

# FURTHER ADVANCING FAST-AND-SLOW THEORIZING

Wim De Neys

Université Paris Cité & CNRS, LaPsyDE, France

Replies to commentaries on Target Article “Advancing Theorizing about Fast-and-Slow Thinking”

(In Press). Behavioral and Brain Sciences.

## Wordcount

Abstract: 64

Main text: 5803

References: 892

## Corresponding author :

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

Email: [wim.de-neys@u-paris.fr](mailto:wim.de-neys@u-paris.fr)

## **ABSTRACT**

The 34 commentaries on the target paper 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.

## INTRODUCTION

The target paper 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. The 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 (R1), links with other models (R2), necessary extensions (R3), and the utility of dual process models (R5). Section R4 deals with some more specific points and clarifications. So let’s get started.

### R1. 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 and Huettig** point to the language processing domain, **Lee and Morewedge** to the anchoring bias, **Petty, Wegener, & Briñol** 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, and Gaertner**). At the same time, two commentaries (**March, Olson, and Gaertner**; **Low, Butterfill, & Michael**) also point to applications where there is allegedly evidence for exclusivity. I comment on these below.

It should be noted that **March, Olson, and Gaertner** 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, Olson, and Gaertner** 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 section 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, Olson, and Gaertner** cite does not argue against my non-exclusivity claim as such.

**Low, Butterfill, and Michael** 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 section 4.2; see also R4 and **Handley, Ghasemi, & Bialek**).

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 section 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, Butterfill, and Michael** 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 **Koleczek & 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 section 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 and 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.

## **R2. 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 and Toplak** stress how they have also started to integrate non-exclusivity in their thinking. **Petty, Wegener, and Briñol** 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 and Klein** note similarities to their Quad model (Sherman et al., 2008). **Ackerman and Morsanyi** and **Newman and Thompson** stress links to their work on meta-reasoning (Ackerman & Thompson, 2017; Thompson, 2009). **Braem, Held, Shenhav, and Frömer** 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) while **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 section 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., Kahneman, 2011; Evans & Stanovich, 2013; Stanovich & West, 2000). **Stanovich and Toplak** note that they no longer make the exclusivity assumption. This is correct. I disagree with **Stanovich and Toplak** about the exact timing (see section R4) 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 paper is certainly not the first to highlight that traditional dual process theories have a switch problem (**Newman and 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 paper 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 sections 3 and 4), **Petty, Wegener, and Briñol** 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, Wegener, and Briñol; Sherman and 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 and Thompson; Ackerman & Morsanyi**). I will discuss these issues in more detail in section R4.

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.

### **R3. 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** 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 section 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, Teoh, Cunningham, & Hutcherson; Koleczek & Sekerdej; Newman and Thompson; Stupple & Ball; Oyserman; Pennycook**) noted that the model does not explicitly incorporate an external System 2 or deliberation engagement mechanism. This is correct. In the target paper 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 and 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, Teoh, Cunningham, and Hutcherson’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 section 3; **Ackerman & Morsanyi; Baron**). I postulated that deliberation will operate on the intuition strength to affect the uncertainty parameter through a feedback loop. I also pointed to the possibility of incorporating an opportunity cost factor (Target Article section 4.3 and 4.5) 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 and Capri**, 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, Held, Shenhav, and Frömer** on how this may be achieved through caching and learning, e.g., Kool et al., 2018).

As I noted in the target paper (section 4.7), the commentators also indicated how further development of the framework may benefit from a computational modeling approach using evidence accumulation models (**Braem, Held, Shenhav, and Frömer; Cho, Teoh, Cunningham, and Hutcherson; Hayes, Stephens, & Dunn**). 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 section 4.6) is also relevant in this respect.

**Grüning and 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 and 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, in press, for an excellent overview) might prove especially relevant here. I agree that this remains an important challenge for the framework I presented.

#### **R4. 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, Koleczek & 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 (Target Article Sections 3.2 and 4.4). **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 Section 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 Section 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; Koleczek & Sekerdej; Handley, Ghasemi, & Bialek**). 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 and 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 and Toplak** note that they have postulated non-exclusivity at least since 2000. As I stressed above (R2), both Stanovich (2018) and Evans (2019) have integrated non-exclusivity in their recent publications. However, I'm puzzled by Stanovich and Toplak's claim that they gave up on exclusivity so long ago<sup>1</sup>. 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

---

<sup>1</sup> 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 section 1.1.1 or De Neys, 2020). But the exclusivity question at stake here is whether System 1 also generates the correct response.

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 and 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 and Toplak** are on board with the key claim in the target article and do not defend exclusivity.

Contrary to what **De Houwer, Boddez, and Van Dessel** 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, Boddez, and Van Dessel** 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, Wegener, and Briñol** and **Sherman and Klein** I do not believe that dual process models need or should posit a parallel processing architecture (Target Article sections 2.1.4 and 3.5). Clearly, in practice, thinking will typically involve a dynamic interaction between System 1 and 2 (see Target Article section 3.4). I do also not deny that a reasoning process can be initiated by System 2 (R3). 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 view of human cognition were to be correct, there would be no need for research on effort or control allocation (**Braem, Held, Shenhav, and Frömer; Cho, Teoh, Cunningham, and Hutcherson; 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 equivalent will not necessarily have the same features (Target Article section 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, Ghasemi, and Bialek** below).

To answer a specific question raised by **Newman and 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 and 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 and 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, Boddez, and Van Dessel**) 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), etc. Hence, the simple working model is more parsimonious. However, ultimately, this is an empirical discussion. If Newman and 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, Ghasemi, and Bialek** 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, Ghasemi, and Bialek** 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, Ghasemi, and Bialek** 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 to 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 section 4.2).

## **R5. DUAL SCHMOSES?**

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, Stephens, and Dunn; Melnikoff & Bargh; Oaksford**) strongly oppose the dual process framework, others fiercely support it (**Shiffrin, Schneider, & Logan; Tinghög, Koppel, & Västfjäll**). I have discussed my views on this wider debate in the target paper (section 4.7) 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, Held, Shenhav, and Frömer** may seem to assume). I do not believe the quantitative-qualitative question is tractable or consequential (Dewey, 2021; De Neys, 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, Schneider, and Logan** seem to assume, the target paper does not argue against dual process theory or does not call to abandon it. The paper tries to correct conceptual unclarity and build a more credible version of the framework. To be clear, I do agree with dual process critics (**Melnikoff & Bargh**) 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, Koppel, and Västfjäll** and **Sherman and 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, Koppel, and Västfjäll**<sup>2</sup> 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.

## REFERENCES

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

---

<sup>2</sup> Extra bonus points for the “Barbapapa” analogy

- 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. (2019). 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. (2017). Fast logic?: Examining the time course assumption of dual process theory. *Cognition*, *158*, 90–109.
- 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.
- Dewey, C. (2021). Reframing single- and dual-process theories as cognitive models: Commentary on De Neys (2021). *Perspectives on Psychological Science*, *16*, 1428–1431.
- 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*. Oxon, UK: Routledge.
- De Neys, W. (2020). Morality, normativity, and the good system 2 fallacy. *Diametros*, *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., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition*, *106*, 1248–1299.
- 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. St. 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*. New York, NY: 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. (in press). A framework for understanding reasoning errors: From fake news to climate change and beyond. *Advances in Experimental Social Psychology*.
- Pennycook, G. (2017). A perspective on the theoretical foundation of dual-process models. In W. De Neys (Ed.), *Dual Process Theory 2.0* (pp. 5-27). Oxon, UK: Routledge.



- Pennycook, G., 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://dx.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.