flor, R. F. (1997). The importance of automaticity for developing expertise in reading. *Reading and Writing Quarterly*, *13*, 107–121.
Schneider, W., & Chein, J. M. (2003). Controlled & automatic processing: From mechanisms to biology. *Cognitive Science*, *27*, 525–559.
Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*, 1–66.
Shiffrin, R. M. (1988). Attention. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey, & R. D. Luce (Eds.), *Stevens' handbook of experimental psychology: Perception and motivation; learning and cognition* (pp. 739–811). Wiley.
Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, *84*, 127–190.

When a thinker does not want to think: Adding meta-control into the working model

Miroslav Sirota

Department of Psychology, University of Essex, Colchester, UK
msirota@essex.ac.uk

doi:10.1017/S0140525X22003181, e141

Abstract

De Neys proposes an elegant solution to several theoretical problems of the dual-process theories but underspecifies the role of motivation in initiating, intensifying, and ceasing deliberation. Therefore, I suggest including a meta-cognitive control component in the working model that can moderate deliberation, for instance by affecting the deliberation threshold.

I applaud De Neys for proposing a new working model of the dual-process theory that solves its two theoretical conundrums. Admirably, the proposed model integrates recent evidence, offers precise, testable hypotheses, and can be computationally implemented. It provides an elegant answer to the questions of what makes us think and what makes us stop thinking. However, it primarily focuses on bottom-up processes and underspecifies top-down processes, such as the role of motivation in initiating, intensifying, and ceasing deliberation. In other words, the working model should have a suite of mechanisms that help us decide when hard thinking is needed but also when it is worthwhile.

Imagine, for example, a situation where a person faces a complex mathematical problem, which does not trigger any initial intuition, and deliberation has been activated. According to the working model, it ceases only if the uncertainty parameter, U, decreases under the critical deliberation threshold, d (e.g., reaches the conflict resolution). So, deliberation cannot stop if a person cannot achieve a sufficiently significant decrease in the uncertainty parameter (e.g., it does not resolve the conflict between the two conflicting intuitions). But a thinker cannot deliberate endlessly because deliberation is costly. Simply put, the current working model does not account for situations when a thinker does not want to think – so hard, so long, or at all – about the problem. Yet, prior research in higher cognition identified empirical and theoretical arguments supporting the critical importance of motivation to deliberate (e.g., Evans, 2011; Stanovich & West, 1998). For instance, in one dual-process model, motivational factors regulate the level of critical effort, which determines whether a reasoner will endorse the default answer as justified or try to correct it (Evans, 2011).

To resolve these issues, I propose expanding the "opportunity cost factor" suggestion presented in the target article (sect. 4.3, para. 3) and including a meta-cognitive component of control allocation into the working model. Such control allocation mechanisms have been proposed in the literature investigating control allocation over lower cognition tasks, such as Stroop tasks, and have been supported by behavioural and neuropsychological evidence (e.g., Kool, Shenhav, & Botvinick, 2017; Shenhav et al., 2017; Shenhav, Prater Fahey, & Grahek, 2021). For instance, the control allocation component can compute the efficiency of the deliberation to achieve the desired outcome while taking the cost and benefits of deliberation into account. Some initial evidence points to the fact that people consider the costs and benefits of deliberation when correcting reasoning (Sirota et al., in press). For instance, the performance reward and imposed cost affect how much time individuals allocate to correcting their initial errors and, in turn, problem-solving accuracy. So, the meta-cognitive control component is involved in the switching (on and off) of thinking by considering the efficacy of deliberation and its cost and benefits.

There might be different pathways by which meta-cognitive control can interact with uncertainty monitoring; for instance, it can directly affect uncertainty (De Neys, target article). It can also modulate the deliberation threshold: It might decrease or increase the critical deliberation threshold while not affecting the uncertainty parameter. For instance, it can make the deliberation threshold high and, in turn, make deliberation more challenging to switch off if the overall value of reaching the correct answer by deliberating is big (e.g., a maths problem solved during an important exam). So, the uncertainty parameter must be minimal to reach the deliberation threshold. On the other hand, the meta-cognitive control can make the threshold low and, in turn, deliberation easy to switch off if the overall value is small (e.g., a maths problem solved during an anonymous experimental session that participants found tedious). Thus, even weak intuitions generating high uncertainty can pass it. For instance, if the uncertainty initiated deliberation, but the deliberation was not as efficient as assumed with the type of problem, or the costs of deliberation were too high, then the threshold might be lowered. Here, the control's overall value is driven not only by the cost (whether intrinsic or opportunity costs) but also by the control efficacy and the reward one can ascribe to deliberation. Furthermore, to avoid the same theoretical traps outlined in the target article, one can assume that this component computes such values more or less effortlessly, whether by retrieving cached information about the reward and cost associated with the task or by estimating the value heuristically from task cues (see Kool, Gershman, & Cushman, 2018).

Finally, one can also speculate whether such a meta-cognitive component can help to resolve other open questions concerning deliberation listed in section 4.3. First, the control allocation component can modify the deliberation intensity – not only the duration. For instance, with high-stakes outcomes, control allocation can intensify, not just prolong deliberation. Second, it can also assist with deciding which type of deliberation processes are carried out (e.g., default answer justification, default answer correction). For instance, a reasoner might compare the overall values of deliberation needed to justify and correct the default answer and decide that justification is a more beneficial use of deliberation resources.

Thus, including the meta-cognitive component of control allocation into the working model can resolve several open questions of the working model. It can also better integrate research and theory on the role of motivation in thinking and be combined with the other model components.

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Competing interest. None.

References

Evans, J. S. B. T. (2011). Dual-process theories of reasoning: Contemporary issues and developmental applications. *Developmental Review*, *31*(2–3), 86–102. https://doi.org/10.1016/j.dr.2011.07.007

Kool, W., Gershman, S. J., & Cushman, F. A. (2018). Planning complexity registers as a cost in metacontrol. *Journal of Cognitive Neuroscience*, *30*(10), 1391–1404. https://doi.org/10.1162/jocn_a_01263

Kool, W., Shenhav, A., & Botvinick, M. M. (2017). Cognitive control as cost–benefit decision making. In T. Egner (Ed.), *The Wiley handbook of cognitive control* (pp. 167–189). Wiley.

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience*, *40*(1), 99–124. https://doi.org/10.1146/annurev-neuro-072116-031526

Shenhav, A., Prater Fahey, M., & Grahek, I. (2021). Decomposing the motivation to exert mental effort. *Current Directions in Psychological Science*, *30*(4), 307–314. https://doi.org/10.1177/09637214211009510

Sirota, M., Juanchich, M., & Holford, D. (in press). Rationally irrational: When people do not correct their reasoning errors even if they could. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0001375

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General*, *127*(2), 161.

A good architecture for fast and slow thinking, but exclusivity is exclusively in the past

Keith E. Stanovich[a] and Maggie E. Toplak[b]

[a]Department of Applied Psychology and Human Development, University of Toronto, Toronto, ON, Canada and [b]Department of Psychology, LaMarsh Centre for Child and Youth Research, York University, Toronto, ON, Canada
keith.stanovich@utoronto.ca
mtoplak@yorku.ca
http://keithstanovich.com/Site/Home.html
http://maggietoplak.com/

doi:10.1017/S0140525X22002904, e142

Abstract

No doubt older work in the dual-process tradition overemphasized the importance and frequency of the override function, and the working model in this target article provides a useful corrective. The attempt to motivate the model using the so-called exclusivity assumption is unnecessary, because no recent dual-process model in the reasoning literature has rested strongly on this assumption.

The target article provides a valuable summary of the current state of play in dual-process theorizing and presents a working model that provides a basic architecture that incorporates most recent research. The working model has much to recommend it whether or not one endorses the historical narrative of developments in this area.

One of the prime motivations for the working model is said to be the correction of a mistaken assumption in the dual-process literature – the assumption of exclusivity. This assumption is that "traditional dual-process models have typically conceived intuition and deliberation as generating unique responses such that one type of response is exclusively tied to deliberation and is assumed to be beyond the reach of the intuitive system" and it is said to be a "foundational dual-process assumption" (target article, sect. 5, para. 1). The target article omits citation of any particular dual-process theory that contained this assumption and that was published after 2000.

Some of us are old enough to have grown up with the dual-process theories of information processing that were so popular in the 1970s such as those of Posner and Snyder (1975) and Shiffrin and Schneider (1977), both of which made clear that information repeatedly transformed by control process operations could become automatized in (what is now called) system 1. Likewise, those of us enamored with the LaBerge and Samuels'

(1974) automaticity theory of reading were captured by the idea of higher and higher levels of text structure becoming automatized with practice as a young child developed.

Certainly by the time that Stanovich and West (2000; see Stanovich, 1999) introduced the system 1/system 2 terminology into the psychology of reasoning, it was well established that both information and strategies originally used by system 2 could also become instantiated in system 1. Stanovich (2004) made "the possibility of the higher-level goal states of the analytic system becoming installed in the more rigid and inflexible System 1 through practice" (p. 66) one of the themes of a book-length treatment of dual-process theory (see Fig. 2.2 and 7.2 in that volume). Other dual-process theorists followed suit in the early part of this century (Evans, 2003).

Exclusivity as a background assumption of most theorists in reasoning had disappeared as far back as two decades ago. Has any major, influential theorist clearly defended the exclusivity assumption since 2000? There is no quote or citation to this effect in the target article. We must clarify here that our focus and expertise is solely on the reasoning literature.

To be clear, there is some inconsistency in the target article concerning the historical role of the exclusivity assumption. Late in the essay (sect. 4.3, para. 1), De Neys describes how "the basic idea that an originally deliberate response may be automatized through practice, is theoretically sound (e.g., Shiffrin & Schneider, 1977) and well-integrated in traditional dual-process models (e.g., Evans & Stanovich, 2013; Rand et al., 2012)." The citation of Shiffrin and Schneider and the phrase "well-integrated in traditional dual-process models" (target article, sect. 4.3, para. 1) is consistent with the history we have been describing in this commentary. In short, the field moved past the exclusivity assumption some time ago. Yet this is somewhat inconsistent with the later part of the essay when it is called a "foundational dual process assumption" that creates "paradoxes that plague the traditional model" (target article, sect. 5, para. 2).

Earlier in the essay there is a puzzling attempt to finesse the conclusions we are drawing here. The target article allows that with repeated exposure any response that might initially require deliberation can become highly compiled and automatized, but claims that "although such a claim is uncontroversial for the alleged system 1 response in traditional dual-process models … it is assumed here that it also applies to the alleged system 2 response" (target article, sect. 3.1, para. 3). This discussion is very confused by the ill-advised term "alleged system 2 response" (and likewise confused by the term "alleged deliberative response"). Response labels shouldn't make reference to the mental state of an imaginary theorist. In a typical heuristics and biases task, two potential responses are usually pitted against one another – one normative and one non-normative. The normative response is the normative response – regardless of how it arose from a processing sequence point of view. The latter is what theories of internal processing are designed to explain.

That the automatization process included normative responses deriving from high-level mindware being repeatedly executed by system 2 processing has also been well established for a while now. Over a decade ago, when describing the domains to which Shiffrin and Schneider-type automatized learning applied, Stanovich (2009) stressed that system 1 contained high-level mindware: "decision-making principles that have been practiced to automaticity" (p. 57). These would include the probabilistic reasoning principles, such as the importance of sample size and the multiplicative probability rule, that those tutored in statistics come to think of as second nature. Indeed, some statistics instructors become unable to empathize with their students for whom the basic probability axioms are not transparent. The instructor can no longer remember when these axioms were not primary intuitions.

More so than dual-process theorists themselves, many *critics* of dual-process theory have been stuck in the past – focusing on straw man assumptions that were left by the wayside decades before (see Evans & Stanovich, 2013). The synthesis in the target article rightly focuses the field on the future. The architecture presented in the target article is motivated by both theory (De Neys & Pennycook, 2019; Evans, 2019; Stanovich, 2018) and recent empirical work (Bago & De Neys, 2017; Newman, Gibb, & Thompson, 2017; Trippas, Thompson, & Handley, 2017). The author rightly points out that critical aspects of the architecture are orthogonal to the single- versus dual-process debate. It does not need to rely on a straw man motivation. It stands on its own as a valid synthesis of the state-of-play of reasoning work that uses the fast/slow distinction in whatever manner. No doubt older work in the dual-process tradition overemphasized the importance and frequency of the override function, and this target article provides a useful corrective.

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Competing interest. None.

References

Bago, B., & De Neys, W. (2017). Fast logic? Examining the time course assumption of dual process theory. *Cognition*, *158*, 90–109.
De Neys, W., & Pennycook, G. (2019). Logic, fast and slow: Advances in dual-process theorizing. *Current Directions in Psychological Science*, *28*, 503–509.
Evans, J. S. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*, 454–459.
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*, 383–415.
Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*, 223–241.
LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*, 293–323.
Newman, I. R., Gibb, M., & Thompson, V. A. (2017). Rule-based reasoning is fast and belief-based reasoning can be slow. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*, 1154–1170.
Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 55–85). Wiley.
Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature*, *489*, 427–430.
Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, *84*, 127–190.
Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Erlbaum.
Stanovich, K. E. (2004). *The robot's rebellion: Finding meaning in the age of Darwin*. University of Chicago Press.
Stanovich, K. E. (2009). Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory? In J. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 55–88). Oxford University Press.
Stanovich, K. E. (2018). Miserliness in human cognition: The interaction of detection, override and mindware. *Thinking & Reasoning*, *24*, 423–444.
Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*, 645–726.
Trippas, D., Thompson, V. A., & Handley, S. J. (2017). When fast logic meets slow belief: Evidence for a parallel-processing model of belief bias. *Memory & Cognition*, *45*, 539–552.

Explaining normative–deliberative gaps is essential to dual-process theorizing

Edward J. N. Stupple[a] and Linden J. Ball[b]

[a]School of Psychology, College of Health, Psychology and Social Care, University of Derby, Derby, UK and [b]School of Psychology & Computer Science, University of Central Lancashire, Preston, UK
e.j.n.stupple@derby.ac.uk;
https://www.derby.ac.uk/staff/ed-stupple/
lball@uclan.ac.uk;
https://www.uclan.ac.uk/academics/professor-linden-ball

doi:10.1017/S0140525X22003107, e143

Abstract

We discuss significant challenges to assumptions of exclusivity and highlight methodological and conceptual pitfalls in inferring deliberative processes from reasoning responses. Causes of normative–deliberative gaps are considered (e.g., disputed or misunderstood normative standards, strategy preferences, task interpretations, cognitive ability, mindware and thinking dispositions) and a soft normativist approach is recommended for developing the dual-process 2.0 architecture.

Dual-process 2.0 accounts are increasingly compelling, and we welcome De Neys's proposed model, which we bolster here by noting further challenges to assumptions of "exclusivity" (the notion that intuition and deliberation generate unique responses). We additionally argue for considerable methodological care when exploring the nature of deliberative processing.

Among the most crucial considerations when devising thinking, reasoning, and decision-making tasks is determining what constitutes a "correct" answer and what it means when participants produce this answer. Indeed, De Neys cautions against an "ought-is fallacy" (Elqayam & Evans, 2011), which arises when responses aligning with "normative" theories (e.g., predicate logic or Bayes' theorem) are viewed as being *diagnostic* of deliberation. We contend that normative standards, although useful for performance benchmarking, can present blind spots for experimental design and theory building. As such, we concur with Elqayam and Evans (2011) that constructing theories of reasoning around normative standards is problematic for understanding psychological processes.

To evaluate deliberative processing successfully, it seems prudent to adopt a "soft normativist" (Stupple & Ball, 2014) or "descriptivist" (Elqayam & Evans, 2011) approach. Accordingly, research programs should acknowledge the distorting lens of normative standards (while also avoiding the trap of relativism), recognizing that although normative standards may be correlated with deliberation, they are not causally linked to it (Stupple & Ball, 2014). From a soft-normativism perspective, "normative–deliberative gaps" are expected for many reasons (e.g., disputed or misunderstood norms, strategy preferences, alternative task interpretations, cognitive ability and mindware constraints, and impoverished thinking dispositions), necessitating careful consideration.

Normative standards should also be contested and evaluated whenever multiple, candidate standards exist (Stenning & Varga, 2018). For some tasks, the normative response is uncontroversial, but for others, participants must make sense of task requirements and may not construe the task as intended. For example, Oaksford and Chater (2009) proposed an alternative normative standard for the Wason selection task based upon "information gain," which is consistent with the most common responses (contrasting with Wason's [1966] logicist proposals). Oaksford and Chater (2009) extend this perspective to demonstrate that logical fallacies can be rationally persuasive. Indeed, caution is advised for researchers who associate endorsement of fallacies with a *lack* of deliberation. It is prudent not simplistically to equate standard normative responses with deliberative thinking without also considering individual goals.

In most thinking tasks, participants are not explicitly prescribed a goal or norm. Indeed, Cohen (1981) famously argued that reasoning research presents "trick" questions with minimalist instructions to naïve participants. The assumption that participants identify tasks as requiring deliberation may itself be naïve. Stupple and Ball (2014) proposed that when naïve participants attempt novel reasoning problems, they determine an appropriate normative standard and select a strategy through a process of "informal reflective equilibrium." Through this, increasing familiarity with problem forms – even in the absence of feedback – can result in participants aligning with normative responses assumed to require deliberation (Ball, 2013; Dames, Klauer, & Ragni, 2022). This alignment need not be deliberative, however, but could instead entail detection of *patterns* in problems and an increasing intuition strength for normatively aligned heuristic responses.

These variations in participants' goals and strategies are captured by Markovits, Brisson, and de Chantal (2017) (cf. Verschueren, Schaeken, & d'Ydewalle, 2005), who demonstrated individual differences in strategy preferences (probabilistic vs. counterexample) that are orthogonal to preferences for intuitive versus deliberative thinking. These strategies have implications for the interplay between deliberation and normative standards. Participants adopting a counterexample strategy (based on mental models) versus a probabilistic strategy (based on information gain or probability heuristics; Beeson, Stupple, Schofield, & Staples, 2019; Oaksford & Chater, 2009; Verschueren et al., 2005) may differ in their task construal and understanding of "correct" answers. Although it is unclear whether strategies necessarily entail adoption of particular normative standards, responding to a problem in terms of information gain versus a necessary truth derived from a mental model would reasonably be assumed to require differing degrees of deliberation and differing use of intuitive cues.

When judging whether deliberation has occurred, we also suggest that responses can be less reliable than response times. For example, for the lily-pad cognitive reflective task (CRT) problem, incorrect non-intuitive answers averaged longer response times than incorrect intuitive or correct answers (Stupple, Pitchford, Ball, Hunt, & Steel, 2017), which is inconsistent with "cognitive miserliness" and the absence of deliberation. Such outcomes can arise from task misinterpretation, lack of mindware, or the strategy selected. When relying on responses to judge a process, we cannot know if a participant has reasoned deliberatively unless we presume the task was understood as intended, and we cannot know they understood the task as intended unless we presume they reasoned deliberatively (cf. Smedslund, 1990).

We also note that meta-reasoning studies offer vital insights into individual differences in uncertainty monitoring, facilitating a more nuanced understanding of deliberative processing on a task. For example, when participants determine how long to persevere, they may be optimizing or satisficing, and those of a miserly disposition may simply be looking to bail out through a "computational escape hatch" (Ackerman, Douven, Elqayam, & Teodorescu, 2020; Ball & Quayle, 2000). Low confidence responses after an "impasse" can also decouple the link between response time and deliberative thinking, as can uncertainty about the intended "correct" answer. As such an array of individual-differences measures are necessary to understand the nature of deliberative processes. Furthermore, unpicking such deliberative processes goes beyond the observation of fast and slow thinking. Intuitive processes are always necessary for a participant to respond and sometimes they are sufficient. Participants who understand a task as requiring the alleged system 1 response will not be prompted into deliberation by an awareness of an alleged system 2 response (as this is not necessarily the normative response or the participant's goal).

In sum, we advocate for an approach that follows dual-process model 2.0, but which triangulates task responses with response times, metacognitive measures and individual-difference variables, while aligning with a soft or agnostic view of normative standards. When deciding which responses are the product of deliberative thinking, researchers must be mindful of the myriad individual differences in task interpretations, strategies, and perceived "normative" responses.

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Competing interest. None.

References

Ackerman, R., Douven, I., Elqayam, S., & Teodorescu, K. (2020). Satisficing, metareasoning, and the rationality of further deliberation. In S. Elqayam, I. Douven, J. St. B. T. Evans, & N. Cruz (Eds.), *Logic and uncertainty in the human mind: A tribute to D. E. Over* (pp. 10–26). Routledge.

Ball, L. J. (2013). Microgenetic evidence for the beneficial effects of feedback and practice on belief bias. *Journal of Cognitive Psychology*, *25*(2), 183–191.

Ball, L. J., & Quayle, J. D. (2000). Alternative task construals, computational escape hatches, and dual-system theories of reasoning. *Behavioral & Brain Sciences*, *23*(5), 667–668.

Beeson, N., Stupple, E. J. N., Schofield, M. B., & Staples, P. (2019). Mental models or probabilistic reasoning or both: Reviewing the evidence for and implications of dual-strategy models of deductive reasoning. *Psihologijske Teme*, *28*(1), 21–35.

Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral & Brain Sciences*, *4*(3), 317–331.

Dames, H., Klauer, K. C., & Ragni, M. (2022). The stability of syllogistic reasoning performance over time. *Thinking & Reasoning*, *28*(4), 529–568.

Elqayam, S., & Evans, J. St. B. T. (2011). Subtracting "ought" from "is": Descriptivism versus normativism in the study of human thinking. *Behavioral & Brain Sciences*, *34*(5), 233–248.

Markovits, H., Brisson, J., & de Chantal, P. L. (2017). Logical reasoning versus information processing in the dual-strategy model of reasoning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *43*(1), 72–80.

Oaksford, M., & Chater, N. (2009). Précis of Bayesian rationality: The probabilistic approach to human reasoning. *Behavioral & Brain Sciences*, *32*(1), 69–84.

Smedslund, J. (1990). A critique of Tversky and Kahneman's distinction between fallacy and misunderstanding. *Scandinavian Journal of Psychology*, *31*(2), 110–120.

Stenning, K., & Varga, A. (2018). Several logics for the many things that people do in reasoning. In L. J. Ball & V. A. Thompson (Eds.), *International handbook of thinking and reasoning* (pp. 523–541). Routledge.

Stupple, E. J. N., & Ball, L. J. (2014). The intersection between descriptivism and meliorism in reasoning research: Further proposals in support of "soft normativism." *Frontiers in Psychology*, *5* (1269), 1–13.

Stupple, E. J. N., 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, 1–18.

Verschueren, N., Schaeken, W., & d'Ydewalle, G. (2005). A dual process specification of causal conditional reasoning. *Thinking & Reasoning*, *11*(3), 239–278.

Dual-process theory is Barbapapa

Gustav Tinghög[a,b], Lina Koppel[a] and Daniel Västfjäll[c,d]

[a]JediLab, Department of Management and Engineering, Division of Economics, Linköping University, Linköping, Sweden; [b]Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden; [c]JediLab, Department of Behavioural Sciences and Learning, Division of Psychology, Linköping University, Linköping, Sweden and [d]Decision Research, Eugene, OR, USA
gustav.tinghog@liu.se
lina.koppel@liu.se
daniel.vastfjall@liu.se

doi:10.1017/S0140525X22003211, e144

Abstract

The biggest benefit of dual-process theory lies in its role as a benchmark theory that, regardless of its empirical plausibility, serves as a starting point for better and more domain-specific models. In this sense, dual-process theory is the Barbapapa of psychological theory – a blob-shaped creature that can be reshaped and adapted to fit in the context of any human behavior.

There is much to like about how De Neys analyzes and seeks to advance theories about fast and slow thinking. One is the emphasis on that there is no good–bad or rational–irrational analogy that can be made based on the distinction between intuition and deliberation. Another is the promotion of a non-exclusive view of dual-process theory, where intuition and deliberation do not necessarily need to generate unique responses such that one type of response is exclusively tied to deliberation and is assumed to be beyond the reach of the intuitive system. However, we think that it is important to distinguish the dual purpose of dual-process theories for (1) producing testable predictions and (2) functioning as benchmark theories that, no matter their empirical plausibility, can serve as a starting point for more refined research questions and domain-specific models. We argue that the main benefit of dual-process theories lies in the latter of these two. Thus, dual-process theories have much in common with the fictional character Barbapapa, a blob-shaped creature with the notable ability to shapeshift and thereby smoothly overcome any obstacle. Just like Barbapapa, dual-process theory is liked by many and can easily be reshaped to fit in many contexts.

De Neys argues that a core issue of dual-process theory lies in its exclusivity feature, the assumption that intuition and deliberation should result in unique responses such that one type of response is exclusively tied to deliberation and cannot be reached via intuition. We agree that this exclusivity assumption is not supported by the empirical literature; however, viewing the lack of empirical evidence as a weakness of dual-process theory implicitly assumes that the main purpose of dual-process theory is to provide empirically falsifiable predictions of human decision making and to pinpoint the exact mechanisms that explain why certain behaviors come about. Although making predictions and pinpointing mechanisms are an important ambition when developing further model specifications, we would argue that the perhaps most important function of dual-process theory is in its role as a benchmark theory that, no matter the empirical falsifiability, can serve as an all-embracing framework for thinking about how people process information. In this perspective, dual-process theory does quite well. Arguably a key reason for why dual-process theory has become so popular is that it can easily be reshaped and refined to make sense of the cognitive processes underlying human behavior in specific domains. Thus, we think that dual-process theory has become so influential *thanks to* the fact that it is practically impossible to falsify, not despite it.

In some sense, dual-process theory can be seen as psychological science's equivalent of expected utility theory in economics. Expected utility theory is an all-embracing theory for assessing decision outcomes and can be applied to all contexts and decisions. It is difficult to falsify the claim that a certain chosen behavior maximizes that person's utility, because we typically behave in a way that leads to what we expect to be the most preferred outcome. The key for any model specification is how to define and measure utility. Expected utility theory serves as an intuitive way to organize and make sense of the costs and benefits that people assign to outcomes. However, although expected utility theory focuses on the *outcomes* of decision making, dual-process theory provides a framework for thinking about the *process* of decision making. Dual-process theory thus adds a missing perspective to the outcome-focused framework that economists traditionally work within.

The evolution of dual-process theory has brought economists and psychologists closer together in the quest to improve understanding of human behavior. Still, attempts to merge dual-process theory with utility maximization and general economic models are few. For future developments of dual-process theory it could be worthwhile to start thinking about how the growing literature on belief-based utility (Golman, Loewenstein, Moene, & Zarri, 2016; Grant, Kajii, & Polak, 1998; Loewenstein & Molnar, 2018) relates to and can be incorporated into dual-process theory. These models emphasize that beliefs often fulfill important psychological needs that can influence how people assess information. People hold certain beliefs partly because it makes them feel good, not because they are necessarily correct (Sharot & Sunstein, 2020; Tinghög, Barrafrem, & Västfjäll, 2022). We see the feedback loop presented in De Neys's model as a potential starting point for theoretical work in this direction. In addition, the future research agenda for dual-process theory could benefit from being more inclusive and less narrow-minded in regard to qualitative methods. The workhorse methodological vehicle in the dual-process literature has been behavioral experiments testing hypotheses about the general effect of invoking more intuitive or analytical processing. However, as pointed out by De Neys, dual-process theory runs into problems when it is used as a framework to make predictions about aggregate behavior. To understand individual differences and make sense of contradicting patterns of results, qualitative approaches may be needed to provide deeper insights that cannot be achieved through experiments.

Back to Barbapapa. Who is Barbapapa? He is a blob-shaped creature from a well-known cartoon who tries to fit into the human world. His most notable ability is to shapeshift at will and thereby smoothly overcome any obstacle. After various adventures Barbapapa finally meets Barbamama who has the same notable ability to shapeshift. Together, Barbapapa and Barbamama are able to merge in order to resolve even bigger obstacles. They have seven barbababies who too can shapeshift at will but who have more distinctive individual strengths that can be used to overcome particular obstacles. Barbapapa, Barbamama, and their barbababies are always ready to help and do not fear action. Dual-process theory is the theoretical equivalent of Barbapapa, because it can easily be reshaped to fit the understanding of any human behavior. In the same analogy, expected utility theory is Barbamama, and the barbababies are the more domain-specific dual-process models. Describing dual-process theory in this way highlights its role as a benchmark theory to develop better and more specific models. To conclude, dual-process theory is Barbapapa. The world needs Barbapapa and social science needs dual-process theory.

Competing interest. None.

References

Golman, R., Loewenstein, G., Moene, K. O., & Zarri, L. (2016). The preference for belief consonance. *Journal of Economic Perspectives*, *30*(3), 165–188. doi:10.1257/jep.30.3.165

Grant, S., Kajii, A., & Polak, B. (1998). Intrinsic preference for information. *Journal of Economic Theory*, *83*(2), 233–259. doi:10.1006/jeth.1996.2458

Loewenstein, G., & Molnar, A. (2018). The renaissance of belief-based utility in economics. *Nature Human Behaviour*, *2*(3), 166–167. doi:10.1038/s41562-018-0301-z

Sharot, T., & Sunstein, C. R. (2020). How people decide what they want to know. *Nature Human Behaviour*, *4*(1), 14–19. doi:10.1038/s41562-019-0793-1

Tinghög, G., Barrafrem, K., & Västfjäll, D. (2022). The good, bad and ugly of information (un)processing; *Homo economicus*, *Homo heuristicus* and *Homo ignorans*. *Journal of Economic Psychology*, *94*, 102574. https://doi.org/10.1016/j.joep.2022.102574

A tale of two histories: Dual-system architectures in modular perspective

John Zerilli

Old College, University of Edinburgh, Edinburgh, UK
john.zerilli@ed.ac.uk
https://www.law.ed.ac.uk/people/dr-john-zerilli

doi:10.1017/S0140525X22002977, e145

Abstract

I draw parallels and contrasts between dual-system and modular approaches to cognition, the latter standing to inherit the same problems De Neys identifies regarding the former. Despite these two literatures rarely coming into contact, I provide one example of how he might gain theoretical leverage on the details of his "non-exclusivity" claim by paying closer attention to the modularity debate.

The cleavage between thinking that's fast, intuitive, and stereotyped and thinking that's slow, effortful, and fluid is a defining feature of contemporary dual-system accounts. However, a parallel and largely independent tradition in cognitive science posits domain-specific cognitive systems or "modules" (Chomsky, 1980; Fodor, 1983; Marr, 1976; Mountcastle, 1957, 1978). In the canonical formulation, the existence of modules is thought to hinge on the difference between "central" and "peripheral" operations, where only the latter qualify as modular (Fodor, 1983; cf. Carruthers, 2006; Chomsky, 2018; Sperber, 1994, 2002). Peripheral systems encompass both sensory (input) and motor (output) systems, including those storing procedural knowledge and skill routines. They are characterised by a similar roster of diagnostic features as those commonly ascribed to the fast and intuitive "system 1" within dual-system accounts – in particular, a degree of informational encapsulation, automaticity, and introspective opacity. The main difference is that, with modules being domain-specific, one doesn't encounter an all-purpose "peripheral module," akin to system 1, that's set against the central system/"system 2." Instead, there are at least as many modules as there are input and output systems, and potentially separate modules for acquired skills (Karmiloff-Smith, 1992). Furthermore, being peripheral, the operations of modules map imperfectly onto system 1 functions, with some possible overlap for skills. But even then, in dual-system accounts, the skills in question are more likely to be cognitive biases and rational heuristics – something more like intellectual habits – than perceptuo-motor and procedural skills. Perhaps ironically, the dual-system view has more in common with theories of "massive modularity," in that both view central operations as carved into stereotyped modes of functioning dependent on context (Barrett & Kurzban, 2006). Both dual-system and modular theories are, in turn, distant cousins of the much older physiological division of the nervous system into the central ("voluntary") and peripheral ("autonomic"/"involuntary") nervous systems. According to the physiological classification, brain and spinal cord constitute the central nervous system, meaning that, counterintuitively, modular (peripheral) operations, being largely cortically controlled, fall under the central nervous system, not the peripheral one.

Some philosophers have thought that if peripheral operations are "fast, cheap, and out of control" they will be less vulnerable to epistemically corrosive top-down/doxastic influences (Machery, 2015; Zeimbekis & Raftopoulos, 2015). Indeed, epistemic worries lay partly behind the traditional effort among modularists to show that perception isn't cognitively penetrable – that a visual module, for example, cannot access central information, such as an agent's beliefs and desires, and so operates without interference from what the agent believes or wants the world to be like (Fodor, 1983, 1984). This form of informational encapsulation amounts to a more pronounced form of the system 1/system 2 distinction, albeit pitting perceptuo-motor tasks against system 2. De Neys's non-exclusivity model, for its part, predicts that system 2 responses are available to system 1, itself a highly suggestive claim that runs counter to the modularist's contention about the cognitive impenetrability of perception. For instance, De Neys speculates that "intuitive logical reasoning would serve to calculate a proxy of logical reasoning, but not actual logical reasoning" (target article, sect. 4.2, para. 3). One compelling explanation for this feat is that the brain is able to execute quick, largely involuntary, and *reliable* routines by exploiting some of the same hardware – and information – that runs the slower (more deliberate) routines. If that's true, and generalises to perceptual systems, the epistemic worry would either dissolve (optimistically) or diminish (more likely), because perceptual systems would then still be fast, cheap, and out of control, and hence less vulnerable to interference from central information, despite having access to that information (i.e., being cognitively penetrable). But more importantly for De Neys (and whether or not the idea generalises to perceptual systems), it would offer De Neys a promising source of corroborating detail for his non-exclusivity framework: System 1 might generate system 2 responses efficiently and reliably because it has *access* to system 2 information! As it happens, a proposal along these lines finds support in some of the (anti)modularity literature, which suggests that perceptual systems do have access to central information.

For example, evidence of widespread neural "reuse" or "recycling" demonstrates that the neural communities subserving even our most evolutionally ancient transduction systems also subserve central systems; and it's also likely that transduction dynamics can sometimes be activated by the same domain-general nodes yielding central system dynamics (Anderson, 2010, 2014; Dehaene, 2005). Both findings are significant, because overlapping neural systems are likely to share information (Pessoa, 2016). Further evidence that fast routines can indeed be gotten out of the elements of slower ones comes from research showing that visual processing integrates memories and prior expectations – which feature in slower, classically central, operations – implying that some perceptual processes have access to central information, despite being fast, automatic, and reflex-like (Chanes & Barrett, 2016; Munton, 2022). Take a simple example.

> *Maple Syrup*: A bottle of "Hamptons Maple Syrup" on my kitchen benchtop struck me as "Hampton's Maple Syrup" for quite some time until one day I realised there was no apostrophe. In fact, for some of the time there *was* an apostophe, but it had been expertly occluded by my partner, an amateur lithographer, who gets a kick out of altering labels on household food items when he's bored.

Maple Syrup seems as good an example as any of the cognitive penetration of perceptual experience, and it's the cumulative force of multiple bouts of misremembering what I had previously seen, on top of heavily weighted priors, that plausibly accounts for it. The penetration is fast, automatic, and not readily susceptible to central revision. Crucially, it illustrates that fast and frugal dynamics can sometimes underwrite perceptual fidelity without the added requirement that perception be cognitively impenetrable – after all, there normally *is* an apostrophe on bottles of Hampton's maple syrup! Contextual disambiguations like this are probably ubiquitous (e.g., incorrectly seeing "agnostic" instead of "agonistic" in a context where the former would be more typical, such as an article about religious beliefs in America).

Obviously, De Neys can afford to be noncommital on the epistemic issues surrounding perception. But a fallout from this debate may offer just the lead he needs in gaining a tighter understanding of how his non-exclusivity proposal might work.

Financial support. No funding was received to assist with the preparation of this manuscript. The author has no relevant financial or non-financial interests to disclose.

Competing interest. None.

References

Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioral and Brain Sciences*, *33*, 245–313.
Anderson, M. L. (2014). *After phrenology: Neural reuse and the interactive brain*. MIT Press.
Barrett, H. C., & Kurzban, R. (2006). Modularity in cognition: Framing the debate. *Psychological Review*, *113*, 628–647.
Carruthers, P. (2006). *The architecture of the mind: Massive modularity and the flexibility of thought*. Oxford University Press.
Chanes, L., & Barrett, L. F. (2016). Redefining the role of limbic areas in cortical processing. *Trends in Cognitive Sciences*, *20*, 96–106.
Chomsky, N. (1980). *Rules and representations*. Columbia University Press.
Chomsky, N. (2018). Two notions of modularity. In R. G. de Almeida & L. R. Gleitman (Eds.), *On concepts, modules, and language: Cognitive science at its core* (pp. 25–40). Oxford University Press.
Dehaene, S. (2005). Evolution of human cortical circuits for reading and arithmetic: The "neuronal recycling" hypothesis. In S. Dehaene, J. R. Duhamel, M. D. Hauser, & G. Rizzolatti (Eds.), *From monkey brain to human brain* (pp. 133–157). MIT Press.
Fodor, J. (1983). *The modularity of mind: An essay on faculty psychology*. MIT Press.

Fodor, J. (1984). Observation reconsidered. *Philosophy of Science*, *51*, 23–43.
Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science*. MIT Press.
Machery, E. (2015). Cognitive penetrability: A no-progress report. In J. Zeimbekis & A. Raftopoulos (Eds.), *The cognitive penetrability of perception: New philosophical perspectives* (pp. 59–74). Oxford University Press.
Marr, D. (1976). Early processing of visual information. *Philosophical Transactions of the Royal Society B*, *275*, 483–524.
Mountcastle, V. (1957). Modality and topographic properties of single neurons of cat's somatic sensory cortex. *Journal of Neurophysiology*, *20*, 408–434.
Mountcastle, V. (1978). An organizing principle for cerebral function: The unit module and the distributed system. In G. Edelman & V. B. Mountcastle (Eds.), *The mindful brain* (pp. 7–50). MIT Press.
Munton, J. (2022). How to see invisible objects. *Noûs*, *56*, 343–365.
Pessoa, L. (2016). Beyond disjoint brain networks: Overlapping networks for cognition and emotion. *Behavioral and Brain Sciences*, *39*, 22–24.
Sperber, D. (1994). The modularity of thought and the epidemiology of representations. In L. A. Hirschfield & S. A. Gelman (Eds.), *Mapping the mind* (pp. 39–67). Cambridge University Press.
Sperber, D. (2002). In defense of massive modularity. In I. Dupoux (Ed.), *Language, brain, and cognitive development* (pp. 47–57). MIT Press.
Zeimbekis, J., & Raftopoulos, A. (Ed.) (2015). *The cognitive penetrability of perception: New philosophical perspectives*. Oxford University Press.

Author's Response

Further advancing fast-and-slow theorizing

Wim De Neys

LaPsyDE, CNRS, Université Paris Cité, Paris, France
wim.de-neys@u-paris.fr

doi:10.1017/S0140525X23000559, e146

Abstract

The 34 commentaries on the target article span a broad range of interesting issues. I have organized my reply around five major themes that seemed to emerge: Remarks about the generalizability of the empirical findings, links with other models, necessary extensions, the utility of dual-process models, and more specific points. This allows me to clarify possible misconceptions and identify avenues for further advancement.

R1. Introduction

The target article argued that we need to re-think the popular "fast-and-slow" dual-process model of thinking. I tried to clarify that there is no good support for foundational assumptions concerning the exclusive capacities of the slow-deliberate system and the mechanism that allows us to switch between fast and slow thinking. The paper built on these insights to sketch a more viable dual-process architecture and future research agenda.

Perhaps not surprisingly given the popularity of the dual-process framework, the paper attracted considerable reaction. More than 60 commentators address a breadth of issues in a total of 34 commentaries. I would like to thank all the commentators for their contribution. It is a privilege to get such a large number of experts to take the time to read your work, reflect on it, and voice their concerns. The constructive and balanced commentaries helped me to correct possible misconceptions, clarify key points, and identify further directions in which the working model can be developed. I have no doubts that this will prove instrumental to advance the debate.

My reply is organized around some of the major themes that seemed to emerge in the commentaries: Remarks about the generalizability of the empirical findings (R2), links with other models (R3), necessary extensions (R4), and the utility of dual-process models (R6). Section R5 deals with some more specific points and clarifications. So let's get started.

R2. Empirical generalization

In the first part of the target article I reviewed empirical evidence and showed that there is no solid ground for the exclusivity assumption. I focused on three influential dual-process applications (logical, moral, and prosocial reasoning) in which the claim has been extensively tested. Various commentators note that there is also good evidence outside these fields that supports the exclusivity claim and working model. **Ferreira & Huettig** point to the language processing domain, **Lee & Morewedge** to the anchoring bias, **Petty, Wegener, & Briñol (Petty et al.)** to social psychological work on persuasion, and **Zerilli** to the (anti-)modularity literature. I thank the commentators for highlighting these links. It underscores the generality of the non-exclusivity claim and indicates it is not restricted to the specific domains I happened to review (contra **March, Olson, & Gaertner [March et al.]**). At the same time, two commentaries (March et al.; **Low, Butterfill, & Michael [Low et al.]**) also point to applications where there is allegedly evidence for exclusivity. I comment on these below.

It should be noted that **March et al.** focus on a different issue. I defined exclusivity as the tendency to exclusively tie a specific response to deliberation such that it cannot be generated by the fast-intuitive processing system. March et al. note that work on threat processing indicates that various threat signals are intuitively responded to within a matter of mere milliseconds – which is clearly out of reach of a slow, deliberate reasoning system. Bluntly put, they argue that there are also things the fast system can do that the slow system cannot. I agree but that was not my point. Even in a non-exclusive case, the alleged system 2 intuition and the deliberate system 2 equivalent will not necessarily have the same features (target article, sect. 4.2). The equivalence is situated at the response or outcome level. By definition, the intuitive response will be generated faster than the deliberate one, for example. Consequently, the specific question as to whether or not people can also show a deliberate threat response is orthogonal to the issue I raised. Hence, the evidence March et al. cite does not argue against my non-exclusivity claim as such.

Low et al. argue that there are differences between what they refer to as fast-and-slow mindreading. I do not doubt that this is indeed the case. What matters here is that the cited evidence indicates that (even very young) people have intuitive processes at their disposal that allow them to be successful at mindreading. In line with my key claim, the evidence suggests that perspective-taking and making assumptions about others' beliefs do not necessarily require effortful deliberation. The point that the properties might not completely overlap is not disputed here. Non-exclusivity does not entail that there are no differences between intuitively and deliberately generated responses. In sum, the examples do not argue against my exclusivity claim. If anything, the cited mindreading evidence strengthens it by indicating that mindreading does not necessarily require deliberation. That being said, I do acknowledge that pinpointing the amount of overlap or how the non-exclusive system 1 and 2 responses precisely differ is an important question for the further development of the framework (target article, sect. 4.2; see also R5 and **Handley, Ghasemi, & Bialek [Handley et al.]**).

However, more generally, it is important to underline that I do not argue against exclusivity per se. That is, I do not exclude that there might be certain domains in which exclusivity will be maintained. This is ultimately an empirical question. My key argument concerns the necessary coupling between the exclusivity and switch features (target article, sect. 2.2). If exclusivity is upheld, a dual-process model will have to postulate that reasoners have no internal mechanism that allows them to reliably switch from system 1 to system 2. **Low et al.** do acknowledge this concern and point to processing fluency as a possible alternative mechanism. People would switch to deliberate processing whenever they experience low processing fluency (see also **Kołeczek & Sekerdej**; **Newman & Thompson**). I simply want to warn here that this popular referral to processing fluency begs the question. The key issue for a reliable internal switch mechanism account is to explain why a reasoner experiences low processing fluency in the absence of external cues. For example, the reviewed evidence (target article, sect. 1.1) suggests that reasoners show an intuitive sensitivity to conflict with the alleged system 2 response. They might indeed experience low processing fluency (or a low Feeling of Rightness, Newman & Thompson) in this case. But why would the processing fluency be lowered then? What internal mechanism might account for this conflict sensitivity? To solve this puzzle, we'll ultimately need to introduce some form of non-exclusivity and assume that the alleged system 2 response is also being cued in system 1 so that the conflict with a competing intuition results in decreased processing fluency. Simply positing low processing fluency as switch determinant by itself does not bring us an inch closer to an explanation here.

R3. Link with other models

The target article tried to illustrate how positing non-exclusivity leads to a solution to the switch issue and sketched the building blocks of an elementary model that avoids the conceptual paradoxes of traditional models. Numerous commentators point out that the basic ideas fit well with other approaches. **Stanovich & Toplak** stress how they have also started to integrate non-exclusivity in their thinking. **Petty et al.** point to links with the heuristic-systematic model (Chaiken et al., 1989) and elaboration likelihood model (Petty & Cacioppo, 1986) from the persuasion literature. **Osman** notes that dynamic-value-effort-based decision-making models (e.g., Osman & Wiegmann, 2017) made similar proposals to explain moral reasoning. **Sherman & Klein** note similarities to their quad model (Sherman et al., 2008). **Ackerman & Morsanyi** and **Newman & Thompson** stress links to their work on meta-reasoning (Ackerman & Thompson, 2017; Thompson, 2009). **Braem, Held, Shenhav, & Frömer (Braem et al.)** mention conceptual links to the cognitive control and conflict monitoring literature (e.g., Abrahamse et al., 2016; Botvinick et al., 2001; Lieder et al., 2018) whereas **Oaksford** notes how Bayesian cognitive science models (e.g., Oaksford & Hall, 2016) also capture key features.

These links are no coincidence. I noted how my ideas were inspired by and integrated the recent work of a wide range of authors (target article, sect. 3). In addition, the proposed model was kept simple on purpose and should not require dramatic departures from common knowledge. That is, once one accepts non-exclusivity, the core solution to the switch issue should be fairly self-evident. In this respect, it is important that there is indeed shared support for the basic ideas (e.g., uncertainty monitoring). It lends credence to the working model and underscores that it should not be discarded as an esoteric or idiosyncratic fringe view defended by one single outlier.

Nevertheless, some points need more explicit clarification: In the target article I focused on traditional dual-process models such as they have been put forward in the seminal work of "founding fathers" of the field such as Keith Stanovich, Jonathan Evans, or Daniel Kahneman (e.g., Evans & Stanovich, 2013; Kahneman, 2011; Stanovich & West, 2000). **Stanovich & Toplak** note that they no longer make the exclusivity assumption. This is correct. I disagree with Stanovich & Toplak about the exact timing (see sect. R5) but it is true that in their most recent work, both Stanovich (2018) and Evans (2019) have integrated the emerging empirical evidence and no longer posit exclusivity. I cited these recent papers as inspiration but I could have stressed this point more explicitly. Indeed, it is also important for potential defendants of exclusivity to know that at least some of the founding fathers no longer hold this view.

I also need to clarify that the article is certainly not the first to highlight that traditional dual-process theories have a switch problem (**Newman & Thompson**). Thompson (2009) noted this long ago and I also made this point in my own older work (De Neys & Glumicic, 2008; although the solution I proposed there can – at best – be described as "patchwork"; Pennycook, 2017). The goal of the target article was not to highlight the switch problem per se but rather to illustrate how this problem is intrinsically linked to the exclusivity assumption.

Although I pointed to a range of work that inspired the working model or can be used to extend it (target article, sects. 3 and 4), **Petty et al.** are right that I failed to identify the link with the social psychological work on persuasion. I thank the authors for bringing this to my and the readers' attention. I agree that it is important to be as integrative as possible.

At the same time I think it is also important to notice that despite the common ground, there remain some important differences. For example, I do not believe that a parallel dual-process model can be viable (contra **Petty et al.**; **Sherman & Klein**). I am also not convinced it is wise to postulate that uncertainty monitoring is necessarily affective in nature and that there is a need to distinguish different types of monitoring signals (contra **Newman & Thompson**; **Ackerman & Morsanyi**). I will discuss these issues in more detail in section R5.

However, the key point I wanted to acknowledge and stress in this section is that there is indeed wide overlap and support for many of the postulated core features of the working model. The model is certainly not the first to put these forward.

R4. Extensions

With the working model I tried to present the elementary structure of a viable dual-process framework that addresses the conceptual switch issue. I kept the verbal model as simple and general as possible to provide the minimal architecture that future models in different fields can build on (or as **Cho, Teoh, Cunningham, & Hutcherson [Cho et al.]** put it "to provide a set of domain-agnostic organizing principles"). I have no doubts that the model will need to be extended or revised. I touched upon this issue and sketched some possible directions (target article, sect. 4) but the commentators pointed to various interesting additional extensions. I will try to summarize these and comment on them in this section.

Many commentators (**Ackerman & Morsanyi**; **Cho et al.**; **Kołeczek & Sekerdej**; **Newman & Thompson**; **Oyserman**; **Pennycook**; **Stupple & Ball**) noted that the model does not explicitly incorporate an external system 2 or deliberation engagement mechanism. This is correct. In the target article, I focused on the specification of an internal switching mechanism. The key conceptual problem for traditional dual-process models is to explain how a reasoner can ever reliably determine there is a need to switch to more effortful deliberation in the absence of external cues. If one is explicitly instructed to deliberate or is getting external environmental feedback that indicates that one's intuitive processing is running astray, one can obviously use this as a cue to engage in deliberation. I used the "boy named Sue" expectancy violation example in the target article to illustrate this point. Hence, traditional dual-process models have no conceptual problem when relying on external switching and I consequently did not focus on this case. But clearly, this does not imply I want to claim that there is no possibility of external switching. To paraphrase Oyserman, we need more than an internal thermostat that is turned on and off by an uncertainty threshold. A system that shuts out external cues would not be evolutionary viable indeed. That is, although traditional dual-process models (and the working model) have no conceptual problem to account for external switch cases, it remains important to specify exactly how such external cueing works (e.g., by lowering the deliberation threshold, modulating intuitive strength and increasing uncertainty, etc.). The commentators are right in that it will be important to integrate such an external cueing route in the model. They offer various interesting suggestions. For example, the metacognition and meta-reasoning work that Ackerman & Morsanyi point to has long focused on this case and an integration of these insights should be especially useful. Note that a similar point applies to Cho et al.'s comment about my focus on "bottom-up" reactive control rather than "top-down" proactive control and their suggestion to integrate insights from the sequential sampling modeling literature.

The commentators also rightly point out that it will be important to pay more attention to individual differences (**Ackerman & Morsanyi**; **Baron**; **Pennycook**; **Stupple & Ball**). The working model focuses on the modal or average reasoner. I noted how individuals may differ in the strength of their intuitions but there are numerous additional routes through which individual differences can potentially emerge. For example, thinking dispositions such as the need for cognition or actively open minded thinking (Baron; Pennycook) may affect the height of the deliberation threshold and lead to differential deliberation engagement among individuals who experience a similar level of uncertainty. It should be clear that I do not argue against such possible individual differences. However, the commentators are also right in that the precise mechanisms through which they emerge remain to be explored and specified.

Another general point that surfaced in multiple commentaries was the need to better specify the mechanism that makes us stop deliberating (**Ackerman & Morsanyi**; **Baron**; **Pennycook**; **Sirota**; **Stupple & Ball**). It is indeed important to avoid that we would get stuck in eternal deliberation (target article, sect. 3; Ackerman & Morsanyi; Baron). I postulated that deliberation will operate on the intuition strength and affect the uncertainty parameter through a feedback loop. I also pointed to the possibility of incorporating an opportunity cost factor (target article, sects. 4.4 and 4.6) but did not integrate a specific mechanism yet. Sirota presents numerous interesting suggestions, for example, by having motivation and the costs and benefits of deliberation directly affect the deliberation monitoring threshold. In line with **Fabio & Caprì**, this implies that the model also needs a direct route from deliberation to uncertainty monitoring (rather than an indirect one through a change in intuition activation strength). Ackerman's (2014) diminishing criterion model might also be a good way to formalize my mere hypothetical suggestion that the longer a deliberation process takes, the less we may bother about it. Finally, Pennycook raises an interesting question as to whether the opportunity cost is factored into the decision to engage deliberation or into the extent of deliberation once it has been engaged. In the former case, it will be important to make sure that the cost computation itself can be handled by system 1 (see Sirota and **Braem et al.** on how this may be achieved through caching and learning, e.g., Kool et al., 2018).

As I noted in the target article (sect. 4.6), the commentators also indicated how further development of the framework may benefit from a computational modeling approach using evidence accumulation models (**Braem et al.**; **Cho et al.**; **Hayes, Stephens, & Dunn [Hayes et al.]**). The commentaries clarify how many of the variables I alluded to (e.g., individual differences, opportunity cost, proactive control, etc.) can be captured, specified, and predicted by the parameter settings in the models. **Oaksford**'s point about the potential of a Bayesian modeling approach (e.g., see target article, sect. 4.7) is also relevant in this respect.

Grüning & Krueger make an excellent point about multiple intuitions. I simply assumed that each intuition is identified by the response it cues. Hence, each response is defined by one single intuition. As Grüning & Krueger clarify, there will also be cases in which multiple (conflicting) intuitions cue a single response. It is indeed an open (but testable) question as to how the response strength is integrated (e.g., absolute difference vs. addition) in this case. I thank the authors for pointing this out and believe it nicely illustrates the value of the working model and how it can generate new research questions.

Frankish calls for a further dual-process revision in which system 1 not only initiates system 2 thinking but necessarily also generates it. The idea is that system 2 never generates a response *sui generis*. It always acts upon information that is essentially already present in system 1. Hence, the building blocks of what we deliberate about are always already available in system 1, system 2 merely recombines them. I believe that under this view it would make no sense to argue that deliberation would lead to the generation of a truly "new" response per se. I can see where the conceptual idea is going but it is at this stage not yet clear to me how we would proceed to test it empirically.

Last but not least, **Osman** notes that it will be important to extend the working model to address reasoning about more pressing societal, political, and cultural challenges. There is indeed a tendency in the dual-process field to focus on somewhat artificial "toy" problems (Bonnefon, 2017). Although these are interesting to study thinking in a controlled setting it is also clear that in daily life we typically do not reason about the cost of a bat and a ball or whether we should push a fat man off a footbridge. Focusing on informal argumentation and communication (e.g., **Oaksford**; Hahn, 2020; Mercier, 2021) or attempts to use a dual-process approach to study science misperception and misinformation spreading (e.g., Pennycook [2023], for an excellent overview) might prove especially relevant here. I agree that this remains an important challenge for the framework I presented.

R5. Specific clarifications and points

The empirical evidence I reviewed and the working model mainly focus on situations in which people are faced with two conflicting intuitions. Several commentators (**Feiman**; **Kołeczek & Sekerdej**; **Frankish**) stress that this should not be considered a paradigmatic case. For example, there are also situations in which more than two, only one, or no intuitions will be cued. I agree and also accounted for this in the target article (sects. 3.2 and 4.5). Frankish does offer an interesting theoretical clarification: The situation in which no intuition is being cued presents by definition a case in which exclusivity is upheld. That is, if system 1 does not cue a response, we will be forced to engage in deliberation to arrive at an answer. This is correct. In unfamiliar domains in which people have not had a chance to automatize the system 2 response, exclusivity will be maintained.

In reply to **Dewey**'s call for a weaker formulation of non-exclusivity it should be stressed that I do not posit a strong categorical classification. That is, non-exclusivity implies that the alleged system 2 response *can* be generated by system 1. This does not imply it always will be (e.g., individual strength differences, target article, sect. 3.2). This fits with the empirical observation in the logical reasoning field that despite the possible intuitive generation of the alleged system 2 response, correct answers still tend to be slightly more likely after deliberation (e.g., Bago & De Neys, 2017; target article, sect. 1.2).

The working model explains how reliable internal switching can occur. However, this does not mean that the mechanism cannot fail (contra **Ackerman & Morsanyi**; **Kołeczek & Sekerdej**; **Handley et al.**). I do not contest that there will be false-negative or -positive cases in which people do not engage in deliberation when they should or deliberate when there is no need for it. For example, a typical false negative during logical reasoning might be a situation in which the alleged system 2 intuition is so weak that the uncertainty threshold is not reached and people will stick to a dominant biased intuitive response without having engaged in deliberation. The point is that with the working model the machinery to explain successful switching is in place, not that switching will always be successful.

Lee & Morewedge illustrate how a noise manipulation in anchoring bias research (e.g., Lee & Morewedge, 2022) can be used to test the predictions of the working model. They note that more uncertainty (e.g., as operationalized by considering a wider range of possible response values) does not lead to more deliberation as measured by response accuracy. However, it will be important to look beyond response accuracy as deliberation index here (e.g., Thompson et al., 2011). That is, the fact that people did not correct their answer does not imply that they did not engage in deliberation. As in other reasoning tasks, they might have used deliberation to rationalize the biased response (Pennycook et al., 2015). Hence, to test the working model we will also need to take the anchoring response latencies into account, for example.

Stanovich & Toplak note that they have postulated non-exclusivity at least since 2000. As I stressed above (R3), both Stanovich (2018) and Evans (2019) have integrated non-exclusivity in their recent publications. However, I'm puzzled by Stanovich & Toplak's claim that they gave up on exclusivity so long ago.[1] Consider, for example, Stanovich and West's (2000) seminal individual differences paper. In my reading, it was precisely because correct responding on logical reasoning tasks was assumed to require demanding deliberation that people higher in cognitive capacity were expected to show higher accuracy. Likewise, Evans and Stanovich (2013; see also their reply to Osman, 2013) extensively discuss how load effects as in the De Neys (2006a, 2006b) studies indicate that correct responding necessitates demanding type 2 thinking. In addition, the actual direct empirical evidence that points to the intuitive cueing of the alleged system 2 response is fairly recent. The evidence only started amassing in the last decade (e.g., De Neys, 2012) and meta-analyses were not published until 2019 (e.g., Kvarven et al., 2020; Rand, 2019). My own older dual-process work (De Neys, 2006a) certainly assumed exclusivity, for example. Note also that as Stanovich & Toplak write, automatization is indeed a core principle of traditional dual-process theories (De Neys, 2012, 2014). However, this logical automatization was typically conceived as a mechanism that might impact the logical reasoning of the rare expert and not as a mechanism that would account for the inferencing of the modal reasoner (De Neys, 2012). Nevertheless, the critical point is that Stanovich & Toplak are on board with the key claim in the target article and do not defend exclusivity.

Contrary to what **De Houwer, Boddez, & Van Dessel (De Houwer et al.)** write, strictly speaking, I do not (or no longer, e.g., De Neys & Glumicic, 2008) favor the low effort deliberation switch model. This model (e.g., De Neys & Glumicic, 2008; Kahneman, 2011) assumes that the monitoring of system 1 requires low effort deliberation within system 2. The working model assumes that the monitoring results from effortless system 1 processing. That being said, given that I have no strict demarcating definition of intuitive and deliberate processing – as De Houwer et al. correctly point out – the difference is purely theoretical. In this respect, I do see the merit of taking the descriptive level and our subjective experiences into account and would not object to the integration of this criterion into the definition of what we call intuition and deliberation.

Contra **Petty et al.** and **Sherman & Klein** I do not believe that dual-process models need or should posit a parallel processing architecture (target article, sects. 2.1.4 and 3.5). Clearly, in practice, thinking will typically involve a dynamic interaction between systems 1 and 2 (see target article, sect. 3.4). I do also not deny that a reasoning process can be initiated by system 2 (R4). However, an organism that would continuously spend costly and limited resources in an environment where there is often an accurate, fast, and effortless alternative available would be at a serious evolutionary disadvantage in comparison to a serial competitor. In general, to put it bluntly, if the parallel views of human cognition were to be correct, there would be no need for research on effort or control allocation (**Braem et al.**; **Cho et al.**; **Sirota**), as humans would simply spend effort or control continuously. At the very least it is clear that we do not need a parallel processing architecture to account for the specific empirical evidence that I reviewed in the target article (e.g., see De Neys, 2017, for an extensive discussion).

Greene is largely sympathetic to the working model and agrees that utilitarian decisions do not necessarily require deliberation. I do not make claims about the emotional nature of intuitive processing but simply want to clarify that the working model is consistent with Greene's suggestion that intuitive and deliberate utilitarian decisions can still differ. Even in a non-exclusive case, the alleged system 2 intuition and the deliberate system 2

https://doi.org/10.1017/S0140525X2200142X Published online by Cambridge University Press

equivalent will not necessarily have the same features (target article, sect. 4.2). For example, I noted how Bago and De Neys (2019) already observed that intuitive utilitarian responses were less frequently explicitly justified than utilitarian responses that were given after deliberation (see also my comments on **Handley et al.** below).

To answer a specific question raised by **Newman & Thompson**: I do assume that people experience the uncertainty monitoring signal consciously. People will not necessarily know why they are uncertain about a certain decision, but they will explicitly register and be able to report (e.g., confidence rating) their uncertainty (see also De Neys, 2014). The empirical evidence is clearly consistent with this position. However, contra Newman & Thompson and their "feeling of rightness" account, I do not posit that the uncertainty signal is necessarily affective in nature. There will be cases in which the uncertainty signal can become affectively loaded (**Ainslie**; **Kurth**) but given the empirical evidence I see no reason why we need to postulate an additional and necessary underlying affective mechanism here. People can obviously be affectively aroused by uncertainty but they do not need to be to arrive at system 2 engagement. Relatedly, Newman & Thompson note that the working model's monitoring mechanism needs to distinguish different types of monitoring signals (e.g., the feeling of rightness, the feeling of wrongness, the feeling of error, etc.). Although I see value in the subjective level (**De Houwer et al.**) I am not sure that it is a good idea to a priori postulate a multitude of "feelings" in the absence of empirical data. Because feeling right and wrong about something might "feel" different does not imply we necessarily need to postulate additional machinery. We are continuously adding additional parameters to the model here. Theoretically, the working model's basic architecture suffices to capture these cases. For example, when I am very certain that response A is wrong, this can be modeled as response A having a low activation strength, consequently the resulting uncertainty will be high, and people will start deliberating (e.g., "I know the Capital of Australia is not famous Sydney but what is it again then?"). Being very certain that a response is right can be modeled as it having a high activation strength (and the resulting uncertainty monitoring signal will be low), and so on. Hence, the simple working model is more parsimonious. However, ultimately, this is an empirical discussion. If Newman & Thompson present evidence that a "rightness," "wrongness," "error," or other subjective feeling index results in or predicts differential system 2 engagement (e.g., answer change or re-thinking time in the two-response paradigm), I will be happy to revise my views.

Finally, **Handley et al.** make a critical point about the extent of the overlap between the intuitively generated alleged system 2 response and the proper deliberate system 2 response (i.e., a "logical" intuition and a deliberate logical answer in the logical reasoning domain that they focus on). Non-exclusivity entails that the alleged system 2 response can be generated by system 1 although this intuitive response does not necessarily have the same features. But to what extent are they similar or different? Handley et al. illustrate how one option is that there is no overlap and people are right for the wrong reasons. For example, in the logical reasoning domain, people's intuitive response could be based on superficial surface features (e.g., whether information is repeated in the premises) that have no intrinsic or epistemological link to logical validity. Alternatively, there might be some minimal overlap. For example, as Handley et al. suggest, the intuitive logical response might be based on the probabilistic strength of an argument. During deliberation people may still take other considerations into account but in this case the intuitive process would respect the probabilistic structure of the environment and thereby be linked to validity. People might get it intuitively right in this case for different, but not for wrong reasons.

It is perhaps useful to illustrate the core issue with a more general example from the moral reasoning domain. A no-overlap scenario might be a case in which an intuitive utilitarian response is simply based on the psychopathic pleasure to kill (e.g., "killing more is better"). In this case, one might make a fast utilitarian decision (e.g., "kill 5 instead of 1") but there would be no overlap with the critical utilitarian considerations that are deliberately taken into account (i.e., the saving more lives aspect). In the overlap case, the fast intuitive utilitarian response could simply urge people to only take the numbers of saved lives into account (e.g., "saving more is better"). This may still differ from a more careful weighting and integration that people achieve during deliberation (e.g., "killing a person is unfortunate but given we can save more lives it is nevertheless acceptable") but it at least focuses attention on the critical utilitarian principle in question.

I clearly favor the overlap position. I do not think people are moral or rational psychopaths. I believe it is unlikely that across the wide range of domains I reviewed, people will have accidentally picked up on a cue that co-varies with the system 2 response but has no further association with it. The alternative view seems more parsimonious to me. Because people have been repeatedly exposed to the system 2 response, they will have automatized it to some extent. The fast, intuitive response generation might not have all the features of the deliberate response but the automatization should have guaranteed at least some minimal overlap. However, ultimately this is an empirical question. Like most of the commentators, I believe the empirical evidence clearly calls for non-exclusivity and indicates that reasoners can generate the alleged system 2 response intuitively. At the same time, we all agree that pinpointing the precise nature and features of this system 1 intuition remains an important challenge (target article, sect. 4.2).

R6. Dual schmosses?

Inevitably, various commentators also commented on the role of dual-process theory more broadly and the question whether we should abandon the framework altogether. Perhaps not surprisingly given the long history of the debate, whereas some commentators (**Ainslie**; **Hayes et al.**; **Melnikoff & Bargh**; **Oaksford**) strongly oppose the dual-process framework, others fiercely support it (**Shiffrin, Schneider, & Logan [Shiffrin et al.]**; **Tinghög, Koppel, & Västfjäll [Tinghög et al.]**). I have discussed my views on this wider debate in the target article (sect. 4.8) and elsewhere (De Neys, 2021) so I will keep my replies short and focus on some important clarifications.

I should explicitly stress that the working model does not postulate qualitative differences (contrary to what **Oaksford** and **Braem et al.** may seem to assume). I do not believe the quantitative–qualitative question is tractable or consequential (De Neys, 2021; Dewey, 2021). The question is really orthogonal to the non-exclusivity and switch issues that both single- and dual-process models need to address.

Contrary to what **Shiffrin et al.** seem to assume, the target article does not argue against dual-process theory or does not call to abandon it. The paper tries to correct conceptual unclarities and build a more credible version of the framework. To be clear, I do agree with dual-process critics (**Melnikoff & Bargh**)

https://doi.org/10.1017/S0140525X2200142X Published online by Cambridge University Press

that there is no empirical evidence to argue for a qualitative rather than a quantitative view on the difference between "system 1" and "system 2" processing. I also agree with the critics that we don't necessarily need the dual-process framework per se to study thinking (Melnikoff & Bargh). But I do not agree with the critics that the dual-process framework hampers studying thinking and would thwart scientific progress (Melnikoff & Bargh, 2018). Much to the contrary, as **Tinghög et al.** and **Sherman & Klein**, I find the framework a very valuable heuristic that presents us with a great tool to communicate and organize our thinking about human (and machine, e.g., Bonnefon & Rahwan, 2020) cognition. However, although I agree with Tinghög et al.[2] that the ultimate benefit of the dual-process framework lies in its role as a benchmark or meta-theory (see Evans & Stanovich, 2013), I also believe that our meta-theories should be viable and avoid conceptual paradoxes or homunculi. It is here that the critical contribution of the target article lies.

Notes

1. It might be that they refer to exclusivity as implying that system 2 can also lead to logical bias in classic reasoning tasks – which was indeed never disputed (see my discussion of the application context, target article, sect. 1.1.1 or De Neys, 2020). But the exclusivity question at stake here is whether system 1 also generates the correct response.
2. Extra bonus points for the "Barbapapa" analogy.

References

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

Ackerman, R. (2014). The diminishing criterion model for metacognitive regulation of time investment. *Journal of Experimental Psychology: General*, *143*, 1349.

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 intuitive greater good: Testing the corrective dual process model of moral cognition. *Journal of Experimental Psychology: General*, *48*, 1782–1801.

Bonnefon, J. F. (2017). *Reasoning unbound: Thinking about morality, delusion and democracy*. Springer.

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

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*, 624–652.

Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic information processing within and beyond the persuasion context. In J. S. Uleman & J.A. Bargh (Eds.), *Unintended thought* (pp. 212–252). Guilford Press.

De Neys, W. (2006a). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science*, *17*, 428–433.

De Neys, W. (2006b). Automatic-heuristic and executive-analytic processing in reasoning: Chronometric and dual task considerations. *Quarterly Journal of Experimental Psychology*, *59*, 1070–1100.

De Neys, W. (2012). Bias and conflict a case for logical intuitions. *Perspectives on Psychological Science*, *7*, 28–38.

De Neys, W. (2014). Conflict detection, dual processes, and logical intuitions: Some clarifications. *Thinking & Reasoning*, *20*, 169–187.

De Neys, W. (Ed.) (2017). *Dual process theory 2.0*. Routledge.

De Neys, W. (2020). Morality, normativity, and the good system 2 fallacy. *Diametros*, *17*(64), 90–95. https://doi.org/10.33392/diam.1447

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

De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition*, *106*, 1248–1299.

Dewey, A. R. (2021). Reframing single- and dual-process theories as cognitive models: Commentary on De Neys (2021). *Perspectives on Psychological Science*, *16*, 1428–1431.

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

Hahn, U. (2020). Argument quality in real world argumentation. *Trends in Cognitive Science*, *24*, 363–374.

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

Kool, W., Gershman, S. J., & Cushman, F. A. (2018). Planning complexity registers as a cost in metacontrol. *Journal of Cognitive Neuroscience*, *30*, 1391–1404.

Kvarven, A., Strømland, E., Wollbrant, C., Andersson, D., Johannesson, M., Tinghög, G., Västfjäll, D., & Myrseth, K. O. R. (2020). The intuitive cooperation hypothesis revisited: A meta-analytic examination of effect size and between-study heterogeneity. *Journal of the Economic Science Association*, *6*, 26–42.

Lee, C. Y., & Morewedge, C. K. (2022). Noise increases anchoring effects. *Psychological Science*, *33*, 60–75.

Lieder, F., Shenhav, A., Musslick, S., & Griffiths, T. L. (2018). Rational metareasoning and the plasticity of cognitive control. *PLoS Computational Biology*, *14*, e1006043.

Melnikoff, D. E., & Bargh, J. A. (2018). The mythical number two. *Trends in Cognitive Sciences*, *22*, 280–293.

Mercier, H. (2021). How good are we at evaluating communicated information? *Royal Institute of Philosophy Supplement*, *89*, 257–272.

Oaksford, M., & Hall, S. (2016). On the source of human irrationality. *Trends in Cognitive Sciences*, *20*, 336–344.

Osman, M. (2013). A case study dual-process theories of higher cognition – Commentary on Evans & Stanovich (2013). *Perspectives on Psychological Science*, *8*, 248–252.

Osman, M., & Wiegmann, A. (2017). Explaining moral behavior: A minimal moral model. *Experimental Psychology*, *64*, 68–81.

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). Routledge.

Pennycook, G. (2023). A framework for understanding reasoning errors: From fake news to climate change and beyond. *Advances in Experimental Social Psychology*, *67*, 131–208.

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.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology*, *19*, 123–205.

Rand, D. G. (2019). Intuition, deliberation, and cooperation: Further meta-analytic evidence from 91 experiments on pure cooperation. Available at SSRN 3390018. Retrieved from http://doi.org/10.2139/ssrn.3390018

Sherman, J. W., Gawronski, B., Gonsalkorale, K., Hugenberg, K., Allen, T. J., & Groom, C. J. (2008). The self-regulation of automatic associations and behavioral impulses. *Psychological Review*, *115*, 314–335.

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

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences*, *23*, 645–665.

Thompson, V. A. (2009). Dual-process theories: A metacognitive perspective. In J. S. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 171–195). Oxford University Press.

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

https://doi.org/10.1017/S0140525X2200142X Published online by Cambridge University Press