Computational models of dual system reasoning

Zoe A. Purcell[1], Jérémie Beucler[1], Wim De Neys[1] & Kobe Desender[2]

1. Université Paris Cité, France
2. KU Leuven, Belgium

Abstract

Dual system theories propose that human reasoning arises from two interacting systems: a fast, automatic process (System 1) and a slower, deliberative process (System 2; Kahneman, 2011; De Neys, 2022). Despite their influence across domains, including economics and moral psychology, these accounts remain largely verbal and underspecified. Existing computational models of reasoning tend to be narrow in scope, lack rigorous fitting, and inadequately capture interactions between System 1 and System 2—particularly how and when deliberative processes are triggered. We develop six computational models formalizing dual system assumptions, including mechanisms such as inhibition and metacognitive monitoring. Preliminary results suggest that intuitive and deliberative reasoning may rely on similar underlying processes.

Keywords: dual system reasoning; bat and ball; base rate; reasoning; confidence

While the functional mapping of processes, like inhibition or working memory, to each System varies across theories, a consistent operational distinction has emerged: System 1 responses are feasible under time pressure and dual-task conditions; System 2 responses are not. Thus we are able to operationally distinguish between the Systems and test, for example, whether a decision was made intuitively or deliberatively. However, whether these are associated with unique functions, e.g., inhibition, remains difficult to determine from standard behavioural methods.

In addition to distinguishing between fast and slow processes, recent dual system accounts emphasize the functional role of metacognitive confidence. Confidence is assumed to reflect the *relative* activation strength of competing solutions with smaller differences leading to lower confidence and prompting further deliberation and larger differences yielding higher confidence and faster responses. Many contemporary dual system theories converge on this "confidence-deliberation hypothesis" (De Neys, 2018; Purcell et al., 2022), yet – as for the functions underlying each System – testing this assumption using behavioral methods remains challenging.

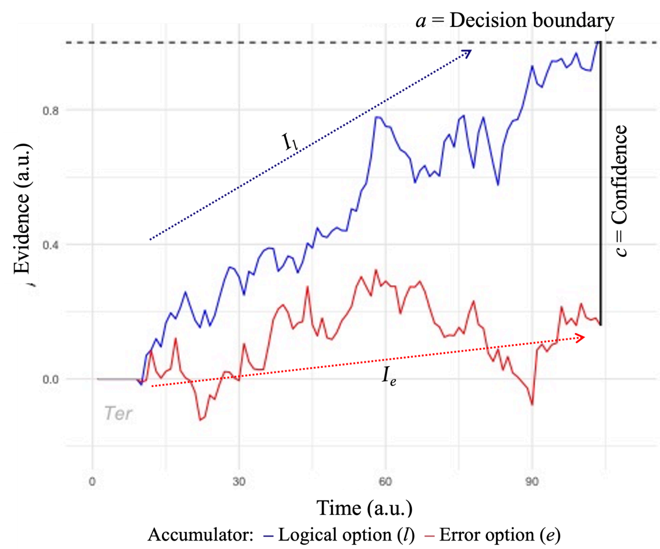

Fig 1. Example of a decision model where the first of two accumulators (logical, error) to reach the boundary *a* wins.

To test predictions about the features and interaction of Systems 1 and 2, we developed six computational models grounded in evidence accumulation frameworks. Evidence accumulation theories describe decisions as a process of gathering evidence over time until a threshold is reached (Usher, 2001; De Martino et al., 2013; Desender et al., 2021; Kiani et al., 2014). Among these, race models propose that each response option (e.g., a logical versus erroneous response to a reasoning problem) is associated with an independent accumulator, and the first to reach a decision threshold determines the choice. This formulation accommodates key features such as confidence, inhibition, and independent evidence strengths for competing solutions (see Fig 1).

Box 1: Default Model

In this "default" model, the competitor terms (w_1, w_2) and the drift terms (I_1, I_2) remain unchanged from speeded at Step 1 to unspeeded decisions at Step 2. This model reflects earlier decision making models of lower-level perceptual decisions, in particular, the leaky, competitor, accumulation model (Usher, 2001).

Step 1:

$$\begin{cases} dy_e = (-w_e y_l + I_e)dt + cdW, & y_e(0) = 0, \\ dy_l = (-w_l y_e + I_l)dt + cdW, & y_l(0) = 0 \end{cases} \tag{1}$$

The process continues until one of the accumulators, $y_1(t)$ or $y_2(t)$, reaches the first decision boundary a_1. Mathematically, this can be written as:

$$\text{Stop when: } \max\big(y_e(t), y_l(t)\big) \geq a_1.$$

At this point (t), confidence is defined as the distance between the two accumulators at the point when the first is made.

Confidence is defined as:

$$\text{Confidence} = |y_e(t) - y_l(t)|\,.$$

The decision process continues (unspeeded) as follows:

Step 2:

$$\begin{cases} dy_e = (-w_e y_l + I_e)dt + cdW, \\ dy_l = (-w_l y_e + I_l)dt + cdW. \end{cases} \tag{2}$$

The process continues until one of the accumulators, $y_e(t)$ or $y_l(t)$, reaches the second decision boundary a_2. Mathematically, this can be written as:

$$\text{Stop when: } \max\big(y_e(t), y_l(t)\big) \geq a_2.$$

We implemented dual system features within the race model framework and tested six variants on behavioral data from studies using the two-response paradigm (Bago & De Neys, 2017; Thompson et al., 2011). Two-response paradigms build from the operational distinction between intuitive and deliberative thinking. Participants are presented with reasoning problems twice: first, under a strict time limit (*t*) to encourage intuitive processing (Step 1) and second, without any restrictions to allow deliberation (Step 2). After each response, participants indicate their confidence in the correctness of their answer. We used this design to examine whether, in line with dual system theory, functions such as inhibition differed at each Step and whether those differences were associated with confidence.

Box 2. Confidence Switch Models

In several models, in line with dual system theory, the parameters were allowed to change from Step 1 to 2. For some of these, that change was dependent on confidence (as defined in Box 1). For example, in the "confidence, boundary shift model", Steps 1 and 2 are defined as in equations (1) and (2), however, the boundary changes such that:

$$\text{Stop when: } \max\big(y_e(t), y_l(t)\big) \geq a_2'.$$

Where $a_2' = \beta_0 + \beta_1 conf$

We developed six two-step models to capture the differences between intuitive (Step 1) and deliberative (Step 2) thinking. The first, default, model is described in Box 1. Here, the process for Step 1 and 2 are identical in that they are governed by the same parameters. In contrast, the second group of models allowed parameters (*w* controlling inhibition, *I* controlling rate of evidence accumulation) to change from Step 1 to Step 2, capturing the core dual system assumptions. The third group of models also allowed parameter changes (*w, I,* and additionally, the boundary *a*) but, critically, these changes were functions of confidence at Step 1 (e.g., Box 2). This group captured the core assumptions as well as the confidence-deliberation hypothesis.

During model fitting, model comparisons based on Bayesian Information Criterion (BIC) consistently favored the default and confidence boundary models (see Boxes 1 and 2). The default model yielded the lowest mean BIC and accounted for around 70% of individual best fits, capturing key patterns in reaction time, accuracy, and confidence across conflict conditions. However, the confidence boundary model provided a better fit for a minority of subjects. So far, these results imply that the same underlying mechanisms may be responsible for intuitive and deliberative processes. However, parameter recovery and tests with alternative tasks are ongoing.

By transforming verbal dual system hypotheses into mathematical functions, we formalize the dual system theories and, as such, are able to examine the cognitive mechanisms proposed to underlie these theories which were previously untestable. We compare the ability of competing models to account for behavioural data, evaluating their ability to capture patterns of confidence, reaction times, accuracy and answer change. Therein, we assess whether the mechanisms behind each System are in fact unique and whether confidence mediates their interaction. This work significantly advances the formalization and specificity of dual system theories, offering new insights into the cognitive mechanisms underlying human reasoning.

References

Bago, B., & De Neys, W. (2017). Fast logic?: Examining the time course assumption of dual process theory. Cognition, 158, 90–109. https://doi.org/10.1016/j.cognition.2016.10.014

De Martino, B., Fleming, S. M., Garrett, N., & Dolan, R. J. (2013). Confidence in value-based choice. Nature Neuroscience, 16(1), Article 1. https://doi.org/10.1038/nn.3279

De Neys, W. (2018). Dual Process Theory 2.0. Routledge. https://doi.org/10.4324/9781315204550

De Neys, W. (2022). Advancing theorizing about fast-and-slow thinking. Behavioral and Brain Sciences, 1–68. https://doi.org/10.1017/S0140525X2200142X

Desender, K., Ridderinkhof, K. R., & Murphy, P. R. (2021). Understanding neural signals of post-decisional performance monitoring: An integrative review. eLife, 10, e67556. https://doi.org/10.7554/eLife.67556

Kahneman, D. (2012). Thinking, fast and slow. Penguin Books.

Kiani, R., Corthell, L., & Shadlen, M. N. (2014). Choice Certainty Is Informed by Both Evidence and Decision Time. Neuron, 84(6), 1329–1342. https://doi.org/10.1016/j.neuron.2014.12.015

Purcell, Z. A., Wastell, C. A., & Sweller, N. (2022). Eye movements reveal that low confidence precedes deliberation. Quarterly Journal of Experimental Psychology, 76(7), 1539–1546. https://doi.org/10.1177/17470218221126505

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

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. Psychological Review, 108(3), 550–592. https://doi.org/10.1037/0033-295X.108.3.550