BRIEF REPORT
Computational Models of Dual-Process Reasoning

Purcell, Z. A., Beucler, J., De Neys, W. & Desender, K.

October 12, 2025

1 Abstract

Dual-process theories posit that reasoning emerges from the interplay of fast, intuitive Type 1 processes and slower, deliberative Type 2 processes, regulated by metacognitive control. Yet despite their influence, these frameworks remain largely verbal and underspecified, limiting their capacity for falsifiable predictions. Here, we introduce a set of computational models that formalise key dual-process assumptions within an evidence-accumulation framework. Each model instantiates distinct mechanisms of working memory, inhibition, and confidence-based regulation and is evaluated against data from two-response bat-and-ball reasoning tasks. Six model variants—ranging from single-process "default" to confidence-regulated architectures—were compared using indices of accuracy, response time, and confidence. All models reproduced core reasoning signatures, but even the simplest captured many qualitative patterns, revealing that common verbal assumptions can be implemented by multiple mechanistic architectures. These findings underscore that while dual-process theories capture essential intuitions about intuitive–analytic dynamics, they remain underdetermined without formal specification. Computational modelling provides the necessary precision to advance dual-process theory from a descriptive dichotomy toward mechanistic explanation.

Dual-process theories propose that human reasoning arises from the interaction between two distinct cognitive processes: an intuitive, fast, and automatic Type 1, and a reflective, slow, and effortful Type 2 (Evans and Stanovich [2013], Kahneman [2012]). The engagement and influence of each type are thought to be regulated by metacognitive processes, such as confidence (Thompson et al. [2011], Pennycook et al. [2015], De Neys [2012]). These theories have shaped research across multiple domains, including economic decision-making, moral reasoning, and the human-AI interaction (Greene [2013], Bago et al. [2021], De Neys and Raoelison [2025]. Yet despite their broad influence, few efforts have been made to formalise these verbal models into testable computational frameworks. Existing models of higher-level decisions tend to target narrow effects—such as framing or risk-taking—and often fail to (or simply do not aim to) explicitly model the interaction between the two types of processes (e.g., Zhou and Pitt [2024], Mukherjee [2010].

In this article, we develop a set of computational models that instantiate key assumptions of dual-process theories, translating verbal characterisations of dual-process frameworks into formal models, and comparing them against empirical data from the ubiquitous reasoning task: the bat-and-ball problem (see Box 1). This task has been central to the dual-process literature, and provides robust behavioural markers—such as response times, confidence, and accuracy—that can constrain and inform model development.

Dual-process accounts have been motivated by empirical findings showing that despite having the sufficient mindware to reach a correct solution, we often err–giving intuitive but incorrect responses (Kahneman [2012], Evans [2010]). In bat-and-ball tasks, for instance, participants frequently miss the 'more than' cue, giving the intuitive 10c response (see Box 1; Frederick [2005]). However, some individuals do succeed at this task—this success, according to dual process proponents, may be due to reflecting upon and inhibiting the intuitive 5c response (Frederick [2005]). However, these claims remain poorly specified, and as such remain largely speculative. The question of why some people succeed on this, and other bias-inducing tasks—and the mechanisms that underlie it—remains open.

Box 1: The Classic Bat-and-Ball Problem

A bat and a ball cost \$1.10 together.
The bat costs \$1.00 more than the ball.
How much does the ball cost?
(1) 5c (incorrect, heuristic response)
(2) 10c (correct, logical response)

Traditional behavioural methods are ill-suited to address these mechanistic questions. But, computational modelling provides a way forward by allowing us to specify and quantiatively compare competing accounts of how underlying reasoning processes unfold over time (van Rooij and Blokpoel [2020], Devezer et al. [2020]. For example, one hypothesis is that deliberation affords success on heuristic and bias tasks because it enables one to override or inhibit an initial intuitive response (Kahneman [2012], Frederick [2005]). Others posit that this success is a result of engaging working memory and allowing the successful creation and integration of probabilistic or logical information (Purcell et al. [2021], Bago and De Neys [2020]). These hypotheses are difficult to adjudicate using behavioural data alone, but can be directly instantiated and compared in formal models.

In addition to aiding the specification of and comparison between competing mechanistic hypotheses, computational modelling is well-positioned to help us evaluate causal claims that are otherwise difficult to test. A widely reported pattern in reasoning research is the inverse relationship between confidence and response time—low confidence is consistently associated with longer response times (e.g., Bago and De Neys [2019], Purcell et al. [2022]). While this pattern is often interpreted as evidence that confidence causally regulates the engagement of deliberative thinking, experimental tests have been unable to disentangle its direction and causality. By constructing models that do or do not include causal links between confidence and response dynamics, and examining how well these competing models capture observed behavioural data, we can begin to test the causality of this mechanism.

Although modelling offers many solutions to the issues facing reasoning research, developing computational dual-process models poses significant challenges. It requires formalising (creating mathematical functions for) the distinct processing characteristics of Type 1 and 2, specifying their interaction dynamics, and incorporating relevant metacognitive signals. In this article, we proceed in three steps. First, we distil core assumptions from prominent verbal dual-process theories. Second, we construct six formal models that differ in how they implement various dual-process assumptions, such as inhibition, working memory engagement, and confidence-based regulation. Third, we compare these models against behavioural data from two-response versions of the bat-and-ball task.

2 DUAL-PROCESS THEORY

Contemporary dual-process theories—sometimes grouped under the label "Dual Process Theory 2.0" (De Neys [2023, 2018])—extend classical dual-system models by refining their assumptions and integrating evidence across domains (De Neys [2023], Ackerman and Thompson [2017], Trippas et al. [2016], Pennycook et al. [2015], Stanovich [2018], Reyna et al. [2017], Handley and Trippas [2015a], Thompson and Newman [2017]). While these models differ in scope and emphasis, they converge on several key hypotheses about the distinct capacities of Type 1 and Type 2 processes, and the metacognitive and goal-directed mechanisms that regulate their interaction. The present work builds on these shared assumptions, focusing on five central constructs: working memory, inhibition, metacognition, metacognitive and goal-directed regulation.

Working memory. A foundational claim across dual-process models is that Type 2 processes require working memory, whereas Type 1 processes do not (for a discussion see Thompson [2013]). Working memory enables integration of multiple problem elements and supports inferential reasoning (Baddeley [1992, 2010], Dehaene et al. [2001]). For example, in bat-and-ball tasks (see Box 1), Type 2 processing may allow an individual to combine mathematical knowledge (e.g., how to perform substitution) with contextual cues (e.g., the phrase 'more than'). In contrast, Type 1 relies on automatic associations or heuristics—such as linking $1.10 and 10c, and concept of subtraction—without integrating stored mathematical knowledge or the 'more than' cue. The capacity of Type 2 to combine external task features with internal knowledge is critical for generating novel inferences and (potentially) revising initial responses.

Inhibition. Dual-process models posit that Type 2 processes can override initial responses generated by Type 1 via inhibitory control (Kahneman [2012], Evans [2010]. In classic reasoning tasks, such as the bat-and-ball problem, Type 1 often generates a quick, stereotypical response (e.g., 10c), neglecting logical information. Successful reasoning may require inhibiting this default response, allowing Type 2 to engage and consider competing cues, or the alternative solutions (Frederick [2005]). This inhibition process is central to the default-interventionist account, wherein intuitive responses are produced first and subsequently subject to monitoring and correction by Type 2 if necessary (Evans and Stanovich [2013], De Neys and Glumicic [2008]). Similarly, parallel models also entail that inhibition is necessary to curb incorrect Type 1 processes (Handley and Trippas [2015b]).

Metacognition. Metacognitive processes—including confidence and error monitoring—are increasingly central to dual-process theories (Koriat et al. [2006], Thompson and Johnson [2014]). Recent 2.0 models suggest that confidence arises from the relative activation strength of competing response processes (Bago and De Neys [2019], De Neys [2023]). When one process (e.g., that leading to the heuristic response) dominates, confidence is high; when activation is balanced across competing processes (e.g., heuristic vs. logical), confidence is lower. This framework provides a verbal account of why people report low confidence in reasoning tasks that elicit conflict and has been supported by data linking confidence with longer response times and answer revision (e.g., Bago and De Neys [2019], Thompson et al. [2011]).

Regulation. Dual-process models propose that the engagement of Type 2, deliberative reasoning is regulated by both metacognitive signals—such as confidence and goal-directed factors (De Neys [2023], Ackerman [2014]). Low confidence in an initial response increases the likelihood of further deliberation, as reflected in longer response times and higher rates of answer revision (Thompson et al. [2011], Purcell et al. [2022]). This has been demonstrated using two-response paradigms, where confidence after an initial, time-pressured answer (Response 1) predicts subsequent revision behaviour during an unconstrained phase (Response 2; Thompson et al. [2011]). However, recent models argue that this regulation is also shaped by internal goals and external constraints. According to the diminishing criterion model (Ackerman [2014], Ackerman and Thompson [2017]), individuals become more willing to accept low-confidence responses over time, particularly under conditions of time pressure or low incentives. Thus, Type 2 engagement and cessation reflects a dynamic interplay between metacognitive monitoring and goal-directed control.

In sum, contemporary dual-process theories propose that reasoning reflects the dynamic interaction of two types of processes, distinguished by functions such as working memory and inhibition, and regulated by metacognitive confidence. Although these frameworks offer rich verbal hypotheses about the processes that govern reasoning, they remain largely informal. Key questions remain unanswered: How do different components—such as inhibition and confidence—jointly determine behavioural outcomes? And, how are these processes causally regulated? In the next section, we describe how computational models of decision making can help us begin to address these questions.

3 DECISION MAKING

Decision-making research has developed detailed computational models that describe how people make choices. One of the most widely used frameworks—known as evidence accumulation models—treat decisions as a process of gradually collecting information until one option wins out over another (Ratcliff [1978], Busemeyer and Rapoport [1988]). They describe decisions as the outcome of a stochastic process in which noisy evidence is integrated over time until a decision threshold is reached. When the task invites a decision to be made between two alternatives, this can be envisioned as a race between two independent accumulators, each accumulating evidence for their respective alternative (Logan and Cowan [1984]).

These models are especially well suited to testing the regulatory role of confidence because they can capture how much evidence a person has for each possible answer, and how that balance shapes their confidence (Usher and McClelland [2001], De Martino et al. [2013]). Echoing the verbal assumptions of contemporary dual process theories, confidence in these computational frameworks is defined as the distance between the two options being considered: the larger the difference in evidence, the higher the confidence in that decision.

In addition to defining confidence, race models offer several structural features that align closely with those central to dual-process theories of reasoning. As illustrated in Figure 1, each accumulator (y_l, y_e) is governed by its own drift rate (I_l, I_e) which describes the rate of evidence accumulation. Dual process models assume that working memory is a core feature of Type 2 processing. Working memory facilitates both inference, through the combination of pieces of information, and increased access to stored information and processes via broadcast. As such, we might expect that, as a person engages more Type 2 thinking, this may be reflected as increases to the rate of evidence accumulation which can be captured by changes to the accumulator drift rates.

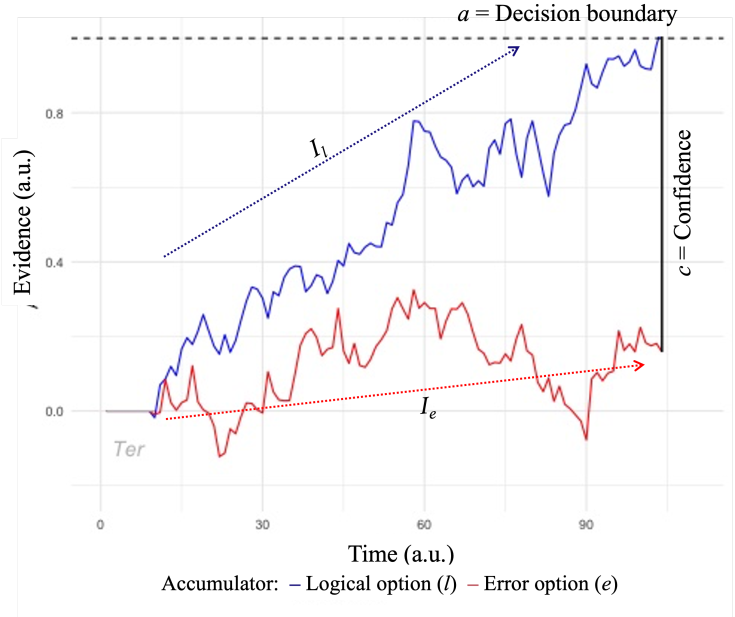
Some race models include a parameter governing the strength of the competition between the accumulators (w_l, w_e; (Usher and McClelland [2001]). Many dual process theories assert that inhibition—the propensity of one solution process to suppress another—is a key Type 2 feature. Therefore, as one thinks more effortfully about a decision, we might expect that the strength of the inhibition, w, between the two accumulators will increase.

A feature of all evidence accumulation models is the evidence boundary, (a). Once one the of two accumulators reaches the boundary, a decision is taken in favour of that accumulator. Dual process models assert that different decisions require different levels of activation depending or evidence depending on the individual's goals and the task context. For example, decisions made under time pressure or that have little consequence for the reasoner may be taken with less evidence than decisions that can be made more slowly or for which the consequences are greater. As such, we might expect that when people are given more time to think about a problem, the evidence boundary increases.

Evidence accumulation models provide useful parameters that can be adapted to account for several dual process assumptions about: working memory (via modulation of drift), inhibition (via cross-suppression), confidence (via the difference between the accumulators), and evidence requirements (via the decision boundary). However, generalising these models to higher-level reasoning tasks requires important extensions. In particular, that 1) as a person moves from more Type 1 to Type 2 processing, we may observe changes to one, some, or all of these parameters and 2) that confidence plays a role in regulating these changes. In the present framework, we extend race models to include these components.

4 DEVELOPING COMPUTATIONAL MODELS OF REASONING

We formalise key components of dual-process reasoning—*working memory*, *inhibition*, *metacognition*, and *regulation*—by linking these processes to key pa-



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

Figure 1: Example of an evidence accumulation model with two accumulators that 'race' toward the decision boundary (a). Each of the accumulators has an independent rate of evidence accumulation (drift rates I_l and I_e) and the difference between the accumulators at boundary crossing reflects confidence (c) in that decision. 'a.u.' $=$ arbitrary units. As depicted in the equation, changes in evidence (dy) for each accumulator are suppressed by the strength of the other at a rate of w. Ter $=$ Non-decision time e.g., for motor response.

rameters from race models adapted to a two-response paradigm. As outlined earlier, we will use a paradigm that involves two responses per problem: an initial answer under time pressure (Step 1), followed by a chance to revise it without constraints (Step 2); after each response, participants rate their confidence (Thompson et al. [2011]). This structure is not intended to mimic naturalistic decision-making, but to offer a controlled framework for identifying whether intuitive (Type 1) and deliberative (Type 2) processes exhibit distinct computational signatures, and whether early confidence regulates subsequent Type 2 engagement. This approach allows us to examine two central questions in dual-process theory:

- **Duality**: Are Type 1 and Type 2 processes mechanistically distinct, and if so, how?
- **Regulation**: Does confidence regulate the engagement of Type 2 processes, and if so, how?

To address these questions, we develop *two-step computational models.* For the two steps—matching Response 1 and Response 2 in two-response paradigms—we implement decisions as a race between two accumulators: one for the *erroneous* response (e), typically aligned with intuitive but incorrect answers, and one for the *logical* response (l), corresponding to the normatively correct choice. Accumulators have independent drift rates (I_e, I_l) and inhibition weights (w_e, w_l). Confidence (c) and decision boundaries (a) are also explicitly parameterised (see Figure 1 for a single-step depiction). We introduce three model classes:

1. **Default model**: A single process governs both steps with fixed parameters, assuming no qualitative shift between intuition and deliberation.
2. **Shift models**: Parameters vary between steps, capturing a transition in processing dynamics once deliberation becomes possible—consistent with the *duality* hypothesis.
3. **Confidence-Regulation models**: These extend the shift models by linking *confidence* at Step 1 to parameter changes between steps, thereby instantiating the *regulation* hypothesis.

Each model is described in detail below.

3.1 Default model

In the *default model*, the parameters governing evidence accumulation remain constant across both decision phases. That is, the drift rates (I_1, I_2) and inhibitory competition terms (w_1, w_2) are fixed between Step 1 (timed) and Step 2 (untimed) responses. This model assumes that the same underlying cognitive process supports both intuitive and deliberative responses, and thus reflects a single decision system. The default model is intended to provide a baseline against which to evaluate *shift* and *confidence* models that explicitly incorporate dual-process dynamics and confidence-driven regulation.

Step 1:

$$\begin{cases} dy_e = (-w_e y_l + I_e)dt + cdW, & y_e(0) = 0, \\ dy_l = (-w_l y_e + I_l)dt + cdW, & y_l(0) = 0 \end{cases} \tag{2}$$

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

$$\text{Stop when: } \max\big(y_e(t), y_l(t)\big) \geq a_1. \tag{3}$$

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

Confidence in the first response is then defined as:

$$c = |y_e(t) - y_l(t)|\,. \tag{4}$$

The decision process continues (untimed) as follows:

Step 2:

$$\begin{cases} dy_e = (-w_e y_l + I_e)dt + cdW, \\ dy_l = (-w_l y_e + I_l)dt + cdW. \end{cases} \tag{5}$$

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:

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

In the following models, decision processes for Step 1 are always defined as in equation (2) above. However, in line with dual-process assumptions, in each of the next models, a change occurs in one parameter in Step 2. Primes $(')$ indicate changed parameters generated for the second decision period.

3.2 Shift Models

3.2.1 Drift Shift

In the *drift shift* model, reasoning at Step 2 reflects enhanced access to task-relevant information through increased working memory engagement. This is operationalised as a change in drift rates (I_l', I_e') relative to that for Step 1.

The logic is that when time and cognitive resources are available, participants can more effectively maintain and manipulate problem-relevant representations, they may also have increased access to more distant cognitive resources, thereby increasing the rate at which evidence is accumulated for the logical response.

For the drift-shift model, Step 2 is defined as:

$$\begin{cases} dy_e = (-w_e y_l + I'_e)dt + cdW, \\ dy_l = (-w_l y_e + I'_l)dt + cdW. \end{cases} \tag{6}$$

Where stimulus input is scaled by fixed multipliers:

$$\begin{cases} I'_e = I_e * I_e m, \\ I'_l = I_l * I_l m. \end{cases} \tag{7}$$

3.2.2 Inhibition Shift

The *inhibition shift* model assumes that increased cognitive control at Step 2 enables more effective suppression of competing or prepotent responses. This is modelled as a change in the suppression terms (w'_e, w'_l), which reflect the degree to which each accumulator inhibits its competitor. Enhanced inhibition may facilitate the eventual selection of the final response by reducing activity of the competing accumulator, even when drift rates remain constant.

Step 2 is defined as:

$$\begin{cases} dy_e = (-w'_e y_l + I_e)dt + cdW, \\ dy_l = (-w'_l y_e + I_l)dt + cdW. \end{cases} \tag{8}$$

Where inhibition weights are scaled by fixed multipliers:

$$\begin{cases} w'_e = w_e * w_e m, \\ w'_l = w_l * w_l m. \end{cases} \tag{9}$$

3.3 Confidence-Regulation Models

In the first two *confidence-regulation* models, the parameters that were allowed to change in the shift models above (e.g., drift rate, inhibition) again change from Step 1 to 2. However, in this case, the change in the parameter is dependent on confidence at the time when the first response is made t_1. In the third *confidence-regulation* model, the boundary parameter at Step 2 changes as a function of confidence at t_1. These models formalise different specifications of the hypothesis that confidence regulates whether and to what extent deliberative processing is engaged.

3.3.2 Confidence-Drift Model

This model assumes that confidence influences the rate of evidence accumulation. Specifically, lower confidence facilitates faster accumulation, while higher confidence slows it, reflecting the proposed relationship between confidence and deliberation as working memory resources.

Confidence is defined as the absolute difference between the two accumulators at the moment of the initial decision, see (3). However, to allow positive and negative modulation of the drift rate we rescale the confidence term to the range $[-1, 1]$ as follows:

$$c_{\text{rescaled}} = 2\left(\frac{c}{a}\right) - 1, \tag{10}$$

where a is the decision threshold used at Step 1. When $c_{\text{rescaled}} \approx +1$, the participant was highly confident; when $c_{\text{rescaled}} \approx -1$, confidence was low. This linear transformation enables symmetric modulation of evidence accumulation across trials.

The drift rates at Step 2 are defined as:

$$\begin{cases} I'_e = I_e * I_e m + c_e * c_{\text{rescaled}}, \\ I'_l = I_l * I_l m + c_l * c_{\text{rescaled}}, \end{cases} \tag{11}$$

where c_e and c_l are gain parameters that determine the extent to which confidence modulates the accumulation rates for each accumulator.

3.3.2 Confidence-Inhibition Model

This model assumes that confidence modulates the level of inhibition between accumulators. This reflects the verbal assumption that low confidence leads to stronger inhibition and high confidence reduces the inhibitory interaction. Unlike the confidence–drift model, we do not rescale confidence to $[-1, 1]$, as allowing negative values would imply excitation rather than inhibition.

Inhibition parameters at Step 2 are modulated by raw confidence, see (3):

$$\begin{cases} w'_e = w_e * w_e m + c_e * c, \\ w'_l = w_l * w_e m + c_l * c, \end{cases} \tag{12}$$

where c_e and c_l are scaling parameters that govern the extent to which confidence modulates inhibition.

3.3.3 Confidence-Boundary Model

This model assumes that confidence modulates the final decision boundary. The decision boundary reflects how much evidence must accumulate before a response is made; thus, changes to the boundary reflect changes in response caution. When confidence is low, a higher boundary may be set to enable more evidence accumulation before responding (i.e., more cautious responding). Conversely, when confidence is high, a lower boundary may be used, reflecting reduced need for caution or deliberation. This formulation reflects the dual-process assertions that (a) low confidence leads to greater deliberation and (b) task-goal trade-offs where low confidence leads to greater caution and a need for more evidence before committing to a decision.

As for the confidence-drift model, here we also rescale confidence to allow positive and negative modulation of the boundary (relative to a); see (10).

The decision process at Step 2 proceeds identically to previous models (as defined in Equations 1 and 2), but with a modified stopping rule:

$$\max\big(y_e(t), y_l(t)\big) \geq a_2', \tag{13}$$

where the new boundary is defined as a linear function of scaled confidence:

$$a_2' = a_2 - c_m * c_{\text{rescaled}}. \tag{14}$$

Here, a_2 represents the baseline boundary level (e.g., matched to Step 1), and c governs how much the boundary increases or decreases based on confidence. When $c < 0$, lower confidence leads to more cautious responding (higher boundary), while high confidence yields a reduced threshold. This model formalises the hypothesis that confidence acts as a metacognitive signal regulating the amount of evidence required (either due to similar activations of competing response options or higher confidence standard due to individual goals) before committing to a decision in subsequent reasoning.

5 METHOD

4.1 Participants

Ninety-nine US participants (47% female; $M_{\text{age}} = 31.06$ years, $SD = 10.07$) were recruited via Prolific. Participants were compensated £4.50 for their time. All participants were located in the United States and provided informed consent prior to participation. No participants were excluded from the behavioural analyses, while eight were excluded from the computational analyses (see below).

4.2 Materials

Participants completed 100 bat-and-ball style reasoning problems (Raoelison and Neys [2019]). The item set comprised 50 conflict and 50 no-conflict problems, presented in randomised order. Content was adapted for variation in surface features while preserving the underlying logical structure. The conflict items followed the underlying structure of the canonical bat-and-ball problem, but their surface features were modified to vary across trials (Bago and De Neys [2019]). For instance, one item read: "A building has 370 pets in total, consisting of dogs and cats. There are 300 more dogs than cats. How many cats are there?" [35, 70]. These items were designed to simultaneously cue an intuitively appealing but incorrect heuristic response, as well as the correct response derivable via a basic algebraic formulation (e.g., $300 + 2x = 370$). A key linguistic feature of these items was the inclusion of the comparative phrase "more than," which often triggers misinterpretation. In contrast, no-conflict items were structurally similar but omitted the comparative phrasing and did not elicit a conflict between heuristic and analytical responses. For example: "A building has 370 pets in total, consisting of dogs and cats. There are 300 dogs. How many cats are there?" [35, 70]. In these cases, the intuitive answer aligned with the correct response (e.g., $300 + x = 370$), and thus no response conflict was expected.

4.3 Procedure

Participants were instructed that each problem would be answered twice. They were first asked to provide their intuitive response—the first answer that came to mind—within a limited time, then given the opportunity to reconsider their answer without time constraints. After each response, they rated their confidence on a six-point scale (1 = not at all confident, 6 = very confident). Two practice trials preceded the main set of 100 items. Responses were made using the keyboard: "C" and "N" keys selected the left and right options, respectively. Confidence ratings from 1 to 3 were entered using the number keys 1, 2, and 3, and ratings from 4 to 6 using the 8, 9, and 0 keys. At the end of the session, participants reported their age, gender, and country of residence.

Each trial followed a fixed sequence adapted from Thompson et al. (2011; see Figure 2). A fixation cross appeared for 1000 ms, followed by a partial problem stem for 2000 ms (e.g., "In a building residents have 370 dogs and cats in total."). The full problem and two response options were then presented for up to 4000 ms or until a response was made (e.g., "There are 300 more dogs than cats. How many cats are there?" [35, 70]). During the final 2000 ms, the screen background turned yellow to signal the approaching deadline. If participants failed to respond in time, they were reminded to answer more quickly on subsequent trials. All trials where the participant did not respond before the initial time-limit were removed before analysis. Participants were then asked to give their confidence in that response, then answer the problem again–this time with no time constraints–finally, they gave their confidence in their second response.

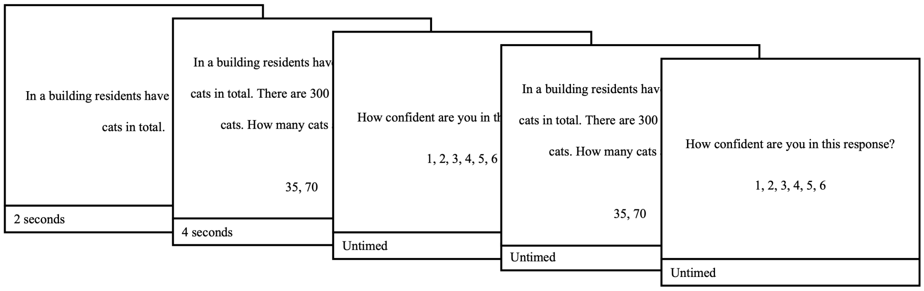

In a building residents have
cats in total.
2 seconds
In a building residents hav
cats in total. There are 300
cats. How many cats
35, 70
4 seconds
How confident are you in th
1, 2, 3, 4, 5, 6
Untimed
In a building residents hav
cats in total. There are 300
cats. How many cats
35, 70
Untimed
How confident are you in this response?
1, 2, 3, 4, 5, 6
Untimed

Figure 2: Two-response trial format.

4.4 Statistical and Model Fitting Approach

We used mixed-effects models to analyse accuracy, response times, confidence, and their interrelations. All models included subject as a random intercept. Accuracy (0, 1) was analysed using a logistic model with fixed effects of response stage (initial vs. final), question type (conflict vs. no-conflict), and their interaction. Response times were log-transformed and analysed with linear models using the same predictors. Confidence ratings were also analysed with linear models. We then examined response times and confidence (separately) as functions of response stage, question type and accuracy, and all their interaction. To examine the role of confidence in subsequent behaviour, we fit additional models predicting response times—using linear mixed model—and answer change—using a logistic mixed model—from initial confidence, question type, and their interaction, with subject as a random intercept. All model analyses used the lmerTest package (Kuznetsova et al. [2013]) in RStudio (RStudio Team [2019]) using an alpha level of .05 for significance testing.

We modelled our data using each of the six two-step computational models developed above. For all models, noisy evidence is accumulated separately for two response options (error and logical). For Step 1—reflecting the Response 1 process—the rate of evidence accumulation for each option is determined by independent drift rates I_e and I_l, and inhibition rates w_e and w_l—until a first decision boundary, a_1, is reached. Changes between the proposed models only emerge at Step 2. For the *default* model, Step 2 is governed by the same weights and parameters as in Step 1, except the boundary, a_2, which is able to change freely. Under this model, we assume that the same decision process is responsible for the decision at both Step 1 (initial, under time constraint) and Step 2 (final, no time constraint)—only at Step 2 this decision process continues until a new boundary is reached. Otherwise, there is no mechanistic difference between intuitive and deliberative stages.

The remaining five models allow the weights to change at Step 2 according to different dual-process principles. The two *shift* models allow the drift and inhibition parameter weights (I and w) to shift at Step 2, respectively. These models align with the dual process assertions that deliberation allows greater access to cognitive resources and increased capacity for inhibition. The three *confidence-regulation* models allow the drift, inhibition, and decision threshold parameter weights (I, w, a_2) to change at Step 2, respectively. In these models, the weights change as a function of confidence. A latent confidence parameter (c) is calculated by taking the absolute difference between the accumulators at the end of Step 1. For models in which confidence is allowed to have a negative impact (i.e., the confidence-drift and the confidence-boundary models), the confidence term is scaled to -1 to 1. These *confidence-regulation* models reflect the contemporary dual process assumption that confidence regulates the engagement and duration of deliberation. For all models, non-decision components were captured by non-decision parameters ter_1 and ter_2 at each Step.

Before fitting the models, we explored the range of behavioural patterns each model can produce by simulating them across wide parameter ranges, focusing on Step 2 outcomes (as all models behave identically in Step 1). This allowed us to verify the ability of the models to reproduce key behavioural patterns, namely, response choice (error or logical) and response time distributions across conflict and no-conflict trials, as well as getting a better understanding of the differences between our models. We conducted posterior predictive checks to assess whether the fitted models could reproduce key qualitative patterns in the data (Palminteri et al. [2017]). These analyses examined how well each model matched the empirical patterns of confidence, response times and accuracies, both overall and as a function of conflict condition and response stage.

6 RESULTS

5.1 Accuracy, Response Times, and Confidence

Behavioural. The accuracy analysis revealed a significant interaction between response stage and question type, $\chi^2(1) = 7.06, p = .008$). For no conflict items, accuracy was higher for final than initial responses ($OR = 1.26, p = .001$), while for conflict items, accuracy did not differ across responses ($p = .800$; see Figure 3). Response times were significantly longer at Response 1 than Response 2, ($F(18104) = 33.02$, $p < .001$). No other effects were significant (see Figure 4). A linear mixed-effects model predicting log-transformed response times from accuracy response times, and question type revealed several significant main and interaction effects. There was a significant three-way interaction between Response, Question, and Accuracy, $F(1, 17, 734.37) = 7.14$, $p = .008$, indicating that the effect of accuracy on response times differed depending on both the response stage and question type. In pairwise comparisons, accuracy effects varied by response stage and question type: there was a significant effect of accuracy in the initial response to conflict problems, $t(17, 811) = 2.88, p = .004$, and in the final response to conflict problems, $t(17, 816) = -5.33, p < .001$. No significant accuracy effects were found in no-conflict problems, either at the initial, $t(17, 771) = 0.27, p = .786$, or final response stage, $t(17, 782) = -1.44, p = .150$.Confidence was higher for final than initial responses, $F(1) = 531.13, p < .001$, and for no-conflict than conflict items, $F(1) = 12.29, p < .001$, as indicated by main effects, however, the interaction was not significant.

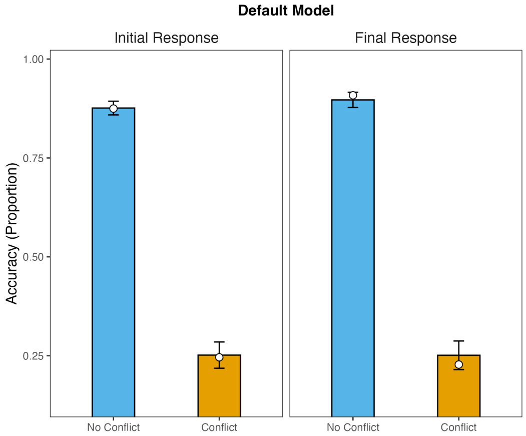


Figure 3: Accuracy as a function of response stage and conflict. Bars show participants' mean accuracy with standard errors, and white dots indicate predictions from the Default model.

Computational. Results were quite consistent across models. All models were able to capture: 1) the lower accuracy for conflict compared to no-conflict responses (see Figure 3), 2) the longer response times being more frequent in the final compared to the initial response stage (see Figure 4), and 3) the more subtle interaction between conflict and accuracy for reaction times, with reaction times being higher for correct responses compared to incorrect responses for conflict items (see Figure 4). However, the models tended to systematically underestimate final response times.

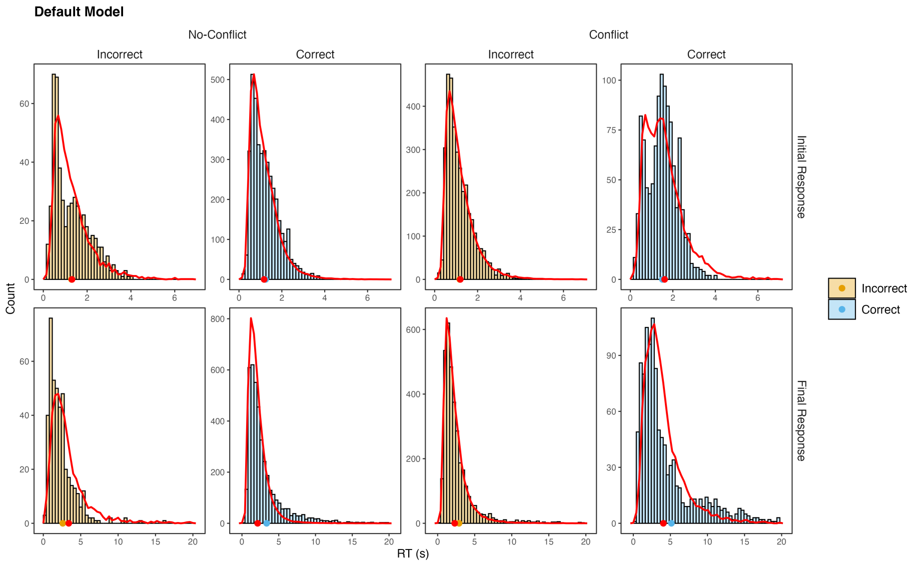


Figure 4: Response time distributions as a function of response stage, conflict, and accuracy. Histograms show trial distributions across all participants, the red line shows the normalised Default model distribution, and overlaid dots indicate mean reaction times for participants (black) and the model (red).

5.2 Confidence and Deliberation Indices

Behavioural. Regarding confidence and indices of deliberation, initial confidence predicted subsequent behaviour: higher confidence at Response 1 was associated with shorter Response 2 times ($B = -0.29, p < .001$) and lower likelihood of answer change ($B = -1.48, p < .001$; see Figure 5).

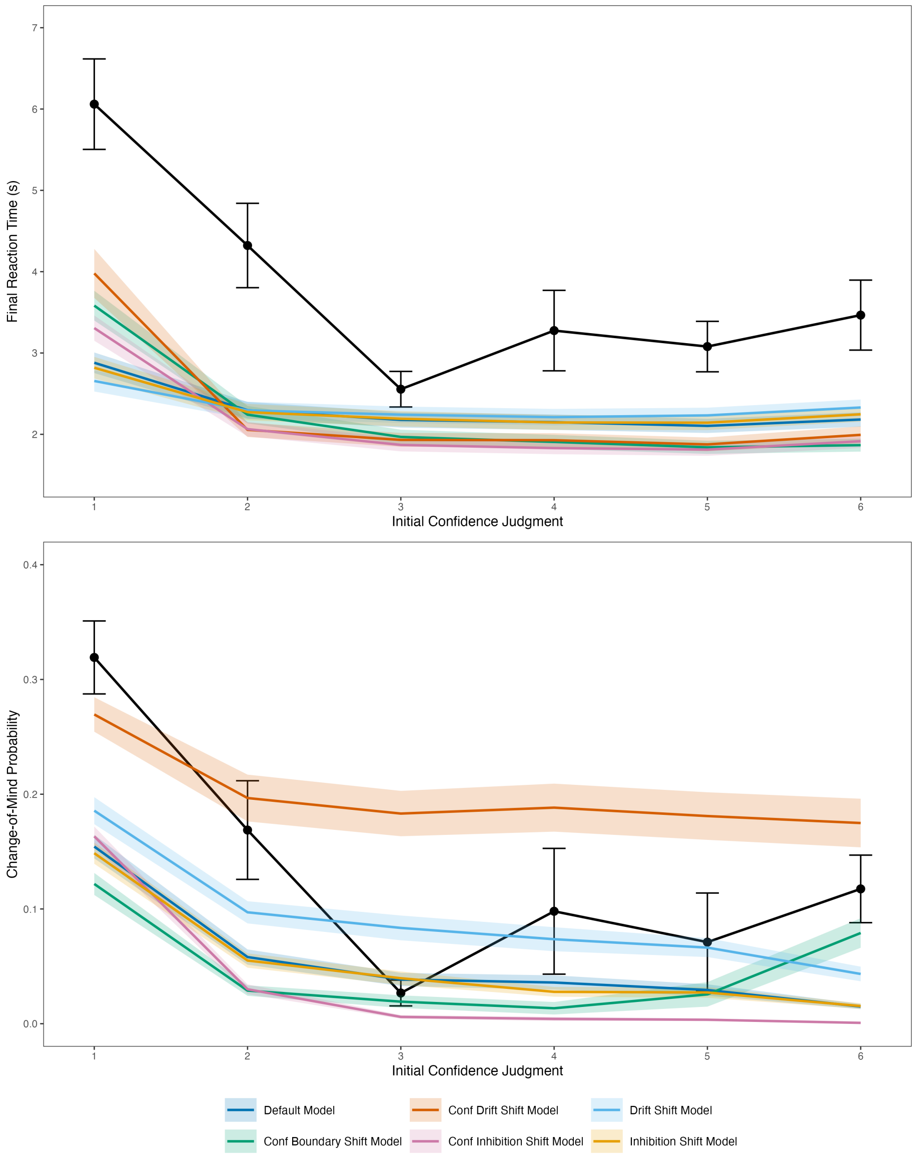


Figure 5: Confidence–deliberation link. The upper panel shows the relationship between initial confidence and final reaction time, and the lower panel shows the relationship between initial confidence and the proportion of changes of mind in the final response stage. Participant data are shown in black (discrete confidence ratings), and model confidence is binned into sextiles of the latent initial confidence variable.

Computational. We tested whether the models could reproduce the confidence patterns found in the data, as well as the relationship between initial confidence and deliberative indices: (a) response times and (b) answer changes in Step 2. In the models, confidence is represented as a latent variable emerging from the internal decision process; here, we directly compare these latent confidence values to the explicit confidence ratings provided by participants. Because confidence measures were not used during fitting, these analyses provide an out-of-sample test on aspects of the data not involved in parameter estimation. This approach reduces the risk of overfitting and offers a stronger assessment of each model's ability to capture latent reasoning processes, allowing us to systematically compare the adequacy of each computational account of dual-process reasoning. As shown in Figure 5, all models were able to capture: 1) the negative relationship between confidence at Response 1 and response times at Response 2, and 2) the negative relationship between confidence at Response 1 and change-of-mind at Response 2.

The Default model provided a good overall qualitative account of participants' behaviour across variables. It captured the main patterns observed in the data, including the relationships between initial confidence, final reaction time, and changes of mind, despite not explicitly modelling these effects. Notably, the simpler architecture of the Default model reproduced the same qualitative trends captured by more complex alternatives, suggesting that much of the observed behaviour in the bat-and-ball task under two-response conditions can be explained without additional model mechanisms.

7 GENERAL DISCUSSION

Dual-process theories of reasoning are highly influential verbal accounts of human reasoning, positing that intuitive and deliberative processes interact under the guidance of metacognitive control. However, despite their theoretical richness, such accounts often lack the precision needed to generate falsifiable predictions or specify concrete mechanisms. In the present work, we translated these verbal theories into a set of computational models, each formalising core constructs—working memory, inhibition, confidence, and regulation—within a race-based evidence accumulation framework. This modelling effort serves not only to evaluate the plausibility of key dual-process assumptions but also to demonstrate how computational instantiations can clarify, sharpen, and sometimes challenge verbal theorising.

6.1 Limitations of Verbal Dual-Process Theories

Despite their theoretical richness, verbal dual-process theories have long faced major conceptual challenges. First, they often lack the specificity required for falsifiable predictions. Terms such as "inhibition," "analytic override," or "confidence regulation" are used descriptively but without formal definitions of their dynamics or interactions. As a result, the same empirical pattern—such as longer response times before correct answers—can be taken as support for entirely different mechanisms. This interpretive flexibility has made it difficult to disconfirm verbal theories and, consequently, to build cumulative progress across studies.

Second, existing dual-process frameworks remain ambiguous about how their key components—such as working memory, inhibition, and metacognition—interact over time. For example, many accounts posit that inhibition allows Type 2 reasoning to suppress Type 1 output, yet they rarely specify how or when this occurs, or how it depends on other factors like confidence or task difficulty. Similarly, the hypothesised role of confidence as a regulator of analytic engagement remains underspecified: some accounts treat confidence as an output of processing, others as an input controlling it. Without explicit mechanistic structure, these verbal models cannot reveal the causal architecture linking these processes.

Third, even the most recent "Dual Process 2.0" frameworks—though more nuanced than classical two-system accounts—retain an essentially verbal format that makes quantitative prediction nearly impossible. They articulate rich qualitative hypotheses about metacognitive regulation, but stop short of defining how changes in working memory, inhibition, or decision thresholds produce measurable behavioural outcomes. Consequently, competing theories can accommodate similar data, leaving open the fundamental questions of what, when, and how dual processes interact during reasoning.

6.2 The Need for Computational Specification

The very fact that we were able to construct five distinct computational models from these theories highlights their vagueness. Each model represents a minimal instantiation of a particular verbal claim—whether that deliberation enhances working memory (drift shift), increases inhibition (inhibition shift), or whether confidence modulates these processes directly. Importantly, these five models are not meant as exhaustive formalisations but as boundary cases that delineate the minimal commitments required to implement common verbal hypotheses. That so many plausible formulations can be drawn from a single set of verbal claims suggests that the theories themselves are underdetermined.

Moreover, each of these models could, in principle, be elaborated into more complex hybrids. For example, a model could combine confidence-driven changes in drift rate with simultaneous adjustments to inhibition or evidence boundaries. The ease with which such extensions can be conceived underscores the flexibility—and hence the indeterminacy—of verbal theorising in this domain. The modelling exercise therefore does not merely translate dual-process assumptions into equations; it exposes their current underspecification and points toward the necessity of formalisation as a means of theoretical refinement.

6.3 What the Models Capture

Despite these conceptual limitations, the modelling results are encouraging. The relatively simple models captured several core behavioural signatures of reasoning. Across models, they reproduced the characteristic differences between conflict and no-conflict items, the longer response times for correct versus incorrect responses in conflict trials, and the inverse relationship between confidence and deliberation (as indexed by subsequent response times and changes of mind).

The fact that models as simple as the default architecture reproduced many of these qualitative patterns is instructive. It suggests that much of the behaviour observed in two-response reasoning paradigms can be explained by modest shifts in evidence accumulation dynamics—without invoking elaborate control hierarchies. Nevertheless, the more elaborate confidence-regulation models extend this by showing how metacognitive variables could mechanistically drive the transition between intuitive and reflective thought, offering a bridge between traditional dual-process constructs and observable performance data.

6.4 Limitations and Next Steps

This work constitutes a proof of concept rather than a fully validated model set. While the six models provide a principled starting point for formalising dual-process theories, they were not subjected to the rigorous cross-validation and model recovery procedures recommended in recent computational modelling guidelines (Wilson and Collins [2019]). The next step for the field will be to perform systematic model comparison across datasets and tasks, ideally with hierarchical Bayesian fitting and predictive cross-validation to assess generalisability. Such efforts will be critical to move from demonstrating that models can capture the data to establishing which mechanisms are necessary or preferable.

Future work should also seek to adjudicate between these mechanistic accounts. For instance, are reasoning improvements across time better explained by enhanced inhibition or by increases in working memory integration? Does confidence regulate deliberation through changes in evidence thresholds, or by altering accumulation dynamics? Addressing these questions will require experiments that manipulate these mechanisms directly— e.g., through cognitive load, incentive, or metacognitive interventions—and testing whether the corresponding model parameters shift in predictable ways. Only through such iterative cycles of modelling and empirical testing can dual-process theories evolve from verbal descriptive theories into mechanistic frameworks.

6.5 Broader Implications

Beyond the domain of reasoning, formalising dual-process models can help clarify long-standing debates in diverse areas of psychology and cognitive science. For example, similar intuitions about fast versus slow thinking underlie research on misinformation and belief revision, moral and climate argumentation, and gender or age differences in confidence and cognitive effort. A computational framework allows these phenomena to be analysed within a common architecture, specifying whether differences arise from altered evidence accumulation rates, inhibitory control, or confidence thresholds. In this sense, computational dual-process models offer a unifying language for testing claims across domains—from how misinformation resists correction to how confidence biases emerge in educational or gender-related contexts.

6.6 Conclusion

In summary, we have demonstrated that the verbal dual-process framework, while influential, remains too underspecified to guide cumulative theoretical progress. By translating its assumptions into six computational models, we revealed both the flexibility and the limits of current theorising: multiple models can capture key behavioural patterns, yet they embody different mechanistic commitments. This exercise highlights the need to move beyond verbal dichotomies toward formal, testable architectures. We propose that computational models—anchored in explicit equations linking working memory, inhibition, confidence, and regulation—offer the most promising path forward for refining dual-process theories and for bridging the gap between descriptive constructs and mechanistic understanding.

8 APPENDIX

To characterise the cognitive mechanisms underlying dual-stage reasoning, we implemented two-step models, simulating the temporal dynamics of intuitive and deliberative responses. Each model instantiates two sequential decision processes—an early, speeded response under time pressure (Step 1) and a subsequent, reflective response without time constraints (Step 2). The full behavioural trajectory on each trial is modelled as a pair of evidence accumulation processes evolving toward response boundaries. Simulations were executed via a custom C++ engine with Rcpp bindings for integration into the R environment. This ensured efficient simulation over repeated trials while preserving trial-level item structure and conflict labels.

A1. Model Architecture

Each response stage is governed by a race between two accumulators (representing the logical and intuitive responses), with inputs modulated by the type of reasoning problem. Conflict trials assign competing drift inputs to both accumulators, whereas no-conflict trials funnel evidence exclusively toward the logical accumulator, while the other accumulator accumulates only noise. The model includes key parameters for drift rate (I_e, I_l), lateral inhibition (w_e, w_l), decision thresholds (a_1, a_2), and non-decision times (ter_1, ter_2). A confidence signal is computed as the absolute difference in accumulator values at threshold crossing.

The *default model* assumes the same parameter set governs both Step 1 and Step 2, allowing a_2 to change. In contrast, the *modulated models* introduced dynamic changes to parameter weights for I and w at Step 2, implemented in one of two ways. In the *shift models*, parameter weights at Step 2 are shifted relative to their Step 1 counterparts (e.g., increased inhibition or drift). In the *confidence-based models*, Step 2 parameters are adjusted relative to their Step 1 weights *but* as a function of confidence, c. For example, low initial confidence may result in increased drift, greater inhibition, or a higher threshold at Step 2, operationalising the idea that confidence modulates deliberative resource allocation.

A2. Parameter Estimation

Model parameters were estimated individually for each participant using a simulation-based quantile fitting approach (e.g., Desender et al. [2021]). Specifically, we minimized the G^2 deviance—the divergence between observed and simulated response time distributions—computed separately by response stage (Step 1, Step 2) and accuracy (correct, incorrect). The fitting procedure used empirical reaction time quantiles at 0.1, 0.3, 0.5, 0.7, and 0.9, dividing each distribution into six bins defined by these five cut points to capture its overall shape:

$$G^2 = 2N_{\text{obs}} \sum_{i=1}^{q} p_{\text{obs},i} \ln\left(\frac{p_{\text{obs},i}}{p_{\text{pred},i}}\right) \tag{15}$$

where N_{obs} is the number of observed trials, $p_{\text{obs},i}$ and $p_{\text{pred},i}$ are the observed and predicted proportions in bin i, and q is the total number of quantile bins.

We minimized G^2_{total} using differential evolution optimization as implemented in the `DEoptim` package in R (Mullen et al. [2011]). Each parameter set was evaluated by simulating 1,000 trials per participant, with 10 stochastic replicates per observed trial to account for variability in the DDM simulation. The best-fitting parameters were retained for each participant after 1000 iterations, or earlier if the improvement in the deviance fell below a relative tolerance of 10^{-6} for more than 200 consecutive steps.

References

Jonathan St. B. T. Evans and Keith E. Stanovich. Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psycholog-*

ical Science, 8(3):223–241, May 2013. ISSN 1745-6916. doi: 10.1177/1745691612460685. URL https://doi.org/10.1177/1745691612460685. Publisher: SAGE Publications Inc.

Daniel Kahneman. *Thinking, fast and slow*. Penguin psychology. Penguin Books, London, 2012. ISBN 978-0-14-103357-0.

Valerie A. Thompson, Jamie A. Prowse Turner, and Gordon Pennycook. Intuition, reason, and metacognition. *Cognitive Psychology*, 63(3):107–140, November 2011. ISSN 0010-0285. doi: 10.1016/j.cogpsych.2011.06.001. URL https://www.sciencedirect.com/science/article/pii/S0010028511000454.

Gordon Pennycook, Jonathan A. Fugelsang, and Derek J. Koehler. What makes us think? A three-stage dual-process model of analytic engagement. *Cognitive Psychology*, 80:34–72, August 2015. ISSN 0010-0285. doi: 10.1016/j.cogpsych.2015.05.001. URL https://www.sciencedirect.com/science/article/pii/S0010028515000481.

Wim De Neys. Bias and Conflict: A Case for Logical Intuitions. *Perspectives on Psychological Science*, 7(1):28–38, January 2012. ISSN 1745-6916. doi: 10.1177/1745691611429354. URL https://doi.org/10.1177/1745691611429354. Publisher: SAGE Publications Inc.

Joshua D. Greene. *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*. Penguin Press, New York, NY, 2013. URL https://www.amazon.com/Moral-Tribes-Emotion-Reason-Between/dp/1594202605.

Bence Bago, Jean-François Bonnefon, and Wim De Neys. Intuition rather than deliberation determines selfish and prosocial choices. *Journal of Experimental Psychology: General*, 150(6):1081–1094, 2021. ISSN 1939-2222. doi: 10.1037/xge0000968. Place: US Publisher: American Psychological Association.

Wim De Neys and Matthieu Raoelison. Humans and LLMs rate deliberation as superior to intuition on complex reasoning tasks. *Communications Psychology*, 3(1):141, September 2025. ISSN 2731-9121. doi: 10.1038/s44271-025-00320-8. URL https://www.nature.com/articles/s44271-025-00320-8. Publisher: Nature Publishing Group.

Ran Zhou and Mark A. Pitt. Dual-process modeling of sequential decision making in the balloon analogue risk task. *Cognitive Psychology*, 149:101629, March 2024. ISSN 0010-0285. doi: 10.1016/j.cogpsych.2023.101629. URL https://www.sciencedirect.com/science/article/pii/S0010028523000877.

Kanchan Mukherjee. A dual system model of preferences under risk. *Psychological Review*, 117(1):243–255, 2010. ISSN 1939-1471. doi: 10.1037/a0017884. Place: US Publisher: American Psychological Association.

Jonathan St. B. T. Evans. *Thinking twice: Two minds in one brain*. Thinking twice: Two minds in one brain. Oxford University Press, New York, NY, US, 2010. ISBN 978-0-19-954729-6. Pages: viii, 240.

Shane Frederick. Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, 19(4):25–42, December 2005. ISSN 0895-3309. doi: 10.1257/089533005775196732. URL https://www.aeaweb.org/articles?id=10.1257/089533005775196732.

Iris van Rooij and Mark Blokpoel. Formalizing Verbal Theories. *Social Psychology*, 51(5):285–298, September 2020. ISSN 1864-9335. doi: 10.1027/1864-9335/a000428. URL https://econtent.hogrefe.com/doi/abs/10.1027/1864-9335/a000428. Publisher: Hogrefe Publishing.

Berna Devezer, Danielle J. Navarro, Joachim Vandekerckhove, and Erkan Ozge Buzbas. The case for formal methodology in scientific reform. *Royal Society Open Science*, 8(3):200805, 2020. doi: 10.1098/rsos.200805. URL https://royalsocietypublishing.org/doi/full/10.1098/rsos.200805. Publisher: Royal Society.

Zoe A. Purcell, Colin A. Wastell, and Naomi Sweller. Domain-specific experience and dual-process thinking. *Thinking & Reasoning*, 27(2):239–267, April 2021. ISSN 1354-6783, 1464-0708. doi: 10.1080/13546783.2020.1793813. URL https://www.tandfonline.com/doi/full/10.1080/13546783.2020.1793813.

Bence Bago and Wim De Neys. Advancing the specification of dual process models of higher cognition: a critical test of the hybrid model view. *Thinking & Reasoning*, 26(1):1–30, January 2020. ISSN 1354-6783. doi: 10.1080/13546783.2018.1552194. URL https://doi.org/10.1080/13546783.2018.1552194. Publisher: Routledge _eprint: https://doi.org/10.1080/13546783.2018.1552194.

Bence Bago and Wim De Neys. The Smart System 1: evidence for the intuitive nature of correct responding on the bat-and-ball problem. *Thinking & Reasoning*, 25(3):257–299, July 2019. ISSN 1354-6783. doi: 10.1080/13546783.2018.1507949. URL https://doi.org/10.1080/13546783.2018.1507949. Publisher: Routledge _eprint: https://doi.org/10.1080/13546783.2018.1507949.

Zoe A. Purcell, Colin A Wastell, and Naomi Sweller. Eye movements reveal that low confidence precedes deliberation. *Quarterly Journal of Experimental Psychology*, 76(7):1539–1546, September 2022. ISSN 1747-0218. doi: 10.1177/17470218221126505. URL https://doi.org/10.1177/17470218221126505. Publisher: SAGE Publications.

Wim De Neys. Advancing theorizing about fast-and-slow thinking. *Behavioral and Brain Sciences*, 46:e111, 2023. ISSN 0140-525X, 1469-1825. doi: 10.1017/S0140525X2200142X. URL https://www.cambridge.org/core/product/identifier/S0140525X2200142X/type/journal_article.

Wim De Neys. *Dual Process Theory 2.0*. Routledge, London, February 2018. ISBN 978-1-315-20455-0. doi: 10.4324/9781315204550.

Rakefet Ackerman and Valerie A. Thompson. Meta-Reasoning: Monitoring and Control of Thinking and Reasoning. *Trends in Cognitive Sciences*, 21(8):607–617, August 2017. ISSN 1364-6613. doi: 10.1016/j.tics.2017.05.004. URL `https://www.sciencedirect.com/science/article/pii/S1364661317301055`.

Dries Trippas, Simon J. Handley, Michael F. Verde, and Kinga Morsanyi. Logic brightens my day: Evidence for implicit sensitivity to logical validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42:1448–1457, 2016. ISSN 1939-1285. doi: 10.1037/xlm0000248. Place: US Publisher: American Psychological Association.

Keith E. Stanovich. Miserliness in human cognition: the interaction of detection, override and mindware. *Thinking & Reasoning*, 24(4):423–444, October 2018. ISSN 1354-6783. doi: 10.1080/13546783.2018.1459314. URL `https://doi.org/10.1080/13546783.2018.1459314`. Publisher: Routledge _eprint: https://doi.org/10.1080/13546783.2018.1459314.

Valerie F. Reyna, Shahin Rahimi-Golkhandan, David M. N. Garavito, and Rebecca K. Helm. The Fuzzy-Trace Dual Process Model. In *Dual Process Theory 2.0*. Routledge, 2017. Num Pages: 18.

Simon J. Handley and Dries Trippas. Chapter Two - Dual Processes and the Interplay between Knowledge and Structure: A New Parallel Processing Model. In BRIAN H. Ross, editor, *Psychology of Learning and Motivation*, volume 62, pages 33–58. Academic Press, January 2015a. doi: 10.1016/bs.plm.2014.09.002. URL `https://www.sciencedirect.com/science/article/pii/S0079742114000036`.

Valerie A. Thompson and Ian R. Newman. Logical Intuitions and other Conundra for Dual Process Theories. In *Dual Process Theory 2.0*. Routledge, 2017. Num Pages: 16.

Valerie A. Thompson. Why It Matters: The Implications of Autonomous Processes for Dual Process Theories—Commentary on Evans & Stanovich (2013). *Perspectives on Psychological Science*, 8(3):253–256, May 2013. ISSN 1745-6916. doi: 10.1177/1745691613483476. URL `https://doi.org/10.1177/1745691613483476`. Publisher: SAGE Publications Inc.

Alan Baddeley. Working Memory. *Science*, 255(5044):556–559, January 1992. doi: 10.1126/science.1736359. URL `https://www.science.org/doi/abs/10.1126/science.1736359`. Publisher: American Association for the Advancement of Science.

Alan Baddeley. Working memory. *Current Biology*, 20(4):R136–R140, February 2010. ISSN 0960-9822. doi: 10.1016/j.cub.2009.12.014. URL `https://www.`

cell.com/current-biology/abstract/S0960-9822(09)02133-2. Publisher: Elsevier.

Stanislas Dehaene, Michel Kerszberg, and Jean-Pierre Changeux. A Neuronal Model of a Global Workspace in Effortful Cognitive Tasks. *Annals of the New York Academy of Sciences*, 929(1):152–165, 2001. ISSN 1749-6632. doi: 10.1111/j.1749-6632.2001.tb05714.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1749-6632.2001.tb05714.x. _eprint: https://nyaspubs.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1749-6632.2001.tb05714.x.

Wim De Neys and Tamara Glumicic. Conflict monitoring in dual process theories of thinking. *Cognition*, 106(3):1248–1299, March 2008. ISSN 0010-0277. doi: 10.1016/j.cognition.2007.06.002. URL https://www.sciencedirect.com/science/article/pii/S0010027707001576.

Simon J. Handley and Dries Trippas. Dual Processes and the Interplay between Knowledge and Structure: A New Parallel Processing Model. In BRIAN H. Ross, editor, *Psychology of Learning and Motivation*, volume 62, pages 33–58. Academic Press, January 2015b. doi: 10.1016/bs.plm.2014.09.002. URL https://www.sciencedirect.com/science/article/pii/S0079742114000036.

Asher Koriat, Hilit Ma'ayan, and Ravit Nussinson. The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General*, 135:36–69, 2006. ISSN 1939-2222. doi: 10.1037/0096-3445.135.1.36. Place: US Publisher: American Psychological Association.

Valerie A. Thompson and Stephen C. Johnson. Conflict, metacognition, and analytic thinking. *Thinking & Reasoning*, 20(2):215–244, April 2014. ISSN 1354-6783. doi: 10.1080/13546783.2013.869763. URL https://doi.org/10.1080/13546783.2013.869763. Publisher: Routledge _eprint: https://doi.org/10.1080/13546783.2013.869763.

Rakefet Ackerman. The diminishing criterion model for metacognitive regulation of time investment. *Journal of Experimental Psychology: General*, 143 (3):1349–1368, 2014. ISSN 1939-2222. doi: 10.1037/a0035098. Place: US Publisher: American Psychological Association.

Roger Ratcliff. A theory of memory retrieval. *Psychological Review*, 85(2): 59–108, 1978. ISSN 1939-1471. doi: 10.1037/0033-295X.85.2.59. Place: US Publisher: American Psychological Association.

Jerome R Busemeyer and Amnon Rapoport. Psychological models of deferred decision making. *Journal of Mathematical Psychology*, 32(2):91–134, June 1988. ISSN 0022-2496. doi: 10.1016/0022-2496(88)90042-9. URL https://www.sciencedirect.com/science/article/pii/0022249688900429.

Gordon D. Logan and William B. Cowan. On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review*, 91(3):295–327, 1984. ISSN 1939-1471. doi: 10.1037/0033-295X.91.3.295. Place: US Publisher: American Psychological Association.

Marius Usher and James L. McClelland. The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, 108(3):550–592, 2001. ISSN 1939-1471. doi: 10.1037/0033-295X.108.3.550. Place: US Publisher: American Psychological Association.

Benedetto De Martino, Stephen M. Fleming, Neil Garrett, and Raymond J. Dolan. Confidence in value-based choice. *Nature Neuroscience*, 16(1):105–110, January 2013. ISSN 1546-1726. doi: 10.1038/nn.3279. URL `https://www.nature.com/articles/nn.3279`. Publisher: Nature Publishing Group.

Matthieu Raoelison and Wim De Neys. Do we de-bias ourselves?: The impact of repeated presentation on the bat-and-ball problem. *Judgment and Decision Making*, 14(2):170–178, March 2019. ISSN 1930-2975. doi: 10.1017/S1930297500003405.

Alexandra Kuznetsova, Per Bruun Brockhoff, and Rune Haubo Bojesen Christensen. lmerTest: Tests in Linear Mixed Effects Models, January 2013. URL `https://CRAN.R-project.org/package=lmerTest`. Institution: Comprehensive R Archive Network Pages: 3.1-3.

RStudio Team. RStudio: Integrated development environment for R. manual, RStudio, Inc., Boston, MA, 2019. URL `http://www.rstudio.com/`.

Stefano Palminteri, Valentin Wyart, and Etienne Koechlin. The Importance of Falsification in Computational Cognitive Modeling. *Trends in Cognitive Sciences*, 21(6):425–433, June 2017. ISSN 1364-6613, 1879-307X. doi: 10.1016/j.tics.2017.03.011. URL `https://www.cell.com/trends/cognitive-sciences/abstract/S1364-6613(17)30054-2`. Publisher: Elsevier.

Robert C Wilson and Anne GE Collins. Ten simple rules for the computational modeling of behavioral data. *eLife*, 8:e49547, November 2019. ISSN 2050-084X. doi: 10.7554/eLife.49547. URL `https://doi.org/10.7554/eLife.49547`. Publisher: eLife Sciences Publications, Ltd.

Kobe Desender, K Richard Ridderinkhof, and Peter R Murphy. Understanding neural signals of post-decisional performance monitoring: An integrative review. *eLife*, 10:e67556, August 2021. ISSN 2050-084X. doi: 10.7554/eLife.67556. URL `https://elifesciences.org/articles/67556`.

Katharine M. Mullen, David Ardia, David L. Gil, Donald Windover, and James Cline. DEoptim: An R Package for Global Optimization by Differential Evolution. *Journal of Statistical Software*, 40:1–26, April 2011. ISSN 1548-7660. doi: 10.18637/jss.v040.i06. URL `https://doi.org/10.18637/jss.v040.i06`.