# Trading Mental Effort for Confidence:

# The Metacognitive Control of Value-Based Decision-Making

Douglas Lee[1,2], Jean Daunizeau[1,3]

[1] Brain and Spine Institute (ICM), Paris, France

[2] Sorbonne University, Paris, France

[3] Translational Neuromodeling Unit (TNU), ETH, Zurich, Switzerland

Address for correspondence:

Jean Daunizeau

Motivation, Brain, and Behavior Group

Brain and Spine Institute

47, bvd de l'Hopital, 75013, Paris, France

Tel: +33 1 57 27 47 19

E-mail: jean.daunizeau@gmail.com

Keywords:

Word count: 8684

## ABSTRACT

Why do we sometimes opt for actions or items that we do not value the most? Under current neurocomputational theories, such preference reversals are typically interpreted in terms of errors that arise from the unreliable signaling of value to brain decision systems. But, an alternative explanation is that people may change their mind because they are reassessing the value of alternative options while pondering the decision. So, why do we carefully ponder some decisions, but not others? In this work, we derive a computational model of the metacognitive control of decisions or MCD. In brief, we assume that the amount of cognitive resources that is deployed during a decision is controlled by an effort-confidence tradeoff. Importantly, the anticipated benefit of allocating resources varies in a decision-by-decision manner according to decision difficulty and importance. The ensuing MCD model predicts choices, decision time, subjective feeling of effort, choice confidence, and choice-induced preference change. As we will see, these predictions are critically different from accumulation-to-bound models of value-based decisions. We compare and test these predictions in a systematic manner, using a dedicated behavioral paradigm. Our results provides a mechanistic link between mental effort, choice confidence, and preference reversals, which suggests alternative interpretations of existing related neuroimaging findings.

## INTRODUCTION

Why do we carefully ponder some decisions, but not others? Decisions permeate every aspect of our lives—what to eat, where to live, whom to date, etc.—but the amount of effort that we put into different decisions varies tremendously. Rather than processing all decision-relevant information, we often rely on fast habitual and/or intuitive decision policies, which can lead to irrational biases and errors (Kahneman et al., 1982). For example, snap judgments about others are prone to unconscious stereotyping, which often has enduring and detrimental consequences (Greenwald and Banaji, 1995). Yet we don't always follow the fast but negligent lead of habits or intuitions. So, what determines how much time and effort we invest when making decisions?

Biased and/or inaccurate decisions can be triggered by psychobiological determinants such as stress (Porcelli and Delgado, 2009; Porcelli et al., 2012), emotions (Harlé and Sanfey, 2007; Martino et al., 2006; Sokol-Hessner et al., 2013), or fatigue (Blain et al., 2016). But, in fact, they also arise in the absence of such contextual factors. This is why they are sometimes viewed as the outcome of inherent neurocognitive constraints on the brain's decision processes, e.g., limited attentional and/or mnemonic capacity (Giguère and Love, 2013; Lim et al., 2011; Marois and Ivanoff, 2005) or unreliable neural representations of decision-relevant information (Drugowitsch et al., 2016; Wang and Busemeyer, 2016; Wyart and Koechlin, 2016). However, an alternative perspective is that the brain has a preference for efficiency over accuracy (Thorngate, 1980). For example, when making perceptual or motor decisions, people frequently trade accuracy for speed, even when time constraints are not tight (Heitz, 2014; Palmer et al., 2005). Related neural and behavioral data are best explained by "accumulation-to-bound" process models, in which a decision is emitted when the accumulated perceptual evidence reaches a bound (Gold and Shadlen, 2007; O'Connell et al., 2012;

Ratcliff and McKoon, 2008; Ratcliff et al., 2016). Further computational work demonstrated that some variants of these models actually implement an optimal solution to speed-accuracy tradeoff problems (Ditterich, 2006; Drugowitsch et al., 2012). From a theoretical standpoint, this implies that accumulation-to-bound policies can be viewed as an evolutionary adaptation, in response to selective pressure that favors efficiency (Pirrone et al., 2014).

This line of reasoning, however, is not trivial to generalize to value-based decision making, for which objective accuracy remains an elusive notion (Dutilh and Rieskamp, 2016; Rangel et al., 2008). This is because, in contrast to evidence-based (e.g., perceptual) decisions, there are no right or wrong value-based decisions. Nevertheless, people still make choices that deviate from subjective reports of value, with a rate that decreases with value contrast. From the perspective of accumulation-to-bound models, these preference reversals count as errors and arise from the unreliable signaling of value to decision systems in the brain (Lim et al., 2013). That value-based variants of accumulation-to-bound models proved able to capture the neural and behavioral effects of, e.g., overt attention (Krajbich et al., 2010; Lim et al., 2011), external time pressure (Milosavljevic et al., 2010), confidence (De Martino et al., 2012) or default preferences (Lopez-Persem et al., 2016), lent empirical support to this type of interpretation. Further credit also came from theoretical studies showing that these process models, under some simplifying assumptions, optimally solve the problem of efficient value comparison (Tajima et al., 2016, 2019). However, despite the widespread use of these models in decision neuroscience, no evidence of a trial-by-trial accumulation signal has ever been observed in neural recordings in brain systems supporting value-based decisions. In fact, contradictory empirical evidence has even been recently reported in the context of perceptual decisions (Latimer et al., 2015, 2017). In addition, accumulation-to-bound models neglect the possibility that people may reassess the value of alternative options during decisions (Slovic, 1995; Tversky and Thaler, 1990; Warren et al., 2011). For example, contemplating competing possibilities during a choice may highlight features of alternative options that may not have been considered thoroughly before (Sharot et al., 2010). Under this view, apparent preference reversals are not errors: they are deliberate changes of mind. Lastly, accumulation-to-bound models may make nonsensical predictions, in particular with respect to confidence (Lebreton et al., 2015). As we will show below, existing variants of these models that care about choice confidence (De Martino et al., 2013; Tajima et al., 2016) predict that choice confidence should decrease when the reliability of value signals increases! Here, we propose an alternative computational model of value-based decision-making that resolves most of these concerns.

We start with the premise that people are reluctant to make a choice that they are not confident about (De Martino et al., 2013). Thus, when faced with a difficult decision, people reassess option values until they reach a satisfactory level of confidence about their preference. Such effortful mental deliberation engages neurocognitive resources, such as attention and memory, in order to process value-relevant information. In line with recent proposals regarding the strategic deployment of cognitive control (Musslick et al., 2015; Shenhav et al., 2013), we assume that the amount of allocated resources optimizes a tradeoff between expected effort cost and confidence gain. Critically, we show how the system can anticipate the expected benefit of allocating resources before having processed value-relevant information. The ensuing *metacognitive control of decisions* or *MCD* thus adjusts mental effort on a decision-by-decision basis, according to prior decision difficulty and importance. As we will see, the MCD model makes clear quantitative predictions that differ from accumulation-to-bound models. We test these predictions by asking participants to choose between pairs of food items, both before and after having reported their judgment about each item's value and their subjective certainty about value judgements. Note that we also measure choice confidence, decision time, and subjective effort for each decision (cf. Figure 1 below).

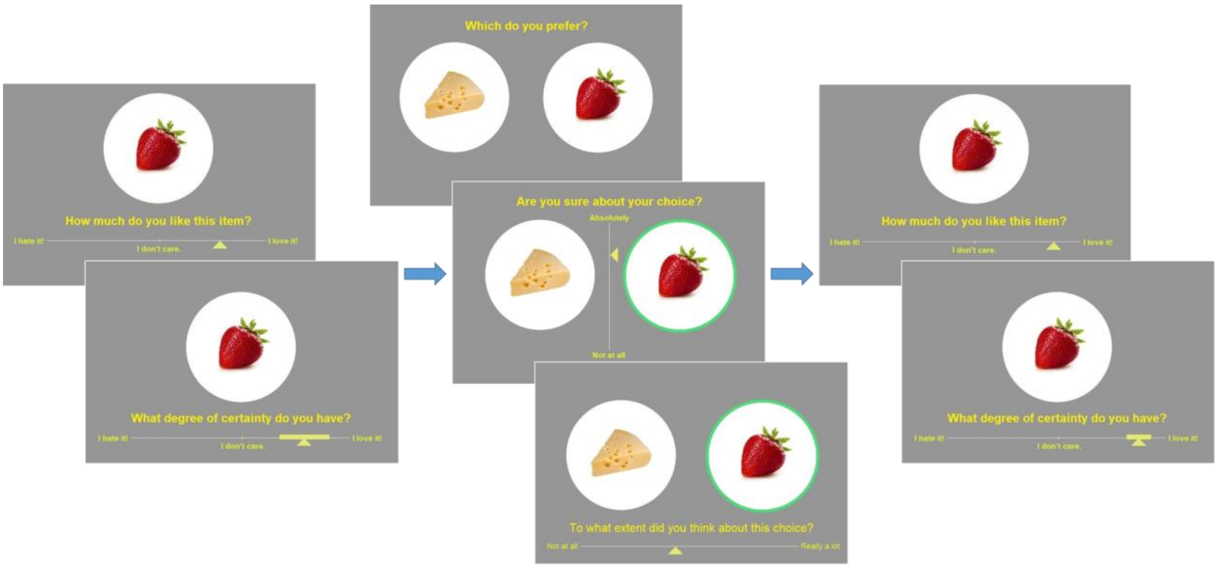


**Figure 1. Experimental design**. Left: pre-choice item rating session: participants are asked to rate how much they like each food item and how certain they are about it (value certainty rating). Center: choice session: participants are asked to choose between two food items, to rate how confident they are about their choice, and to report the feeling of effort associated with the decision. Right: post-choice item rating session (same as pre-choice item rating session).

The MCD model predicts choice, decision time, subjective effort, choice confidence, probability of changing one's mind, and choice-induced preference change on a decision-by-decision basis, out of two properties of pre-choice value representations, namely: value ratings and value certainty ratings. Relevant details regarding the model derivations, as well as the decision-making paradigm we designed to evaluate those predictions, can be found in the Model and Methods sections below. In what follows, we present our main dual computational/behavioral results.

## RESULTS

First, we compare the MCD model to two established models of value-based decision making, namely: an optimal drift-diffusion model with collapsing bounds (Tajima et al., 2016) and a modified race model (De Martino et al., 2013). These two models use variants of the accumulation-to-bound principle, and they can make quantitative predictions regarding the impact of pre-choice value and value certainty ratings (cf. Supplementary Materials). Second, we test a few specific novel predictions that the MCD model makes and that have no analog under alternative frameworks. We note that basic descriptive statistics of our data, including measures of test-retest reliability and replications of previously reported effects on confidence in value-based choices (De Martino et al., 2013), are appended in the Supplementary Materials.

- **Comparing models of decision time and choice confidence**

In what follows, we compare existing computational models of the relationship between choice, value, confidence, and decision time. At this point, suffice it to say that, under accumulation-to-bound models, value uncertainty ratings proxy the magnitude of the stochastic noise in the evidence accumulation process. In contrast, under the MCD model, they simply capture the precision of subjective value representations before the choice. As we will see, all models make rather similar predictions regarding the impact of value ratings. However, they disagree about the impact of value certainty ratings. We will now inspect the three-way relationships between pre-choice value and value certainty ratings and each choice feature (namely: prediction accuracy, decision time, and confidence). Unless stated otherwise, we will focus on both the absolute difference between pre-choice value ratings (hereafter: $|\Delta VR^0|$) and the mean pre-choice value certainty rating across paired choice items (hereafter: $VCR^0$). In each case, we will summarize the empirical data and the corresponding model prediction.

First, we checked how choice prediction accuracy relates to $|\Delta VR^0|$ and $VCR^0$. Here, we measure choice accuracy in terms of the rate of choices that are congruent with preferences derived from pre-choice value ratings $\Delta VR^0$. Under accumulation-to-bound models, choice accuracy should increase with $|\Delta VR^0|$, and decrease with $VCR^0$. This is because the relative impact of stochastic noise on the decision decreases with choice ease, and its magnitude decreases with value certainty ratings. The MCD model makes the same prediction, but for a different reason. In brief, increasing $|\Delta VR^0|$ and/or $VCR^0$ will decrease the demand for effort, which implies that the probability of changing one's mind will be smaller. Figure 2 below shows all quantitative model predictions and summarizes the corresponding empirical data.

One can see that the data seem to conform to the models' predictions. To confirm this, we ran, for each participant, a multiple logistic regression of choice accuracy against $|\Delta VR^0|$ and $VCR^0$. A random effect analysis shows that both have a significant positive effect at the group level ($|\Delta VR^0|$: mean GLM beta=0.17, s.e.m.=0.02, $p<0.001$; $VCR^0$: mean GLM beta=0.07, s.e.m.=0.03, $p=0.004$). Note that people make "inaccurate" choices either because they make mistakes or because they change their mind during the decision. In principle, we can discriminate between these two explanations because we can check whether "inaccurate" choices are congruent with post-choice value ratings (change of mind) or not (error). This is important, because accumulation-to-bound models do not allow for the possibility that value representations change during decisions (hence all "inaccurate" choices would be deemed "errors"). It turns out that, among "inaccurate" choices, mind changes are more frequent than errors (mean rate difference=2.3%, s.e.m.=0.01, $p=0.032$). Note that analyses of mind changes yield qualitatively identical results as choice accuracy (we refer the interested reader to Supplementary Material).

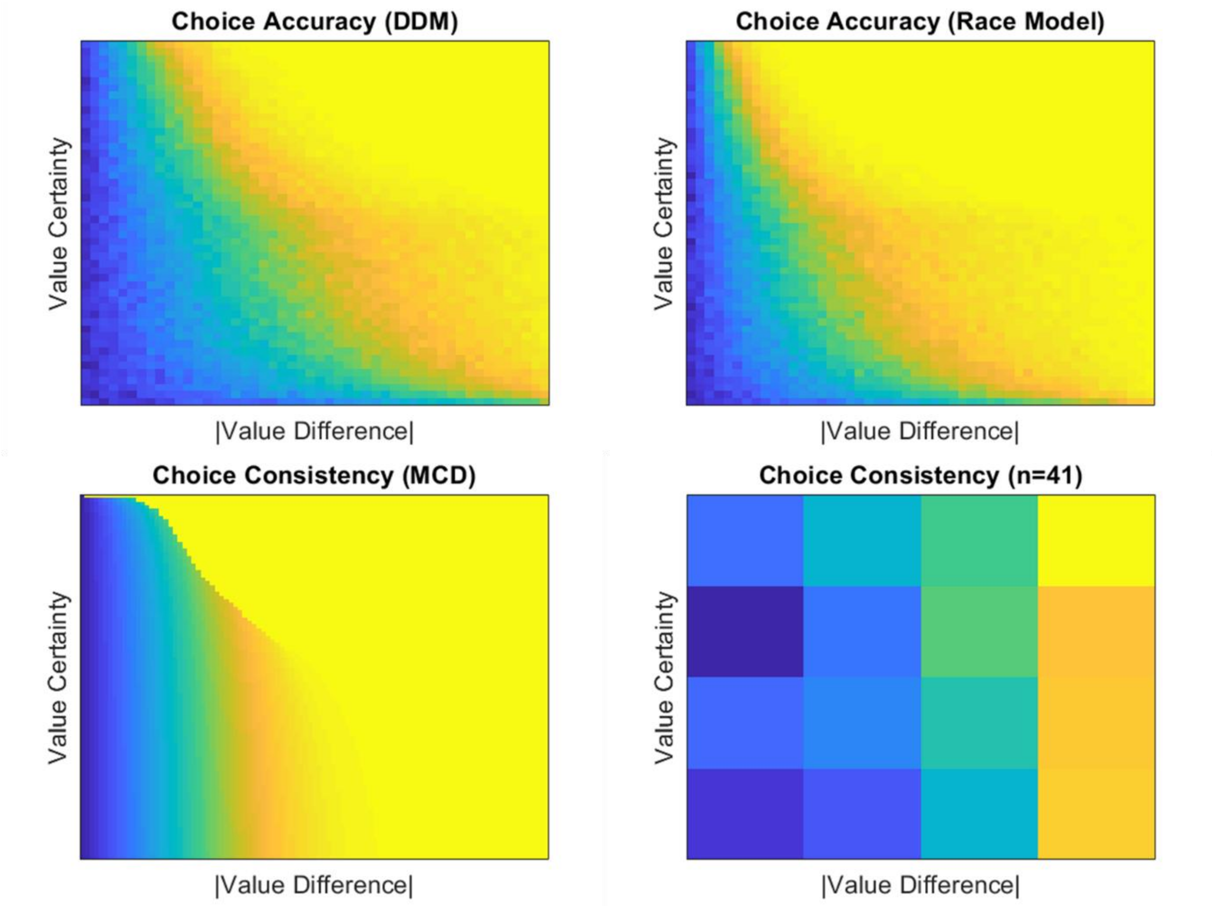


**Figure 2. Three-way relationship between choice accuracy, value, and value certainty**. Upper-left panel: prediction of the DDM model: choice accuracy (color code) is shown as a function of $|\Delta VR^0|$ (x-axis) and $CR^0$ (y-axis). Upper-right panel: prediction of the race model: same format. Lower-left panel: prediction of the MCD model: same format. Lower-right panel: empirical data: same format.

Second, we checked how decision time relates to first- and second-order pre-choice ratings. Under accumulation-to-bound models, decisions are triggered whenever the stochastic evidence accumulation process reaches a predefined threshold. Now, increasing $\Delta VR^0$ effectively increases the drift rate, eventually decreasing the expected decision time. In addition, expected decision time increases with $VCR^0$, because the probability of an early bound hit decreases when the noise magnitude decreases. Under the MCD model, decision time can be thought of as a proxy for effort duration. Here, increasing $|\Delta VR^0|$ and/or $VCR^0$ will decrease the demand for effort, which will result in smaller expected decision time. In other words, the MCD model differs from accumulation-to-bound models with respect to the impact of $VCR^0$ on decision time. Figure 3 below shows all quantitative model predictions and summarizes the corresponding empirical data.

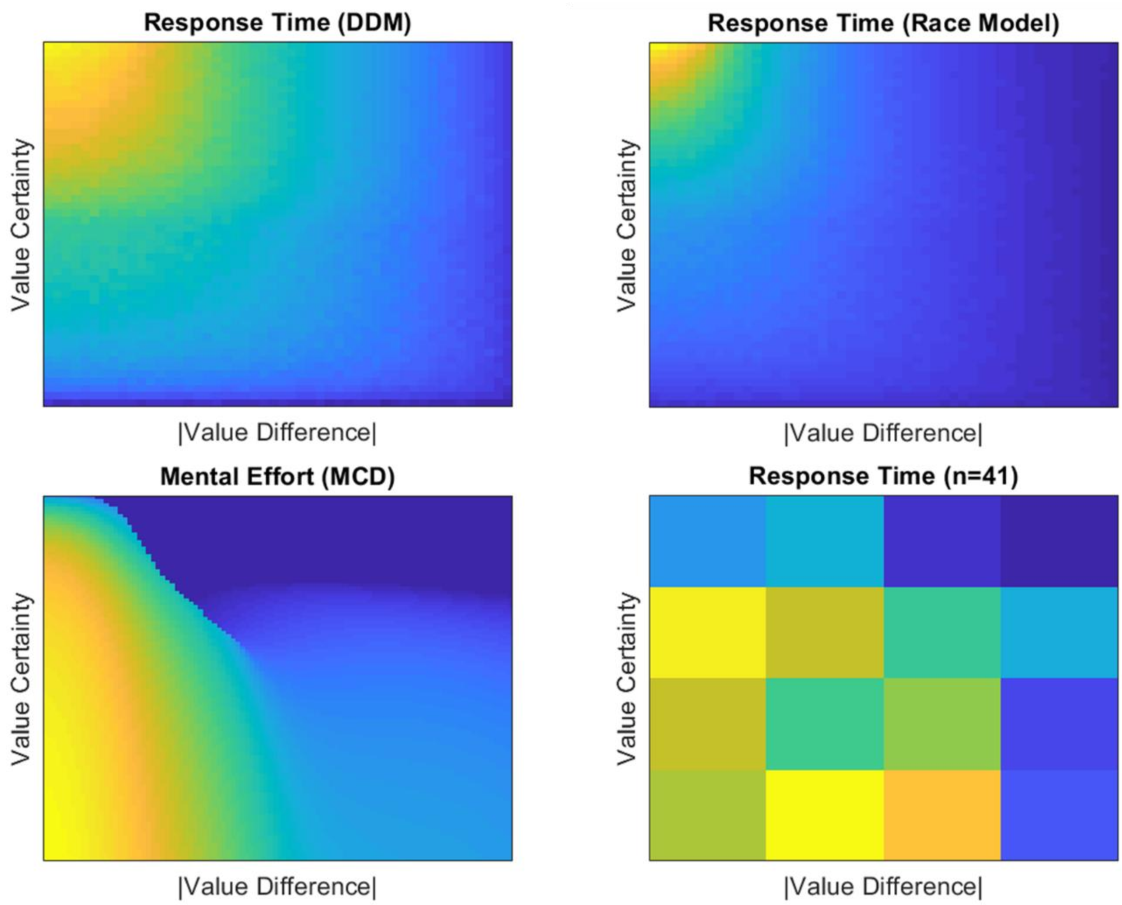


**Figure 3. Three-way relationship between decision time, value, and value certainty.** Upper-left panel: prediction of the DDM model: decision time (color code) is shown as a function of $|\Delta VR^0|$ (x-axis) and $CR^0$ (y-axis). Upper-right panel: prediction of the race model: same format. Lower-left panel: prediction of the MCD model: same format. Lower-right panel: empirical data: same format.

One can see that the decision time data behave as predicted by the MCD model. Here, we also ran, for each participant, a multiple logistic regression of decision times against $|\Delta VR^0|$ and $VCR^0$. A random effect analysis shows that both have a significant and negative effect at the group level ($|\Delta VR^0|$: mean GLM beta=-0.13, s.e.m.=0.02, $p<0.001$; $CR^0$: mean GLM beta=-0.06, s.e.m.=0.02, $p=0.005$).

Third, we checked how choice confidence relates to $|\Delta VR^0|$ and $VCR^0$. Under the DDM model, choice confidence is defined as the height of the optimal collapsing bound when it is hit. Because bounds are collapsing with decision time, confidence increases with $|\Delta VR^0|$ and decreases with $VCR^0$. Under the race model, confidence is defined as the gap between the two value accumulators when the bound is hit. As with the DDM model, increasing $|\Delta VR^0|$ trivially increases confidence. In addition, increasing $VCR^0$ decreases the expected gap between the best and the worst value accumulators (Lebreton et al., 2015). Under the MCD model, confidence reflects the discriminability of value representations after optimal resource allocation. Critically, although more resources are allocated to the decision when either $|\Delta VR^0|$ or $VCR^0$ decrease, this does not overcompensate for decision difficulty, and thus choice confidence decreases. As before, Figure 4 below shows all quantitative model predictions and summarizes the corresponding empirical data.

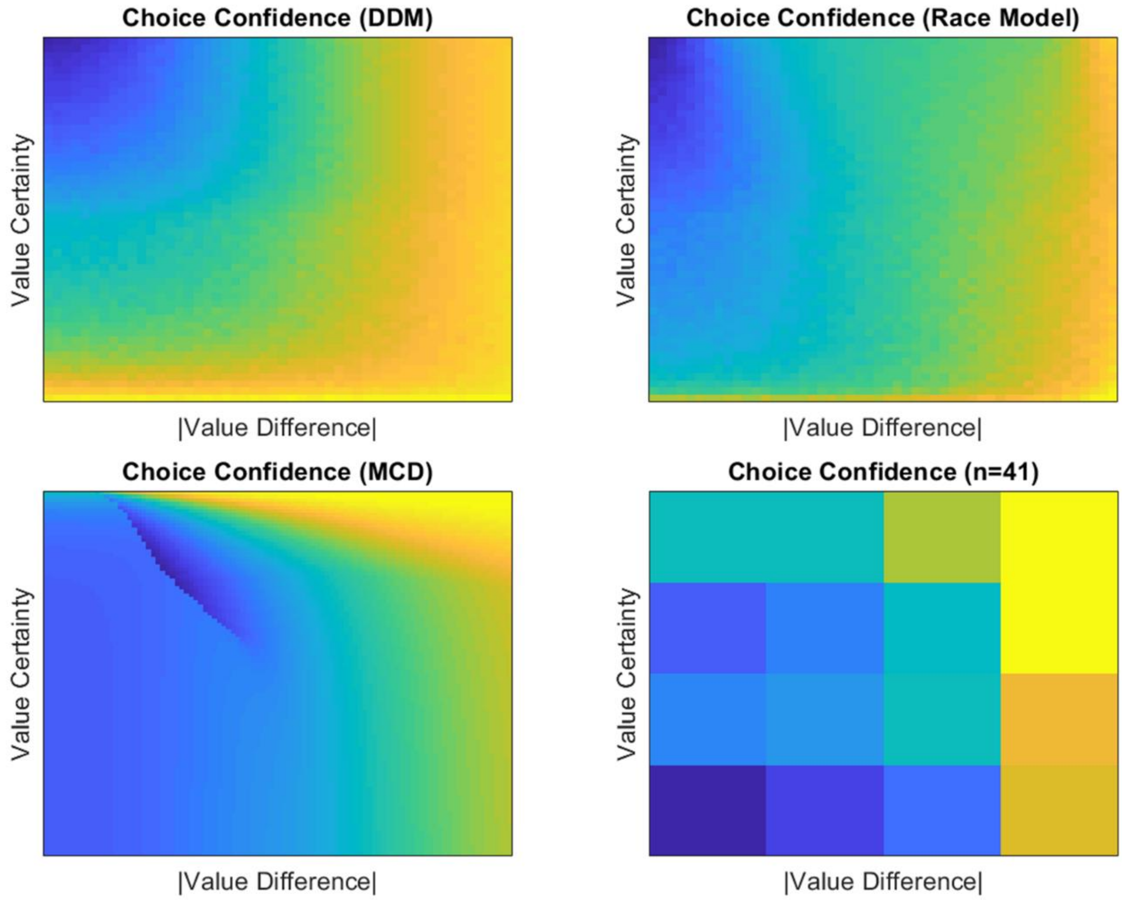


**Figure 4. Three-way relationship between choice confidence, value, and value certainty**. Upper-left panel: prediction of the DDM model: choice confidence (color code) is shown as a function of $|\Delta VR^0|$ (x-axis) and $CR^0$ (y-axis). Upper-right panel: prediction of the race model: same format. Lower-left panel: prediction of the MCD model: same format. Lower-right panel: empirical data: same format.

One can see that the choice confidence follows the MCD model predictions. Again, we ran, for each participant, a multiple logistic regression of confidence against $|\Delta VR^0|$ and $VCR^0$. A random effect analysis shows that both have a significant and positive effect at the group level ($|\Delta VR^0|$: mean GLM beta=0.23, s.e.m.=0.02, p<0.001; $VCR^0$: mean GLM beta=0.15, s.e.m.=0.03, p<0.001). Note that this is unlikely to be a trivial consequence of peoples' decision time readout, since confidence is only mildly correlated with decision time (mean correlation=-0.32, s.e.m.=0.03, p<0.001).

- **Subjective feeling of effort, choice-induced preference change, decision importance, and cost of time**

So far, we have provided evidence that choice confidence and decision time are better explained with the MCD model than with accumulation-to-bound models. In what follows, we will evaluate some additional quantitative predictions that are specific to the MCD model. The derivation of each of these predictions is detailed in the Model section below.

First, recall that MCD really is about the allocation of costly cognitive resources, i.e. mental effort, into the decision process. One may thus ask whether the subjective feeling of effort *per se* follows the MCD predictions. Recall that increasing $|\Delta VR^0|$ and/or $VCR^0$ will decrease the demand for mental resources, which will result in the decision being associated with a lower feeling of effort. To check this, we thus performed a multiple linear regression of subjective effort ratings against $|\Delta VR^0|$ and $VCR^0$. A random effect analysis shows that both have a significant and negative effect at the group level ($|\Delta VR^0|$: mean GLM beta=-0.20, s.e.m.=0.03, $p<0.001$; $CR^0$: mean GLM beta=-0.05, s.e.m.=0.02, $p=0.025$). A graphical summary of the data can be seen in the Supplementary Material.

Second, the MCD model predicts how value representations will be modified during the decision process. In particular, choice-induced preference change should globally follow the optimal effort allocation. More precisely, the reported value of alternative options should spread apart, and the expected spreading of alternatives should be decreasing with $|\Delta VR^0|$ and $VCR^0$. Figure 5 below shows the model predictions and summarizes the corresponding empirical data.

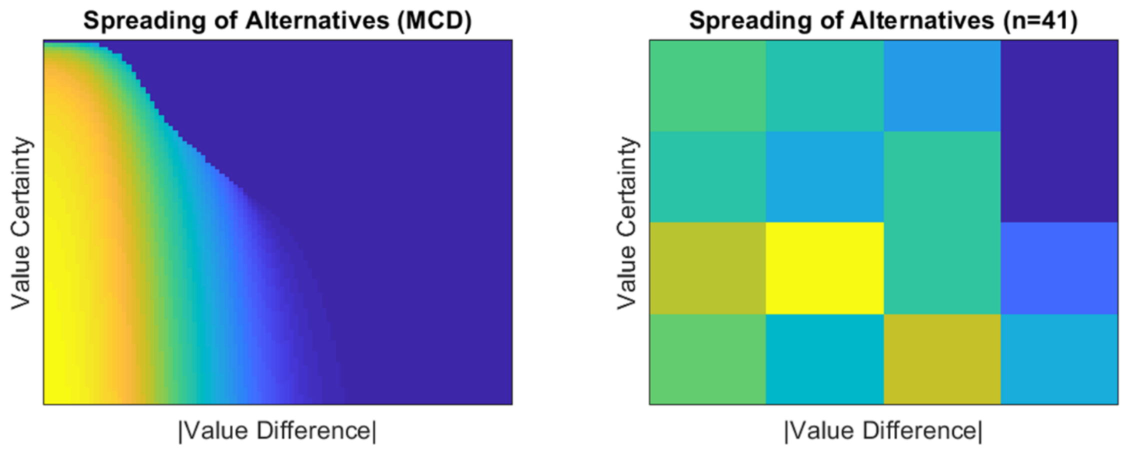


**Figure 5. Three-way relationship between choice-induced preference change, value, and value certainty.** Left panel: prediction of the MCD model: the spreading of alternatives (color code) is shown as a function of $|\Delta VR^0|$ (x-axis) and $VCR^0$ (y-axis). Right panel: empirical data: same format.

One can see that the spreading of alternatives follows the MCD model predictions. A random effect analysis confirms this, showing that both $|\Delta VR^0|$ and $CR^0$ have a significant negative effect at the group level ($|\Delta VR^0|$: mean GLM beta=-0.09, s.e.m.=0.03, p<0.001; $CR^0$: mean GLM beta=-0.04, s.e.m.=0.02, p=0.027). Note that this replicates our previous findings on choice-induced preference change (Lee and Daunizeau, 2019). In addition to expected changes in value ratings, the MCD model predicts that the precision of value representations should increase after the decision has been made (cf. "β-effect" in the Supplementary Material). Indeed, post-choice value certainty ratings are significantly higher than pre-choice value certainty ratings (mean difference=1.34, s.e.m.=0.51, p=0.006). Importantly, under the MCD model, post-choice ratings are simply reports of modified value representations at the time when the choice is triggered. Therefore, choice and its associated confidence level should be better predicted with post-choice ratings than with pre-choice ratings. Indeed, we found that the predictive power of post-choice ratings is significantly higher than that of pre-choice ratings, both for choice (mean prediction accuracy difference=7%, s.e.m.=0.01, p<0.001) and choice confidence (mean prediction accuracy difference=3%, s.e.m.=0.01, p=0.004). Details regarding this analysis can be found in the Supplementary Material.

Third, the MCD model predicts that, all else being equal, effort increases with decision importance and decreases with costs. We checked the former prediction by asking participants to make a few decisions where they knew that the choice would be real, i.e. they would actually have to eat the chosen food item. We refer to these trials as "consequential" decisions. To check the latter prediction, we imposed a financial penalty that increases with decision time. These experimental manipulations are described in the Methods section. Figure 6 below shows subjective effort ratings and decision times for "neutral", "consequential" and "penalized" decisions, when controlling for $|\Delta VR^0|$ and $VCR^0$ (see the Supplementary Material for more details).

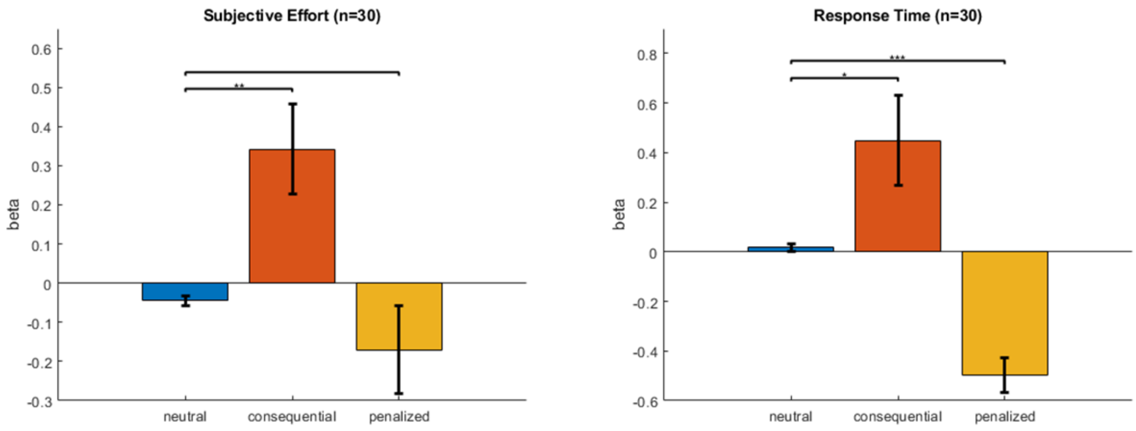


**Figure 6. Comparison of "neutral", "consequential", and "penalized" decisions.** Left: Mean (+/- s.e.m.) effort ratings are shown for "neutral" (blue), "consequential" (red) and "penalized" (yellow) decisions. Right: Mean (+/- s.e.m.) decision time (same format). Both datasets were corrected for $|\Delta VR^0|$ and $VCR^0$.

One can see that subjective effort reports and decision times follow the MCD model predictions. More precisely, both subjective effort reports and decision times were significantly higher for "consequential" decisions than for "neutral" decisions (mean effort difference=0.39, s.e.m.=0.12, p=0.001; mean decision time difference=0.43, s.e.m.=0.19, p=0.017). In addition, decision times are significantly faster for "penalized" than for "neutral" decisions (mean decision time difference=-0.51, s.e.m.=0.08, p<0.001). Note that although the difference in reported effort between "neutral" and "penalized" decisions does not reach statistical significance (mean effort difference=-0.13, s.e.m.=0.12, p=0.147), it goes in the right direction.

## METHODS

Participants for our study were recruited from the RISC (*Relais d'Information sur les Sciences de la Cognition*) subject pool through the ICM (*Institut du Cerveau et de la Moelle épinière*). All participants were native French speakers. All participants were from the non-patient population with no reported history of psychiatric or neurological illness.

Written instructions provided detailed information about the sequence of tasks within the experiment, the mechanics of how participants would perform the tasks, and images illustrating what a typical screen within each task section would look like. The experiment was developed using Matlab and PsychToolbox. The experiment was conducted entirely in French.

Eye gaze position and pupil size were continuously recorded throughout the duration of the experiment using The Eye Tribe eye tracking devices. Participants' head positions were fixed using stationary chinrests. In case of incidental movements, we corrected the pupil size data for distance to screen, separately for each eye.

- **Participants**

A total of 41 people (28 female; age: mean=28, stdev=5, min=20, max=40) participated in this study. The experiment lasted approximately 2 hours, and each participant was paid a flat rate of 20€ as compensation for his time plus an average of 4€ as a bonus. One group of 11 participants was excluded from the cross-condition analysis only (see below), due to technical issues.

- **Materials**

The stimuli for this experiment were 148 digital images, each representing a distinct food item (50 fruits, 50 vegetables, 48 various snack items including nuts, meats, and cheeses). Food items were selected such that most items would be well known to most participants.

- **Procedure**

Prior to commencing the testing session of the experiment, participants underwent a brief training session. The training tasks were identical to the experimental tasks, although different stimuli were used (beverages). The experiment itself began with an initial section where all individual items were displayed in a random sequence for 1.5 seconds each, in order to familiarize the participants with the set of options they would later be considering and form an impression of the range of subjective value for the set. The main experiment was divided into three sections, following the classic Free-Choice Paradigm protocol (Chen and Risen, 2010; Izuma and Murayama, 2013): pre-choice item ratings, choice, and post-choice item ratings (see Figure 1 above). There was no time limit for the overall experiment, nor for the different sections, nor for the individual trials. Item raging and choice sessions are described below.

Item rating (same for pre-choice and post-choice sessions): Participants were asked to rate the entire set of items in terms of how much they liked each item. The items were presented one at a time in a random sequence (pseudo-randomized across participants). At the onset of each trial, a fixation cross appeared at the center of the screen for 750ms. Next, a solitary image of a food item appeared at the center of the screen. Participants had to respond to the question, "How much do you like this item?" using a horizontal slider scale (from "I hate it!" to "I love it!") to indicate their value rating for the item. The middle of the scale was the point of neutrality ("I don't care about it."). Hereafter, we refer to the reported value as the "pre-choice value rating". Participants then had to respond to the question, "How certain are you about the item's value?" by expanding a solid bar symmetrically around the cursor of the value slider scale to indicate the range of possible value ratings that would be compatible with their subjective feeling. We measured participants' certainty about value rating in terms of the percentage of the value scale that is not occupied by the reported range of compatible value ratings. We refer to this as the "pre-choice value certainty rating". At that time, the next trial began.

Choice: Participants were asked to choose between pairs of items in terms of which item they preferred. The entire set of items was presented one pair at a time in a random sequence. Each item appeared in only one pair. At the onset of each trial, a fixation cross appeared at the center of the screen for 750ms. Next, two images of snack items appeared on the screen: one towards the left and one towards the right. Participants had to respond to the question, "Which do you prefer?" using the left or right arrow key. We measured decision time in terms of the delay between the stimulus onset and the response. Participants then had to respond to the question, “Are you sure about your choice?” using a vertical slider scale (from “Not at all!” to “Absolutely!”). We refer to this as the report of choice confidence. Finally, participants had to respond to the question, “To what extent did you think about this choice?” using a horizontal slider scale (from “Not at all!” to “Really a lot!”). We refer to this as the report of subjective effort. At that time, the next trial began.

Note: In the Results section, we refer to $\Delta VR^0$ as the difference between pre-choice value ratings of items composing a choice set. Similarly, $CVR^0$ is the average pre-choice value certainty ratings across items composing a choice set.

- **Conditions**

The choice section of the experiment included trials of three different conditions: "neutral" (60 trials), "consequential" (7 trials), and "penalized" (7 trials), which were randomly intermixed. Immediately prior to each "consequential" trial, participants were instructed that they would be required to eat, at the end of the experiment, a portion of the item that they were about to choose. Immediately prior to each "penalized" trial, participants were instructed that they would lose 0.20€ for each second that they would take to make their choice.

## MODEL

In what follows, we derive a computational model of the metacognitive control of decisions or MCD. In brief, we assume that the amount of cognitive resources that is deployed during a decision is controlled by an effort-confidence tradeoff. Critically, this tradeoff relies on a proactive anticipation of how these resources will perturb the internal representations of subjective values. As we will see, the computational properties of the MCD model are critically different from accumulation-to-bound models of value-based decision-making, which we briefly describe in the Supplementary Material.

- **Deriving the expected value of decision control**

Let $z$ be the amount of cognitive (e.g., executive, mnemonic, or attentional) resources that serve to process value-relevant information. Allocating these resources will be associated with both a benefit $B(z)$, and a cost $C(z)$. As we will see, both are increasing functions of $z$: $B(z)$ derives from the refinement of internal representations of subjective values of alternative options or actions that compose the choice set, and $C(z)$ quantifies how aversive engaging cognitive resources is (mental effort). In line with the framework of expected value of control or EVC (Musslick et al., 2015; Shenhav et al., 2013), we assume that the brain chooses to allocate the amount of resources $\hat{z}$ that optimizes the following cost-benefit trade-off:

$$\hat{z} = \arg\max_{z} E\left[B(z) - C(z)\right] \tag{1}$$

where the expectation accounts for predictable stochastic influences that ensue from allocating resources (this will be more clear below). Note that the benefit term $B(z)$ is the (weighted) choice confidence $P_c(z)$:

$$B(z) = R \times P_c(z) \tag{2}$$

where the weight $R$ is analogous to a reward and quantifies the importance of making a confident decision (see below). As will be made more clear below, $P_c(z)$ plays a pivotal role in the model, in that it captures the efficacy of allocating resources for processing value-relevant information. So, how do we define choice confidence?

We assume that the decision maker may be unsure about how much he likes/wants the alternative options that compose the choice set. In other words, the internal representations of values $V_i$ of alternative options are probabilistic. Such a probabilistic representation of value can be understood in terms of, for example, an uncertain prediction regarding the to-be-experienced value of a given option. Without loss of generality, the probabilistic representation of option values take the form of Gaussian probability density functions, as follows:

$$p(V_i) = N(\mu_i, \sigma_i) \tag{3}$$

where $\mu_i$ and $\sigma_i$ are the mode and the variance of the probabilistic value representations, respectively (and $i$ indexes alternative options in the choice set).

This allows us to define choice confidence $P_c$ as the probability that the (predicted) experienced value of the (to be) chosen item is higher than that of the (to be) unchosen item:

$$\begin{aligned} P_c &= \begin{cases} P(V_1 > V_2) & \text{if item \#1 is chosen} \\ P(V_2 > V_1) & \text{if item \#2 is chosen} \end{cases} \\ &= \begin{cases} P(V_1 > V_2) & \text{if } \Delta\mu > 0 \\ P(V_2 > V_1) & \text{if } \Delta\mu < 0 \end{cases} \\ &\approx s\left( \frac{\pi |\Delta\mu|}{\sqrt{3(\sigma_1 + \sigma_2)}} \right) \end{aligned} \tag{4}$$

where the second line derives from assuming that the choice follows the sign of the preference $\Delta\mu = \mu_1 - \mu_2$, and the last line derives from a moment-matching approximation to the Gaussian cumulative density function (Daunizeau, 2017).

Now, how does the system anticipate the benefit of allocating resources to the decision process? Recall that the purpose of allocating resources is to process (yet unavailable) value-relevant information. The critical issue is thus to predict how both the uncertainty $\sigma_i$ and the modes $\mu_i$ of value representations will change, before having allocated the resources (i.e., without having processed the information). In brief, allocating resources essentially has two impacts: (i) it decreases the uncertainty $\sigma_i$, and (ii) it perturbs the modes $\mu_i$ in a stochastic manner.

The former impact derives from assuming that the amount of information that will be processed increases with the amount of allocated resources. Under simple Bayesian belief update rules, this reduces to stating that the variance of a given probabilistic value representation decreases in proportion to the amount of allocated effort, i.e.:

$$\sigma_i \triangleq \sigma_i(z) = \frac{1}{\frac{1}{\sigma_i^0} + \beta z} \qquad (5)$$

where $\sigma_i^0$ is the prior variance of the representation (before any effort has been allocated), and $\beta$ controls the efficacy with which resources increase the precision of value representations. Formally speaking, Equation 5 has the form of a Bayesian update of the belief variance in a Gaussian-likelihood model, where the precision of the likelihood term is $\beta z$. More precisely, $\beta$ is the precision increase that follows from allocating a unitary amount of resources $z$. In what follows, we will refer to $\beta$ as the "*type #1 effort efficacy*".

The latter impact follows from acknowledging the fact that the system cannot know how processing more value-relevant information will affect its preference before having allocated the corresponding resources. Let $\delta_i(z)$ be the change in the position of the mode of the $i^{\text{th}}$ value representation, having allocated allocating an amount $z$ of resources. The direction of the mode's perturbation $\delta_i(z)$ cannot be predicted because it is tied to the information that would be processed. However, a tenable assumption is to consider that the magnitude of the perturbation increases with the amount of information that will be processed. This reduces to stating that the variance of $\delta_i(z)$ increases in proportion to $z$, i.e.:

$$\begin{aligned} \mu_i(z) &= \mu_i^0 + \delta_i \\ \delta_i &\sim N(0, \gamma z) \end{aligned} \tag{6}$$

where $\mu_i^0$ is the mode of the value representation before any effort has been allocated, and $\gamma$ controls the relationship between the amount of allocated resources and the variance of the perturbation term $\delta$. The higher $\gamma$, the greater the expected perturbation of the mode for a given amount of allocated resources. In what follows, we will refer to $\gamma$ as the "*type #2 effort efficacy*".

Taken together, Equations 5 and 6 imply that predicting the net effect of allocating resources onto choice confidence is not trivial. On the one hand, allocating effort will increase the precision of value representations (cf. Equation 5), which mechanically increases choice confidence, all other things being equal. On the other hand, allocating effort can either increase or decrease the absolute difference $|\Delta\mu(z)|$ between the modes. This, in fact, depends upon the sign of the perturbation terms $\delta$, which are not known in advance. Having said this, it is possible to derive the *expected* absolute difference between the modes that would follow from allocating an amount $z$ of resources:

$$E\left[|\Delta\mu| \| z\right] = 2\sqrt{\frac{\gamma z}{\pi}}\exp\left(-\frac{\left|\Delta\mu^0\right|^2}{4\gamma z}\right) + \Delta\mu^0\left(2\times s\left(\frac{\pi\,\Delta\mu^0}{\sqrt{6\gamma z}}\right) - 1\right) \quad (7)$$

where we have used the expression for the first-order moment of the so-called "folded normal distribution", and the second term in the right-hand side of Equation 7 derives from the same moment-matching approximation to the Gaussian cumulative density function as above. The expected absolute means' difference $E\left[|\Delta\mu| \| z\right]$ depends upon both the absolute prior mean difference $\left|\Delta\mu^0\right|$ and the amount of allocated resources $z$. This is depicted on Figure 7 below.

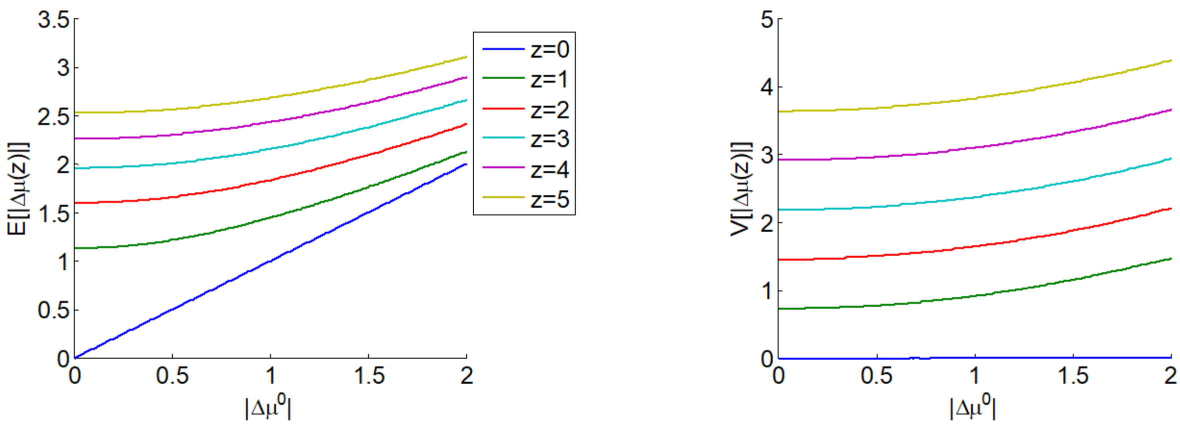


**Figure 7. The expected impact of allocated resources onto value representations.** Left: the expected absolute mean difference $E\left[|\Delta\mu| \| z\right]$ (y-axis) is plotted as a function of the absolute prior mean difference $\left|\Delta\mu^0\right|$ (x-axis) for different amounts $z$ of allocated resources (color code), having set type #2 effort efficacy to unity (i.e. $\gamma = 1$). Right: Variance $V\left[|\Delta\mu| \| z\right]$ of the absolute mean difference ; same format.

One can see that $E\left[|\Delta\mu|\,|\,z\right]-|\Delta\mu^0|$ is always greater than 0 and increases with $z$ (and if $z=0$, then $E\left[|\Delta\mu|\,|\,z\right]=|\Delta\mu^0|$). In other words, allocating resources is expected to increase the value difference, despite the fact that the impact of the perturbation term can go either way. In addition, the expected gain in value difference afforded by allocating resources decreases with the absolute prior means' difference.

Similarly, the variance $V\left[|\Delta\mu|\,|\,z\right]$ of the absolute means' difference is derived from the expression of the second-order moment of the corresponding folded normal distribution:

$$V\left[|\Delta\mu|\,|\,z\right]=2\gamma z+|\Delta\mu^0|^2-E\left[|\Delta\mu|\,|\,z\right]^2 \tag{8}$$

One can see on Figure 7 that $V\left[|\Delta\mu|\,|\,z\right]$ increases with the amount $z$ of allocated resources (but if $z=0$, then $V\left[|\Delta\mu|\,|\,z\right]=0$).

Knowing the moments of the distribution of $|\Delta\mu|$ now enables us to derive the expected confidence level $\bar{P}_c(z)$ that would result from allocating the amount of resource $z$:

$$\begin{aligned}\bar{P}_c(z) &\triangleq E\left[P_c\,|\,z\right]\\ &=E\left[s\left(\frac{\pi|\Delta\mu|}{\sqrt{6\sigma(z)}}\right)\middle|\,z\right]\\ &\approx s\left(\frac{\pi E\left[|\Delta\mu|\,|\,z\right]}{\sqrt{6\left(\sigma(z)+\frac{1}{2}V\left[|\Delta\mu|\,|\,z\right]\right)}}\right)\end{aligned} \tag{9}$$

where we have assumed, for the sake of conciseness, that both prior value representations are similarly uncertain (i.e., $\sigma_1^0 \approx \sigma_2^0 \triangleq \sigma^0$). It turns out that the expected choice confidence $\bar{P}_c(z)$ always increase with $z$, irrespective of the efficacy parameters $\beta$ and $\gamma$. These, however, control the magnitude of the confidence gain that can be expected from allocating an amount $z$ of resources. Equation 9 is important, because it quantifies the expected benefit of resource allocation, before having processed the ensuing value-relevant information. More details regarding the accuracy of Equation 9 can be found in the Supplementary Material.

To complete the cost-benefit model, and without loss of generality, we will assume that the cost of allocating resources to the decision process linearly scales with the amount of resources, i.e.:

$$C(z) = \alpha z \quad (10)$$

where $\alpha$ determines the effort cost of allocating a unitary amount of resources $z$. In what follows, we will refer to $\alpha$ as the "effort unitary cost". We refer to $\alpha$ as the "effort unitary cost".

In brief, the MCD-optimal resource allocation $\hat{z} \triangleq \hat{z}(\alpha, \beta, \gamma)$ is simply given by:

$$\hat{z} = \arg\max_{z} \left[ R \times \bar{P}_c(z) - \alpha z \right] \quad (11)$$

which does not have any closed-form analytic solution. Nevertheless, it can easily be identified numerically, having replaced Equations 7-9 into Equation 11. We refer the readers interested in the impact of model parameters $\{\alpha, \beta, \gamma\}$ on the MCD-optimal control to the Supplementary Material.

<u>Note</u>: at this point, the MCD model is agnostic about what the allocated resource is. Empirically, we relate $\hat{z}$ to two different measures, namely: decision time and the subjective feeling of effort. The former makes sense if one thinks of decision time in terms of effort duration, which increases the cumulative engagement of

neurocognitive resources. The latter relies on the subjective cost incurred when deploying neurocognitive resources, which would be signaled by experiencing mental effort. We will comment on this in the Discussion section. Also, implicit in the above model derivation is the assumption that the allocation of resources is similar for both alternative options in the choice set (i.e. $z_1 \approx z_2 \triangleq z$). This simplifying assumption is justified by eye-tracking data (cf. Supplementary Material). Finally, we investigate the effect of decision importance by comparing effort and decision time in "neutral" versus "consequential" decisions (cf. Methods section).

- **Corollary predictions of the MCD model**

In the previous section, we derived the MCD-optimal resource allocation, which effectively best balances the expected choice confidence with the expected effort costs, given the predictable impact of stochastic perturbations that arise from processing value-relevant information. This quantitative prediction is effectively shown on Figure 3 (and/or Figure S4 of the Supplementary Material), as a function of (empirical proxies for) the prior absolute difference between modes $|\Delta\mu^0|$ and the prior certainty $1/\sigma^0$ of value representations. But, this mechanism has a few interesting corollary implications.

To begin with, note that knowing $\hat{z}$ enables us to predict what confidence level the system should reach. In fact, one can define the MCD-optimal confidence level as the expected confidence evaluated at the MCD-optimal amount of allocated resources, i.e., $\bar{P}_c(\hat{z})$. This is important, because it implies that the model can predict both the effort the system invests and its associated confidence, on a decision-by-decision basis. This quantitative prediction is shown on Figure 4.

Similarly, one can predict the MCD-optimal probability of changing one's mind. Recall that the probability $Q(z)$ of changing one's mind depends on the amount of allocated resources $z$, i.e.:

$$\begin{aligned} Q(z) &\triangleq P\left(sign(\Delta\mu) \neq sign\left(\Delta\mu^0\right) \middle| z\right) \\ &= \begin{cases} P\left(\Delta\mu > 0 \middle| z\right) & \text{if } \Delta\mu^0 < 0 \\ P\left(\Delta\mu < 0 \middle| z\right) & \text{if } \Delta\mu^0 > 0 \end{cases} \\ &\approx s\left(-\frac{\pi\left|\Delta\mu^0\right|}{\sqrt{6\gamma z}}\right) \end{aligned} \quad (12)$$

One can see that the MCD-optimal probability of changing one's mind $Q(\hat{z})$ is a simple monotonic function of the allocated effort $\hat{z}$. Note that, by definition, choice accuracy (i.e., congruence of choice and prior preference $\Delta\mu^0$) is but $1-Q(z)$, which is shown on Figure 2.

Lastly, we can predict choice-induced preference change, i.e., how value representations are supposed to spread apart during the decision. Such an effect is typically measured in terms of the so-called "spreading of alternatives" or SoA, which is defined as follows:

$$\begin{aligned} SOA &= \left(\mu_{chosen}^{(post-choice)} - \mu_{unchosen}^{(post-choice)}\right) - \left(\mu_{chosen}^{(pre-choice)} - \mu_{unchosen}^{(pre-choice)}\right) \\ &= \begin{cases} \Delta\mu(z) - \Delta\mu^0 & \text{if } \Delta\mu(z) > 0 \\ \Delta\mu^0 - \Delta\mu(z) & \text{if } \Delta\mu(z) < 0 \end{cases} \\ &= \begin{cases} \Delta\delta(z) & \text{if } \Delta\delta(z) > -\Delta\mu^0 \\ -\Delta\delta(z) & \text{if } \Delta\delta(z) < -\Delta\mu^0 \end{cases} \end{aligned} \quad (13)$$

where $\Delta\delta(z) \sim N(0, 2\gamma z)$ is the cumulative perturbation term of the modes' difference. Taking the expectation of the right-hand term of Equation 13 under the distribution of $\Delta\delta(z)$ and evaluating it at $z = \hat{z}$ now yields the MCD-optimal spreading of alternatives $\overline{SOA}(\hat{z})$:

$$\begin{aligned} \overline{SOA}(\hat{z}) &= E\left[SOA \middle| \hat{z}\right] \\ &= E\left[\Delta\delta(\hat{z}) \middle| \Delta\delta(\hat{z}) > -\Delta\mu^0\right] P\left(\Delta\delta(\hat{z}) > -\Delta\mu^0\right) \\ &\quad - E\left[\Delta\delta(\hat{z}) \middle| \Delta\delta(\hat{z}) < -\Delta\mu^0\right] P\left(\Delta\delta(\hat{z}) < -\Delta\mu^0\right) \\ &= 2\sqrt{\frac{\gamma\hat{z}}{\pi}} \exp\left(-\frac{\left|\Delta\mu^0\right|^2}{4\gamma\hat{z}}\right) \end{aligned} \tag{14}$$

where the last line derives from the expression of the first-order moment of the truncated Gaussian distribution. Note that the expected preference change also increases monotonically with the allocated effort $\hat{z}$.

In summary, the MCD model predicts, given the prior absolute difference between modes $\left|\Delta\mu^0\right|$ and the prior certainty $1/\sigma^0$ of value representations, choice accuracy, choice confidence, choice-induced preference change, decision time and/or subjective feelings of effort. Note that, when testing the decision-by-decision predictions of the MCD model, we use $\Delta VR^0$ and $CVR^0$ as empirical proxies for $\Delta\mu^0$ and $1/\sigma^0$, respectively.

## DISCUSSION

In this work, we have presented a novel computational model of decision-making which explains the intricate relationships between choice accuracy, decision time, subjective effort, choice confidence, and choice-induced preference change. This model assumes that deciding between alternative options whose values are uncertain induces a demand for allocating cognitive resources to processing value-relevant information. Cognitive resource allocation then optimally trades effort for confidence, given the discriminability of prior value representations. Such metacognitive control of decisions or MCD makes novel predictions that differ from standard accumulation-to-bound models of decision-making, including a drift-diffusion model that was proposed as an optimal policy for value-based decision making (Tajima et al., 2016). But, how can these two frameworks both be optimal? The answer lies in the distinct computational problems that they solve. The MCD solves the problem of finding the optimal amount of effort to invest under the possibility that yet-unprocessed value-relevant information might change the decider's mind. In fact, this resource allocation problem would be vacuous, would it not be possible to reassess preferences during the decision process. In contrast, the DDM provides an optimal solution to the problem of efficiently comparing option values, which may be unreliably signaled, but remain stationary nonetheless. This why the DDM cannot predict choice-induced preference changes. This critical distinction extends to other types of accumulation-to-bound models, including race models ( De Martino et al, 2013; Tajima et al, 2019).

Now, let us highlight that the MCD model offers a plausible alternative interpretation for the two main reported neuroimaging findings regarding confidence in value-based choices (De Martino et al., 2013). First, the ventromedial prefrontal cortex or vmPFC was found to respond positively to both value difference (i.e., $\Delta VR^0$) and choice confidence. Second, the right rostrolateral prefrontal cortex or rRLPFC was more active during low-confidence versus high-confidence choices. These findings were originally interpreted through the framework of the race model that we compared to the MCD model. In brief, rRLPFC was thought to perform a readout of choice confidence (for the purpose of subjective metacognitive report) from the racing value accumulators hosted in the vmPFC. Under the MCD framework, the contribution of the vmPFC to value-based choices might rather be to anticipate and monitor the benefit of effort investment (i.e., confidence). This would be consistent with recent fMRI studies suggesting that vmPFC confidence computations signal the attainment of task goals (Hebscher and Gilboa, 2016; Lebreton et al., 2015). Now, recall that the MCD model predicts that confidence and effort should be anti-correlated. Thus, the puzzling negative correlation between choice confidence and rRLPFC activity could be simply explained under the assumption that rRLPFC provides the neurocognitive resources that are instrumental for processing value-relevant information during decisions. This resonates with the known involvement of rRLPFC in reasoning (Desrochers et al., 2015; Dumontheil, 2014) or memory retrieval (Benoit et al., 2012; Westphal et al., 2019).

At this point, we would like to discuss a few features of the MCD model. First, we did not specify what determines the reward component, which quantifies decision importance and acts as an effective weight for confidence against effort costs (cf. $R$ in Equation 2 of the Model section). We know, from the comparison of “consequential” and “neutral” choices that increasing decision importance eventually increases effort, as predicted by the MCD model. However, decision importance may have many determinants, such as, for example, the commitment time of the decision (cf. partner choices), the breadth of its repercussions (cf. political decisions), or its instrumentality with respect to the achievement of superordinate goals (cf. moral decisions). How these determinants are combined and/or moderated by the decision context is virtually unknown (Locke and Latham, 2002, 2006). In addition, decision importance might also be influenced by the prior (intuitive/emotional/habitual) appraisal of option values. For example, we found that, all else equal, people spent much more time and effort deciding between two disliked items than between two liked items (results not shown). This reproduces recent results regarding the evaluation of choice sets (Shenhav and Karmarkar, 2019). Probing this type of influence will be the focus of forthcoming publications.

Second, our current version of the MCD model relies upon a simple variant of resource costs. We note that rendering the cost term nonlinear (e.g., quadratic) does not change the qualitative nature of the MCD model predictions. More problematic, perhaps, is the fact that we did not consider distinct types of effort, which could, in principle, be associated with different costs. For example, the cost of allocating attention to a given option may depend upon whether this option would be a priori chosen or not. This might eventually explain systematic decision biases and differences in decision times between default and non-default choices (Lopez-Persem et al., 2016). Another possibility is that effort might be optimized along two canonical dimensions, namely: duration and intensity. The former dimension essentially justifies the fact that we used decision time as a proxy for cognitive effort. In fact, as is evident from the comparison between “penalized” and “neutral” choices, imposing an external penalty cost on decision time reduces, as expected, the ensuing subjective effort. More generally, however, the dual optimization of effort dimensions might render the relationship between effort and decision time more complex. For example, beyond memory span or attentional load, effort intensity could be related to processing speed. This would explain why, although "penalized" choices are made much faster than "neutral" choices, the associated feeling of effort is not strongly impacted (cf. Figure 6). In any case, the relationship between effort and decision time might depend upon the relative costs of effort duration and intensity, which might itself be partially driven by external availability constraints (cf. time pressure or multitasking). We note that the essential nature of the cost of mental effort in cognitive tasks (e.g., neurophysiological cost, interferences cost, opportunity cost) is still a matter of intense debate (Kurzban et al., 2013; Musslick et al., 2015; Ozcimder et al., 2017). Progress towards addressing this issue will be highly relevant for future extensions of the MCD model.

Third, we did not consider the issue of identifying plausible neuro-computational implementations of MCD. This issue is tightly linked to the previous one, in that distinct cost types would likely impose different constraints on candidate neural network architectures (Feng et al., 2014; Petri et al., 2017). For example, underlying brain circuits are likely to operate MCD in a more dynamic manner, eventually adjusting resource allocation from the continuous monitoring of relevant decision variables (e.g., experienced costs and benefits). Such a reactive process contrasts with our current, proactive-only, variant of MCD, which sets resource allocation based on anticipated costs and benefits. We already checked that simple reactive scenarios, where the decision is triggered whenever the online monitoring of effort or confidence reaches the optimal threshold, make predictions qualitatively similar to those we have presented here. We tend to think however, that such reactive processes should be based upon a dynamic programming perspective on MCD, as was already done for the problem of optimal efficient value comparison (Tajima et al., 2016, 2019). We will pursue this and related neuro-computational issues in subsequent publications.

## REFERENCES


Benoit, R.G., Gilbert, S.J., Frith, C.D., and Burgess, P.W. (2012). Rostral Prefrontal Cortex and the Focus of Attention in Prospective Memory. Cereb. Cortex *22*, 1876–1886.

Blain, B., Hollard, G., and Pessiglione, M. (2016). Neural mechanisms underlying the impact of daylong cognitive work on economic decisions. Proc. Natl. Acad. Sci. *113*, 6967–6972.

Chen, K.M., and Risen, J.L. (2010). How choice affects and reflects preferences: Revisiting the free-choice paradigm. J. Pers. Soc. Psychol. *99*, 573–594.

Daunizeau, J. (2017). Semi-analytical approximations to statistical moments of sigmoid and softmax mappings of normal variables. ArXiv170300091 Q-Bio Stat.

De Martino, B., Fleming, S.M., Garrett, N., and Dolan, R.J. (2013). Confidence in value-based choice. Nat. Neurosci. *16*, 105–110.

Desrochers, T.M., Chatham, C.H., and Badre, D. (2015). The necessity of rostrolateral prefrontal cortex for higher-level sequential behavior. Neuron *87*, 1357–1368.

Ditterich, J. (2006). Evidence for time-variant decision making. Eur. J. Neurosci. *24*, 3628–3641.

Drugowitsch, J., Moreno-Bote, R., Churchland, A.K., Shadlen, M.N., and Pouget, A. (2012). The Cost of Accumulating Evidence in Perceptual Decision Making. J. Neurosci. *32*, 3612–3628.

Drugowitsch, J., Wyart, V., Devauchelle, A.-D., and Koechlin, E. (2016). Computational Precision of Mental Inference as Critical Source of Human Choice Suboptimality. Neuron *92*, 1398–1411.

Dumontheil, I. (2014). Development of abstract thinking during childhood and adolescence: The role of rostrolateral prefrontal cortex. Dev. Cogn. Neurosci. *10*, 57–76.

Dutilh, G., and Rieskamp, J. (2016). Comparing perceptual and preferential decision making. Psychon. Bull. Rev. *23*, 723–737.

Feng, S.F., Schwemmer, M., Gershman, S.J., and Cohen, J.D. (2014). Multitasking versus multiplexing: Toward a normative account of limitations in the simultaneous execution of control-demanding behaviors. Cogn. Affect. Behav. Neurosci. *14*, 129–146.

Giguère, G., and Love, B.C. (2013). Limits in decision making arise from limits in memory retrieval. Proc. Natl. Acad. Sci. *110*, 7613–7618.

Gold, J.I., and Shadlen, M.N. (2007). The neural basis of decision making. Annu. Rev. Neurosci. *30*, 535–574.

Greenwald, A.G., and Banaji, M.R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. Psychol. Rev. *102*, 4–27.

Harlé, K.M., and Sanfey, A.G. (2007). Incidental sadness biases social economic decisions in the Ultimatum Game. Emot. Wash. DC *7*, 876–881.

Hebscher, M., and Gilboa, A. (2016). A boost of confidence: The role of the ventromedial prefrontal cortex in memory, decision-making, and schemas. Neuropsychologia *90*, 46–58.

Heitz, R.P. (2014). The speed-accuracy tradeoff: history, physiology, methodology, and behavior. Front. Neurosci. *8*.

Izuma, K., and Murayama, K. (2013). Choice-Induced Preference Change in the Free-Choice Paradigm: A Critical Methodological Review. Front. Psychol. *4*.

Kahneman, D., Slovic, P., and Tversky, A. (1982). Judgment Under Uncertainty: Heuristics and Biases (Cambridge University Press).

Krajbich, I., Armel, C., and Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. Nat. Neurosci. *13*, 1292–1298.

Kurzban, R., Duckworth, A., Kable, J.W., and Myers, J. (2013). An opportunity cost model of subjective effort and task performance. Behav. Brain Sci. *36*, 661–679.

Latimer, K.W., Yates, J.L., Meister, M.L.R., Huk, A.C., and Pillow, J.W. (2015). Single-trial Spike Trains in Parietal Cortex Reveal Discrete Steps During Decision-making. Science *349*, 184–187.

Latimer, K.W., Huk, A.C., and Pillow, J.W. (2017). No cause for pause: new analyses of ramping and stepping dynamics in LIP (Rebuttal to Response to Reply to Comment on Latimer et al 2015). BioRxiv 160994.

Lebreton, M., Abitbol, R., Daunizeau, J., and Pessiglione, M. (2015). Automatic integration of confidence in the brain valuation signal. Nat. Neurosci. *18*, 1159–1167.

Lee, D., and Daunizeau, J. (2019). Choosing what we like vs liking what we choose: How choice-induced preference change might actually be instrumental to decision-making. BioRxiv 661116.

Lim, S.-L., O'Doherty, J.P., and Rangel, A. (2011). The Decision Value Computations in the vmPFC and Striatum Use a Relative Value Code That is Guided by Visual Attention. J. Neurosci. *31*, 13214–13223.

Lim, S.-L., O'Doherty, J.P., and Rangel, A. (2013). Stimulus Value Signals in Ventromedial PFC Reflect the Integration of Attribute Value Signals Computed in Fusiform Gyrus and Posterior Superior Temporal Gyrus. J. Neurosci. *33*, 8729–8741.

Locke, E.A., and Latham, G.P. (2002). Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. Am. Psychol. *57*, 705–717.

Locke, E.A., and Latham, G.P. (2006). New Directions in Goal-Setting Theory. Curr. Dir. Psychol. Sci. *15*, 265–268.

Lopez-Persem, A., Domenech, P., and Pessiglione, M. (2016). How prior preferences determine decision-making frames and biases in the human brain. ELife *5*, e20317.

Marois, R., and Ivanoff, J. (2005). Capacity limits of information processing in the brain. Trends Cogn. Sci. *9*, 296–305.

Martino, B.D., Kumaran, D., Seymour, B., and Dolan, R.J. (2006). Frames, Biases, and Rational Decision-Making in the Human Brain. Science *313*, 684–687.

Milosavljevic, M., Malmaud, J., Huth, A., Koch, C., and Rangel, A. (2010). The drift diffusion model can account for value-based choice response times under high and low time pressure. Judgm. Decis. Mak. *5*, 437–449.

Musslick, S., Shenhav, A., Botvinick, M., and D Cohen, J. (2015). A Computational Model of Control Allocation based on the Expected Value of Control. p.

O'Connell, R.G., Dockree, P.M., and Kelly, S.P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. Nat. Neurosci. *15*, 1729–1735.

Ozcimder, K., Dey, B., Musslick, S., Petri, G., Ahmed, N.K., Willke, T.L., and Cohen, J.D. (2017). A Formal Approach to Modeling the Cost of Cognitive Control. ArXiv170600085 Q-Bio.

Palmer, J., Huk, A.C., and Shadlen, M.N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. J. Vis. *5*, 376–404.

Petri, G., Musslick, S., Dey, B., Ozcimder, K., Ahmed, N.K., Willke, T., and Cohen, J.D. (2017). Universal limits to parallel processing capability of network architectures. ArXiv170803263 Q-Bio.

Pirrone, A., Stafford, T., and Marshall, J.A.R. (2014). When natural selection should optimize speed-accuracy trade-offs. Front. Neurosci. *8*.

Porcelli, A.J., and Delgado, M.R. (2009). Acute stress modulates risk taking in financial decision making. Psychol. Sci. *20*, 278–283.

Porcelli, A.J., Lewis, A.H., and Delgado, M.R. (2012). Acute Stress Influences Neural Circuits of Reward Processing. Front. Neurosci. *6*.

Rangel, A., Camerer, C., and Montague, P.R. (2008). A framework for studying the neurobiology of value-based decision making. Nat. Rev. Neurosci. *9*, 545–556.

Ratcliff, R., and McKoon, G. (2008). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. Neural Comput. *20*, 873–922.

Ratcliff, R., Smith, P.L., Brown, S.D., and McKoon, G. (2016). Diffusion Decision Model: Current Issues and History. Trends Cogn. Sci. *20*, 260–281.

Sharot, T., Velasquez, C.M., and Dolan, R.J. (2010). Do Decisions Shape Preference? Evidence From Blind Choice. Psychol. Sci. *21*, 1231–1235.

Shenhav, A., & Karmarkar, U. R. (2019). Dissociable components of the reward circuit are involved in appraisal versus choice. *Scientific reports*, *9*(1), 1958.

Shenhav, A., Botvinick, M.M., and Cohen, J.D. (2013). The Expected Value of Control: An Integrative Theory of Anterior Cingulate Cortex Function. Neuron *79*, 217–240.

Slovic, P. (1995). The construction of preference. Am. Psychol. *50*, 364–371.

Sokol-Hessner, P., Camerer, C.F., and Phelps, E.A. (2013). Emotion regulation reduces loss aversion and decreases amygdala responses to losses. Soc. Cogn. Affect. Neurosci. *8*, 341–350.

Tajima, S., Drugowitsch, J., and Pouget, A. (2016). Optimal policy for value-based decision-making. Nat. Commun. *7*, 12400.

Tajima, S., Drugowitsch, J., Patel, N., and Pouget, A. (2019). Optimal policy for multi-alternative decisions. Nat. Neurosci. *22*, 1503–1511.

Thorngate, W. (1980). Efficient decision heuristics. Behav. Sci. *25*, 219–225.

Tversky, A., and Thaler, R.H. (1990). Anomalies: Preference Reversals. J. Econ. Perspect. *4*, 201–211.

Wang, Z., and Busemeyer, J.R. (2016). Interference effects of categorization on decision making. Cognition *150*, 133–149.

Warren, C., McGraw, A.P., and Van Boven, L. (2011). Values and preferences: defining preference construction. Wiley Interdiscip. Rev. Cogn. Sci. *2*, 193–205.

Westphal, A.J., Chow, T.E., Ngoy, C., Zuo, X., Liao, V., Storozuk, L.A., Peters, M.A.K., Wu, A.D., and Rissman, J. (2019). Anodal Transcranial Direct Current Stimulation to the Left Rostrolateral Prefrontal Cortex Selectively Improves Source Memory Retrieval. J. Cogn. Neurosci. *31*, 1380–1391.

Wyart, V., and Koechlin, E. (2016). Choice variability and suboptimality in uncertain environments. Curr. Opin. Behav. Sci. *11*, 109–115.

# Trading Mental Effort for Confidence: Supplementary Material

## 1. Data descriptive statistics and sanity checks

Recall that we collect value ratings and value certainty ratings both before and after the choice session. We did this for the purpose of validating specific predictions of the MCD model (in particular: choice-induced preference changes: see Figure 5 of the main text). It turns out this also enables us to assess the test-retest reliability of both value and value certainty ratings. We found that both ratings were significantly reproducible (value: mean correlation=0.88, s.e.m.=0.01, p <0.001, value certainty: mean correlation=0.37, s.e.m.=0.04, p <0.001).

We also checked whether choices were consistent with pre-choice ratings. For each participant, we thus preformed a logistic regression of choices against the difference in value ratings. We found that the balanced prediction accuracy was beyond chance level (mean accuracy=0.68, s.e.m.=0.01, p<0.001).

## 2. Does choice confidence moderate the relationship between choice and pre-choice value ratings?

Previous studies regarding confidence in value-base choices showed that choice confidence moderates choice prediction accuracy (De Martino et al., 2013). We thus splat our logistic regression of choices into high- and low-confidence trials, and tested whether higher confidence was consistently associated with increased choice accuracy. A random effect analysis showed that the regression slopes were significantly higher for high- than for low-confidence trials (mean slope difference=0.14, s.e.m.=0.03, p<0.001). For the sake of completeness, the impact of choice confidence on the slope of the logistic regression (of choice onto the difference in pre-choice value ratings) is shown on Figure S1 below.

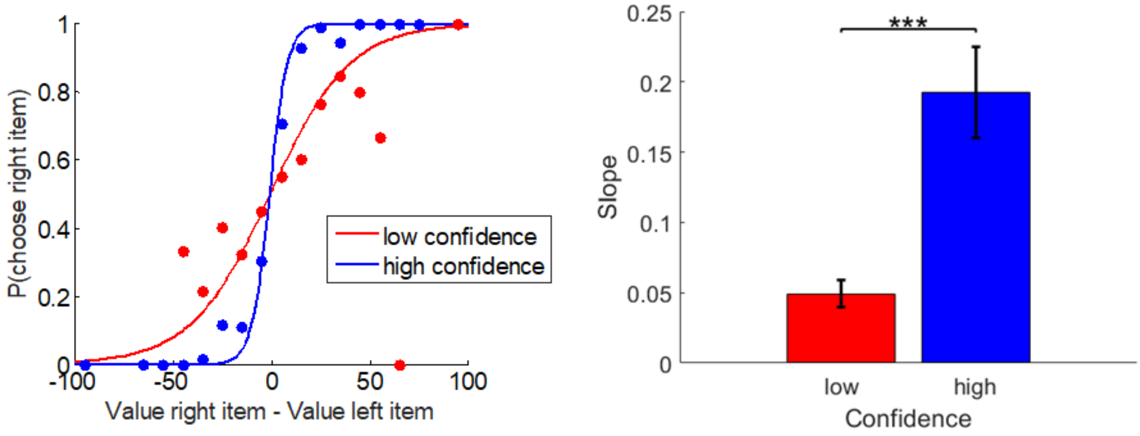


**Figure S1. Relationship between choices, pre-choice value ratings and choice confidence.** Left: the probability of choosing the item on the right (y-axis) is shown as a function of the pre-choice value difference (x-axis), for high- (blue) versus low- (red) confidence trials. The plain lines show the logistic prediction that would follow from group-averages of the corresponding slope estimates. Right: the corresponding logistic regression slope (y-axis) is shown for both high- (blue) and low- (red) confidence trials (group means +/- s.e.m.).

These results clearly replicate the findings of De Martino and colleagues (2013), which were interpreted with a race model variant of the accumulation-to-bound principle. We note, however, that this effect is also predicted by the MCD model. Here, variations in both (i) the prediction accuracy of choice from pre-choice value ratings, and (ii) choice confidence, are driven by variations in resource allocation. In brief, the expected magnitude of the perturbation of value representations increases with the amount of allocated resources. This eventually degrades the prediction accuracy of choice from pre-choice value ratings (which have been changed during the decision process). However, although more resources are allocated to the decision, this does not overcompensate for decision difficulty, and thus choice confidence decreases. Thus, low-confidence choices will be those choices that cannot be well predicted with pre-choice value ratings. We note that the anti-correlation between choice confidence and choice accuracy can be seen by comparing Figures 2 and 4 of the main text.

### 3. How do choice confidence, difference in pre-choice value ratings, and decision time relate to each other?

In the main text, we show that trial-by-trial variation in choice confidence is concurrently explained by both pre-choice value and value certainty ratings. Here, we reproduce previous findings relating choice confidence to both absolute value difference $\Delta VR^0$ and decision time (De Martino et al., 2013). First, we regressed, for each participant, decision time concurrently against both $|\Delta VR^0|$ and choice confidence. A random effect showed that both have a significant main effect on decision time ($\Delta VR^0$: mean GLM beta=-0.016, s.em.=0.003, p<0.001; choice confidence: mean GLM beta=-0.014, s.em.=0.002; p<0.001), without any two-way interaction (p=0.133). This analysis is summarized in Figure S2 below, together with the full three-way relationship between $|\Delta VR^0|$, confidence and decision time.

In brief, confidence increases with the absolute value difference and decreases with decision time. This effect is also predicted by the MCD model, for reasons identical to the explanation of the relationship between confidence and choice accuracy (see above). Recall that, overall, an increase in choice difficulty is expected to yield an increase in decision time and a decrease in choice confidence. This would produce the same data pattern as Figure S2, although the causal relationships implicit in this data representation is partially incongruent with the computational mechanisms underlying MCD.

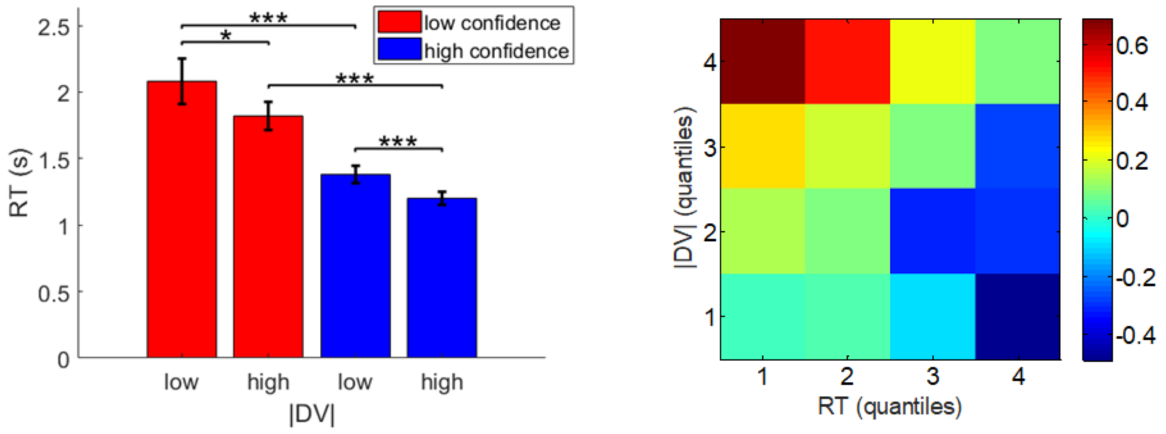


**Figure S2. Relationship between pre-choice value ratings, choice confidence, and decision time.** Left: decision time (y-axis) is plotted as a function of low- and high- $|\Delta VR^0|$ (x-axis) for both low- (red) and high- (blue) confidence trials. Error bars represent s.e.m. Right: A heatmap of mean z-scored confidence is shown as a function of both decision time (x-axis) and $|\Delta VR^0|$ (y-axis).

## 4. Analysis of changes of mind

In the main manuscript, we show that choice accuracy increases with pre-choice value difference $\Delta VR^0$ and pre-choice value certainty $VCR^0$. Recall that choice accuracy was defined in terms of the rate of choices that are congruent with preferences derived from pre-choice value ratings. Now, people make "inaccurate" choices either because they make mistakes or because they change their mind during the decision. In principle, we can discriminate between these two explanations because we can check whether "inaccurate" choices are congruent with post-choice value ratings (change of mind) or not (error). This is important, because accumulation-to-bound models do not allow for the possibility that value representations change during decisions. Hence all "inaccurate" choices would be deemed "errors", which are driven by stochastic noise in the evidence accumulation process. It turns out that most choices are "accurate" (mean choice accuracy =73.3%, s.e.m.=1%), and less than half of the "inaccurate" choices are classified as "errors" (mean error rate=12%, s.e.m.=0.01), which is significantly less than "mind changes" (mean rate difference=2%, s.e.m.=0.01, p=0.032). In addition, choice confidence and (post- versus pre-choice) value certainty gain were significantly higher for "changes of mind" than for "errors" (choice confidence: mean difference=13.7, s.e.m.=2.1, p<0.001; value certainty gain: mean difference=2.6, s.e.m.=1.4, p=0.035).

Thus, one may wonder what would be the impacts of both pre-choice value difference $\Delta VR^0$ and pre-choice value certainty $VCR^0$ on choice accuracy, if one were to remove "errors" from "inaccurate" choices. Figure S3 below shows both the predicted and measured three-way relationship between the probability of changing one's mind, $\Delta VR^0$ and $VCR^0$.

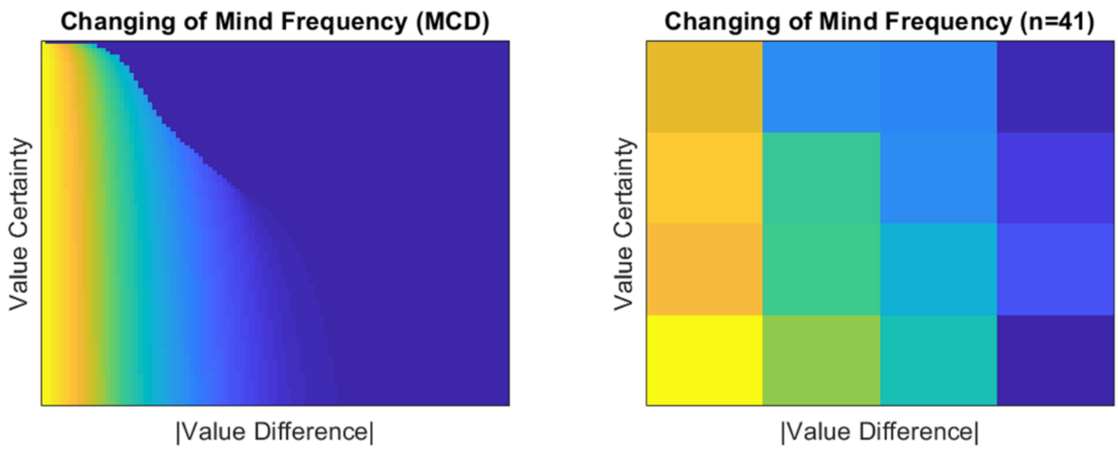


**Figure S3. Relationship between the probability of changing one's mind, value ratings, and certainty ratings.** Left: Prediction under the MCD model: a heatmap of the probability of changing one's mind is shown as a function of both $|\Delta VR^0|$ (x-axis) and $VCR^0$ (y-axis). Right: Empirical data: same format.

Recall that, under the MCD model, the probability of changing one's mind increases with the resource demand, which decreases when either $|\Delta VR^0|$ or $VCR^0$ increase. One can see that the data seem to conform to this prediction. To check this, we ran, for each participant, a multiple logistic regression of change of mind against $|\Delta VR^0|$ and

$VCR^0$. A random effect analysis shows that both have a significant and negative effect at the group level ($\Delta VR^0$: mean GLM beta=-0.16, s.e.m.=-0.02, $p<0.001$; $VCR^0$: mean GLM beta=-0.08, s.e.m.=0.02, $p<0.001$). These results are qualitatively similar to the analysis of choice accuracy (cf. Figure 2 in the main text).

## 5. Analysis of the subjective feeling of effort

In the main manuscript, we show that decision time decreases with pre-choice value difference $\Delta VR^0$ and pre-choice value certainty $VCR^0$. The focus on decision time was motivated by the fact that all models could make quantitative—and thus comparable—predictions. In brief, we found that the effect of $VCR^0$ on decision time was consistent with the MCD model, but not with accumulation-to-bound models. Now, under the MCD model, decision time is but a proxy for effort duration. Here, we ask whether the subjective feeling of effort *per se* follows the MCD model predictions. This is possible because we asked participants to rate how effortful each decision felt. Figure S4 below shows both the predicted and the measured three-way relationship between effort, $|\Delta VR^0|$ and $VCR^0$.

One can see that the reported subjective feeling of effort closely matches model predictions. One may ask whether people's effort reports may be trivial post-choice read-outs of decision time and/or choice confidence. This, however, is unlikely, given that people's subjective effort is reducible neither to decision time (mean correlation=0.39, s.e.m.=0.04), nor to choice confidence (mean correlation=-0.48, s.e.m.=, p=0.05).

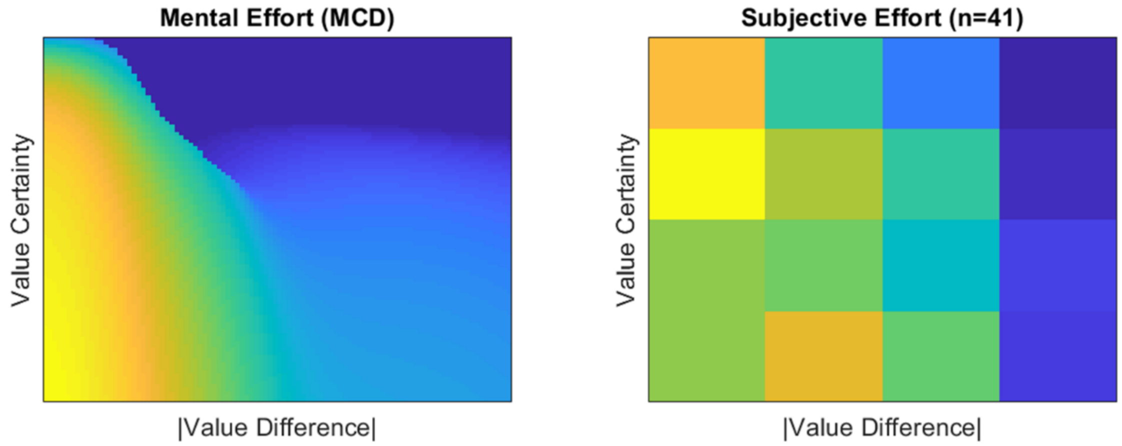


**Figure S4. Relationship between subjective effort, value ratings, and certainty ratings.** Left: Prediction under the MCD model: a heatmap of the MCD-optimal effort allocation is shown as a function of $|\Delta VR^0|$ (x-axis) and $VCR^0$ (y-axis). Note: this prediction is identical to Figure 3 in the main text (MCD model). Right: Empirical data: same format.

## 6. Do post-choice ratings better predict choice and choice confidence than pre-choice ratings?

The MCD model assumes that value representations are modified during the decision process, until the MCD-optimal amount of resources is met. This eventually triggers the decision, whose properties (i.e., which alternative option is eventually preferred, and with which confidence level) then reflects the modified value representations. If post-choice ratings are reports of modified value representations at the time when the choice is triggered, then choice and its associated confidence level should be better predicted with post-choice ratings than with pre-choice ratings. In what follows, we test this prediction.

In the first section of this Supplementary Material, we report the result of a logistic regression of choice against pre-choice value ratings (see also Figure S1). We performed the same regression analysis, but this time against post-choice value ratings. Figure S5 below shows the ensuing predictive power (here, in terms of balanced accuracy or BA) for both pre-choice and post-choice ratings. The main text also features the result of a multiple linear regression of choice confidence ratings onto $|\Delta VR^0|$ and $VCR^0$ (cf. Figure 4). Again, we performed the same regression, this time against post-choice ratings. Figure S5 below shows the ensuing predictive power (here, in terms of percentage of explained variance or $R^2$) for both pre-choice and post-choice ratings.

A simple random effect analysis shows that the predictive power of post-choice ratings is significantly higher than that of pre-choice ratings, both for choice (mean difference in BA=7%, s.e.m.=0.01, $p<0.001$) and choice confidence (mean difference in $R^2$=3%, s.e.m.=0.01, $p=0.004$).

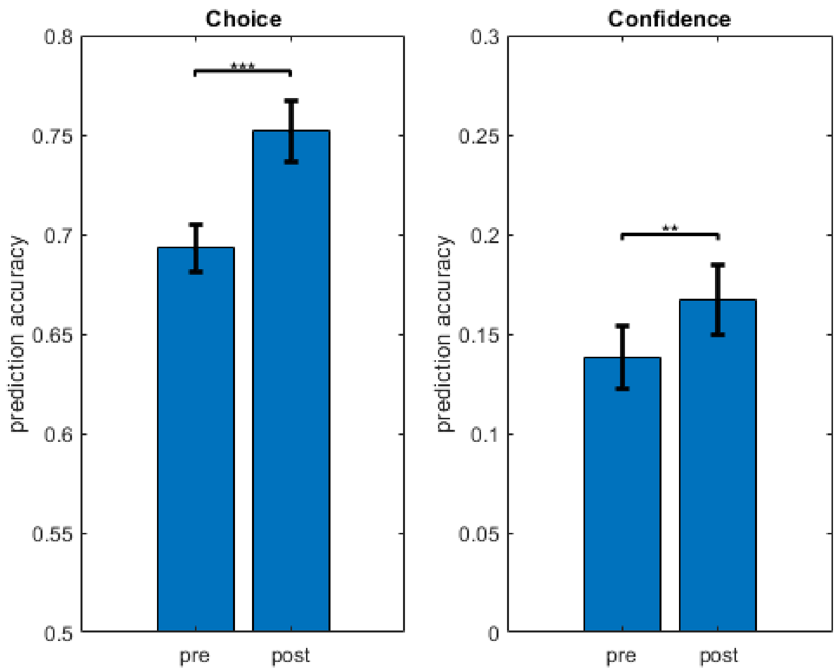


**Figure S5. Comparison of the predictive power of pre-choice versus post-choice ratings.** Left: Mean (+/- s.e.m.) BA of logistic regressions of choice against pre-choice (left) and post-choice (right) value ratings. Right: Mean (+/- s.e.m.) $R^2$ of multiple linear regressions of choice confidence against pre-choice (left) and post-choice (right) ratings.

## 7. Cross-condition analysis: decision importance and cost of decision time

As featured in the main manuscript, we intermixed "neutral" trials with two specific sets of trials, in which we either manipulated decision importance (cf. "consequential" decisions) or the cost of decision time (cf. "penalized" decisions). Figure S6 below shows the mean subjective effort ratings and decision times for "neutral", "consequential" and "penalized" decisions.

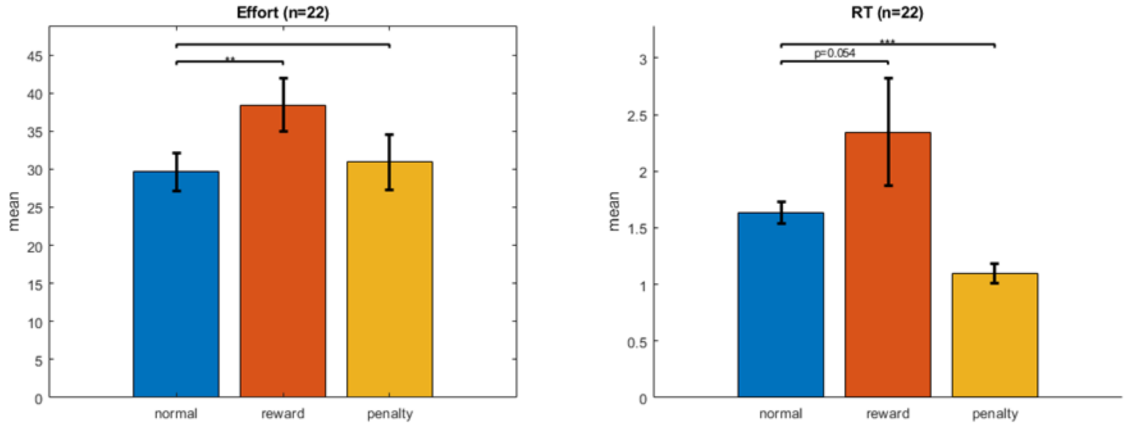


**Figure S5. Comparison of "neutral", "consequential", and "penalized" decisions.** Left: Mean (+/- s.e.m.) effort rating are shown for "neutral" (blue), "consequential" (red) and "penalized" (yellow) decisions. Right: Mean (+/- s.e.m.) decision time (same format).

Overall, the data partially follows the model predictions. In particular, subjective effort and decision time are both significantly higher for "consequential" than for "neutral" decisions (effort: mean difference=9.0, s.e.m.=2.2, $p<0.001$; DT: mean difference=0.56, s.e.m.=0.32, $p=0.043$). In addition, decision time is significantly lower for "penalized" than for "neutral" decision (mean DT difference=-0.46, s.e.m.=0.08, $p<0.001$). However, there is no noticeable difference between reported efforts in "neutral" and "penalized" decisions (mean effort difference=0.6, s.e.m.=2.1, $p=0.604$).

This comparison, however, may be confounded by between-condition differences in $\Delta VR^0$ or $VCR^0$. For each participant, we thus performed a multiple linear regression of effort and DT onto $|\Delta VR^0|$ and $VCR^0$, including all types of trials. Corrected effort and DT can now be compared, after having removed the effects of $|\Delta VR^0|$ and $VCR^0$. This is what Figure 6 of the main text shows. As one can see, the overall pattern is similar to Figure S5. As before, subjective effort and decision time are both significantly higher for "consequential" than for "neutral" decisions (effort: mean GLM beta difference=0.39, s.e.m.=0.12, p=0.001; DT: mean GLM beta difference=0.43, s.e.m.=0.19, p=0.017), and decision time is significantly lower for "penalized" than for "neutral" decisions (mean DT GLM beta difference=-0.51, s.e.m.=0.08, p<0.001). Finally, the difference between reported efforts in "neutral" and "penalized" decisions is now almost significant (mean effort GLM beta difference=-0.13, s.e.m.=0.12, p=0.147).

## 8. Analysis of eye-tracking data

We first checked whether pupil dilation positively correlates with participants' reports of subjective effort. We epoched the pupil size data into trial-by-trial time series, and temporally co-registered the epochs either at stimulus onset (starting 1.5 seconds before the stimulus onset and lasting 5 seconds) or at choice response (starting 3.5 seconds before the choice response and lasting 5 seconds). Data was baseline-corrected at stimulus onset. For each participant, we then regressed, at each time point during the decision, pupil size onto effort ratings (across trials). Time series of regression coefficients were then reported at the group level, and tested for statistical significance (correction for multiple comparison was performed using random field theory 1D-RFT). Figure S6 below summarizes this analysis, in terms of the baseline-corrected time series of regression coefficients.

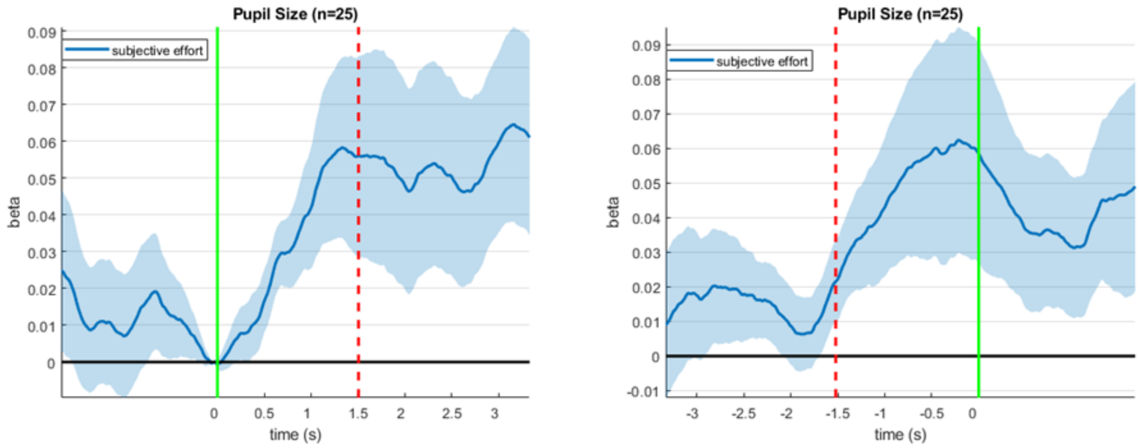


**Figure S6. Correlation between pupil size and reports of subjective effort during decision time.** Left: Mean (+/- s.e.m.) correlation between pupil size and subjective effort (y-axis) is plotted as a function of peristimulus time (x-axis). Here, epochs are co-registered w.r.t. stimulus onset (the green line indicates stimulus onset and the red dotted line indicates the average choice response). Right: Same, but for epochs co-registered w.r.t. choice response (the green line indicates choice response and the red dotted line indicates the average stimulus onset).

We found that the correlation between effort and pupil dilation was becoming significant from 500ms after stimulus onset onwards. Note that, using the same approach, we found a negative correlation between pupil dilation and pre-choice absolute value difference $|\Delta VR^0|$. However, this relationship disappeared when we entered both $|\Delta VR^0|$ and effort into the same regression model.

Our eye-tracking data also allowed us to ascertain which item was being gazed at for each point in peristimulus time (during decisions). Using the choice responses, we classified each time point as a gaze at the (to be) chosen item or at the (to be) rejected item. We then derived, for each decision, the ratio of time spent gazing at chosen/unchosen items versus the total duration of the decision (between stimulus onset and choice response). The difference between these two gaze ratios measures the overt attentional bias towards the chosen item. We refer to this as the gaze bias. Consistent with previous studies, we found that chosen items were gazed at more than rejected items (mean gaze bias=0.02, s.e.m.=0.01, p=0.067). However, we also found that this effect was in fact limited to low effort choices. Figure S7 below shows the gaze bias for low and high effort trials, based upon a median-split of subjective effort.

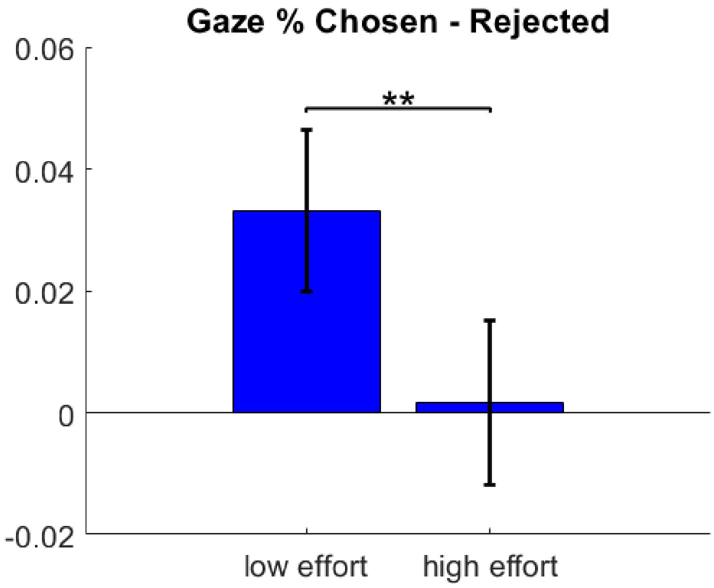


**Figure S7. Gaze bias for low and high effort trials.** Mean (+/- s.e.m.) gaze bias is plotted for both low (left) and high (right) effort trials.

We found that there was a significant gaze bias for low effort choices (mean gaze ratio difference=0.033, s.e.m.=0.013, p=0.009), but not for high effort choices (mean gaze ratio difference=0.002, s.e.m.=0.014, p=0.453). A potential trivial explanation for the fact that the gaze bias is large for low effort trials is that these are the trials where participants immediately recognize their favorite option, which attracts their attention. More interesting is the fact that the gaze bias is null for high effort trials. This may be taken as evidence for the fact that, on average, people allocate the same amount of (attentional) resources on both items. This is important, because we use this simplifying assumption in our MCD model derivations.

## 9. On the accuracy of the predicted confidence gain

The MCD model relies on the system's ability to anticipate the benefit of allocating resources to the decision process. Given the mathematical expression of choice confidence (cf. Equation 4 in the main text), this reduces to finding an analytical approximation to the following expression:

$$\bar{P} = E\left[ s\left( \lambda |x| \right) \right]$$

(S1)

where $x \rightarrow s(x) = 1/1+\mathrm{e}^{-x}$ is the sigmoid mapping, $\lambda$ is an arbitrary constant, and the expectation is taken under the Gaussian distribution of $x \sim N\left(\mu, \sigma^2\right)$, whose mean and variance are $\mu$ and $\sigma^2$, respectively.

Note that the absolute value mapping $x \rightarrow |x|$ follows a folded normal distribution, whose first two moments $E\left[|x|\right]$ and $V\left[|x|\right]$ have known expressions:

$$\begin{cases} E\left[|x|\right] = \sigma\sqrt{\dfrac{2}{\pi}}\exp\left(-\dfrac{|\mu|^2}{2\sigma^2}\right) + \mu\left(2\times s\left(\dfrac{\pi\mu}{\sigma\sqrt{3}}\right) - 1\right) \\ V\left[|x|\right] = \mu^2 + \sigma^2 - E\left[|x|\right]^2 \end{cases} \quad \text{(S2)}$$

where the first line relies on a moment-matching approximation to the cumulative normal distribution function (Daunizeau, 2017). This allows us to derive the following analytical approximation to Equation S1:

$$\bar{P} \approx s\left(\frac{E\left[|x|\right]}{\sqrt{\frac{1}{\lambda^2}+aV\left[|x|\right]}}\right) \quad \text{(S3)}$$

where setting $a \approx 3/\pi^2$ makes this approximation tight (Daunizeau, 2017).

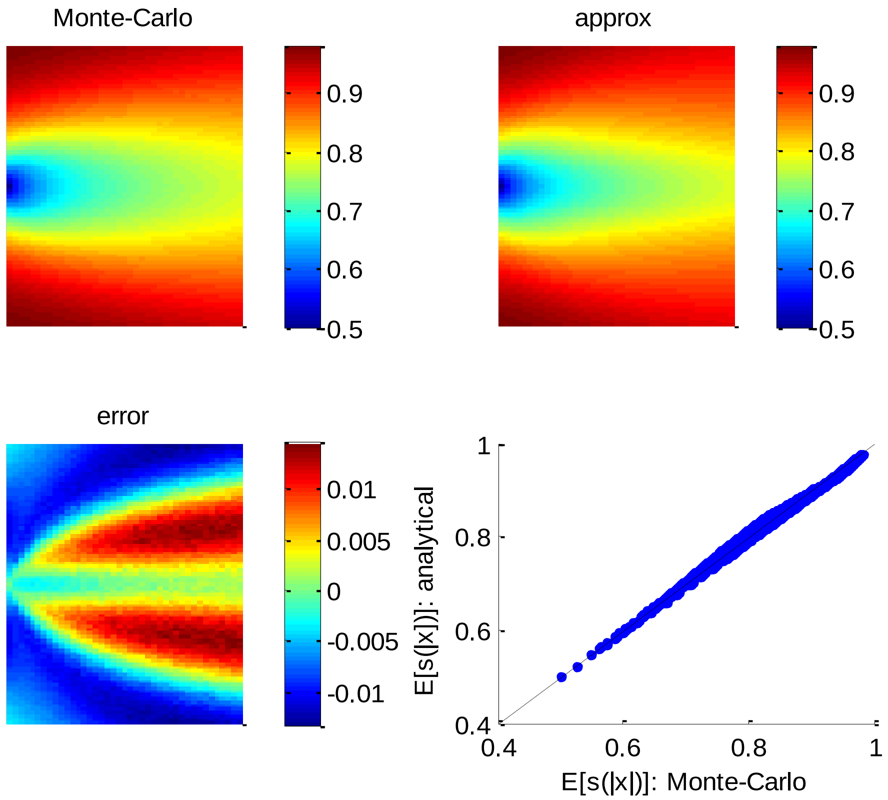


**Figure S8: Quality of the analytical approximation to $\bar{P}$.** Upper left panel: the Monte-Carlo estimate of $\bar{P}$ (colour-coded) is shown as a function of both the mean $\mu \in [-4,4]$ (y-axis) and the variance $\sigma^2 \in [0,4]$ (x-axis) of the parent process $x \sim N(\mu,\sigma^2)$. Upper right panel: analytic approximation to $\bar{P}$ as given by Equation S3 (same format). Lower left panel: the error, i.e. the difference between the Mon-Carlo and the analytic approximation (same format). Lower right panel: the analytic approximation (y-axis) is plotted as a function of the Monte-Carlo estimate (x-axis) for each pair of moments $\{\mu,\sigma^2\}$ of the parent distribution.

The quality of this approximation can be evaluated by drawing samples of $x \sim N\left(\mu,\sigma^2\right)$, and comparing the Monte-Carlo average of $s\left(\lambda|x|\right)$ with the expression given in Equation S3. This is summarized in Figure S8 above, where the range of variation for the moments of $x$ were set as follows: $\mu \in [-4,4]$ and $\sigma^2 \in [0,4]$.

One can see that the error rarely exceeds 5%, across the whole range of moments $\left\{\mu,\sigma^2\right\}$ of the parent distribution. This is how tight the semi-analytic approximation of the expected confidence gain (Equation 9 in the main text) is.

## 10. On the impact of model parameters for the MCD model

First, note that the properties of the metacognitive control of decisions (in terms of effort allocation and/or confidence) actually depends upon the demand for resources, which is itself determined by prior value representations. Now the way the MCD-optimal control responds to the resource demand (which is fully specified by the prior uncertainty $\sigma^0$ and the absolute means' difference $\left|\Delta\mu^0\right|$) is determined by effort efficacy and unitary cost parameters.

Let us first ask what would be the MCD-optimal effort $\hat{z}$ and confidence $\bar{P}_c\left(\hat{z}\right)$ when $\gamma = 0$, i.e. if the only effect of allocating resources is to increase the precision of value representations. We call this the "β-effect". It is depicted on Figure S9 below.

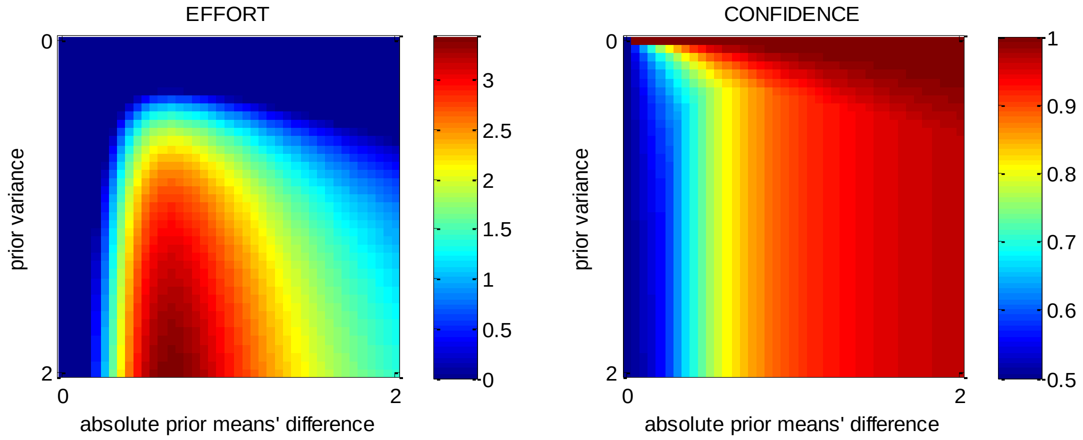


**Figure S9. The β-effect: MCD-optimal effort and confidence when effort has no impact on the value difference.** MCD-optimal effort (left) and confidence (right) are shown as a function of the absolute prior mean difference $|\Delta\mu^0|$ (x-axis) and prior variance $\sigma^0$ (y-axis).

One can see that, overall, increasing the prior variance $\sigma^0$ increases the resource demand, which eventually increases the MCD-optimal allocated effort $\hat{z}$. This, however, does not overcompensate for the loss of confidence incurred when increasing the prior variance. This is why the MCD-optimal confidence $\bar{P}_c(\hat{z})$ always decreases with the prior variance $\sigma^0$. Note that, for the same reason, the MCD-optimal confidence always increases with the absolute prior means' difference $|\Delta\mu^0|$. Now the impact of the absolute prior means' difference $|\Delta\mu^0|$ on $\hat{z}$ is less trivial. In brief, when $|\Delta\mu^0|$ is high, the MCD-optimal allocated effort $\hat{z}$ decreases with $|\Delta\mu^0|$. This is due to the fact that the resource demand decreases with $|\Delta\mu^0|$. However, if $|\Delta\mu^0|$ decreases even more, it eventually reaches a critical point, below which the MCD-optimal allocated effort $\hat{z}$ *increases* with $|\Delta\mu^0|$. This is because, although the resource demand still decreases with $|\Delta\mu^0|$, the cost of allocating resources overcompensates the gain in confidence. For such difficult decisions, the system does not follow the demand anymore, and progressively de-motivates the allocation of resources as $|\Delta\mu^0|$ continues to decrease. In brief, the amount $\hat{z}$ of allocated resources decreases away from a "sweet spot", which is the absolute prior means' difference that yields the maximal confidence gain per effort unit. Critically, the position of this sweet spot decreases with $\beta$ and increases with $\alpha$. This is because the confidence gain increases, by definition, with effort efficacy, whereas it becomes more costly when $\alpha$ increases.

Let us now ask what would be the MCD-optimal effort $\hat{z}$ and confidence $\bar{P}_c(\hat{z})$ when $\beta = 0$, i.e. if the only effect of allocating resources is to perturb the value difference. The ensuing "γ -effect" is depicted on Figure S10 below.

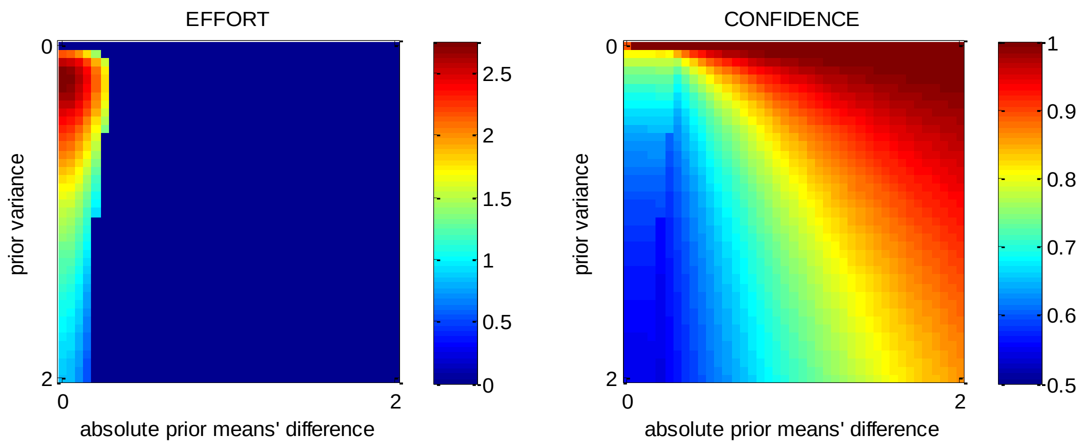


**Figure S10. The γ-effect: MCD-optimal effort and confidence when effort has no impact on value precision.** Same format as Fig S9.

In brief, the overall picture is reversed, with a few minor differences. One can see that increasing the absolute prior means' difference $|\Delta\mu^0|$ decreases the resource demand, which eventually decreases the MCD-optimal allocated effort $\hat{z}$. This does decrease confidence, because the γ-effect of allocated effort overcompensates the effect of variations in $|\Delta\mu^0|$. When no effort is allocated however, confidence is driven by $|\Delta\mu^0|$, i.e. it becomes an increasing function of $|\Delta\mu^0|$. In contrast, variations in the prior variance $\sigma^0$ always overcompensate the ensuing changes in effort, which is why confidence always decreases with $\sigma^0$. In addition, the amount $\hat{z}$ of allocated resources decreases away from a sweet prior variance spot, which is the prior variance $\sigma^0$ that yields the maximal confidence gain per effort unit. Critically, the position of this sweet spot increases with $\gamma$ and decreases with $\alpha$, for reasons similar to the β-effect.

Now one can ask what happens in the presence of both the β-effect and the γ-effect. If the effort unitary cost $\alpha$ is high enough, the MCD-optimal effort allocation is essentially the superposition of both effects. This means that there are two "sweet spots": one around some value of $|\Delta\mu^0|$ at high $\sigma^0$ (β-effect) and one around some value of $\sigma^0$ at high $|\Delta\mu^0|$ (γ-effect). If the effort unitary cost $\alpha$ decreases, then the position of the β-sweet spot increases and that of the β-sweet spot decreases, until they effectively merge together. This is exemplified on Figure S11 below.

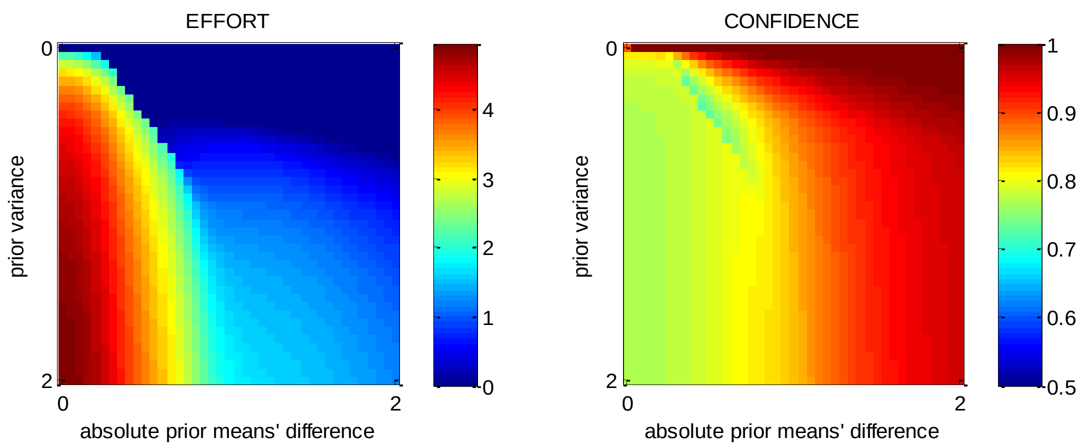


**Figure S11. MCD-optimal effort and confidence when both types of effort efficacy are operant.** Same format as Fig S9.

One can see that, somewhat paradoxically, the effort response is now much simpler. In brief, the MCD-optimal effort allocation $\hat{z}$ increases with the prior variance $\sigma^0$ and decreases with the absolute prior means' difference $|\Delta\mu^0|$. The landscape of the ensuing MCD-optimal confidence level $\bar{P}_c(\hat{z})$ is slightly less trivial, but globally, it can be thought of as increasing with $|\Delta\mu^0|$ and decreasing with $\sigma^0$. Here again, this is because variations in $|\Delta\mu^0|$ and/or $\sigma^0$ almost always overcompensate the ensuing changes in allocated effort.

## 11. Accumulation-to-bound process models

In the main text, we compare the MCD model to two variants of accumulation-to-bound process models, namely: an optimal DDM with collapsing bounds (Tajima et al., 2016) and a modified race model (De Martino et al., 2013). We focus on these two models because both can make quantitative predictions regarding choice, value, decision time, and choice confidence.

Recall that DDMs essentially solve the problem of comparing uncertain values to make accurate choices as quickly as possible. Tajima and colleagues assume that, at the beginning of each trial, the two options have true but unknown values $V_1$ and $V_2$, which the decision maker only indirectly accesses through some instantaneous noisy evidence $x_j^t = V_j + \varepsilon_j^t$, where $t$ indexes time. Here, $\varepsilon_j^t \sim N(0,\sigma)$ is some Gaussian random noise with variance $\sigma$, which partially masks the true values. The noise variance $\sigma$ thus effectively controls the (un)reliability of the evidence signal $x_j^t$. The decision maker then progressively updates his/her posterior estimate $\hat{V}_j^t$ by accumulating past evidence signals. When neglecting, for the sake of simplicity, the decision maker's prior belief about $V_j$ , the decision maker's value estimate is given by: $\hat{V}_j^t \approx \frac{1}{t}\sum_{t'=1}^{t} x_j^{t'} = V_j + \frac{1}{t}\sum_{t'=1}^{t}\varepsilon_j^{t'}$ . As one can see, $\hat{V}_j^t$ fluctuates along with the noise, and behaves as a random walk that eventually converges towards $V_j$ . Now the model also assumes that the decision maker pays a cost $c$ per second of evidence accumulation. Optimal decision making then amounts to finding a policy that maximizes the expected discounted value $\langle V_j \left| x_j^1, x_j^2, ..., x_j^T \right\rangle - c\langle T \rangle$ . It turns out that the optimal policy is to wait until the estimated value difference $\Delta\hat{V}^t = \hat{V}_1^t - \hat{V}_2^t$ eventually hits any of two (upper or lower) collapsing bounds, at which point the decision maker commits to the corresponding choice. Setting these optimal collapsing bounds is done by numerically solving the so-called Bellman equation. In this work, we simply use the code written by Tajima and colleagues to perform their simulations. Let us now make two remarks on this model. First, the height of the bound at the time when it is hit measures choice confidence. This is essentially because choice confidence increases with $\Delta\hat{V}$ .

The race model of De Martino and colleagues was specifically proposed to predict choice confidence in the context of value-based decisions. Here, separate decision variables $\hat{V}_1^t$ and $\hat{V}_2^t$ accumulate evidence for each option, with the decision being determined by which accumulator reaches the threshold first. At each time step, a new evidence sample $x^t \sim N(\Delta V, \sigma)$ is drawn from a Gaussian distribution, which is a noisy measure of the value difference between the two options. The model then assumes that only the decision variable that benefits from the evidence sample is updated. For example, if $x^t > 0$ , then $\hat{V}_1^{t+1} = \hat{V}_1^t + x^t$ and $\hat{V}_2^{t+1} = \hat{V}_2^t$ . This ensures that decision variables can only increase with accumulation time. Here, confidence is defined as the gap $\left|\Delta\hat{V}^t\right|$ between the two variables when the bound is hit.

Critically, both models can make predictions about choice accuracy, decision time, and choice confidence from value ratings... and value certainty ratings (although this was never exploited before). This is because, under both models, the more uncertain people are about option values, the less reliable evidence signals will be, i.e. the higher the noise variance $\sigma$ should be. As we see when simulating the model, increasing $\sigma$ increases the probability of an early bound hit. For the DDM model, this implies that increasing $\sigma$ increases choice confidence (because optimal bounds are collapsing over decision time). In addition, under the race model, increasing $\sigma$ increases the average gap between the two accumulators, eventually yielding the same prediction.

Lastly, let us highlight that one of the core assumption of both these models is that option values do not change during the decision. This is because the models focus on comparing option values, not on constructing them (option values are considered as inputs to the value comparison system). This effectively prevents them from being able to explain choice-induced preference change.