The metacognitive control of decisions predicts whether and how mice override their default policy

C. Schreiweis[1], M. Euvrard[1], E. Burguière[1,], J. Daunizeau[1,2,*]*

[1] Paris Brain Institute - ICM, Hôpital de la Pitié Salpêtrière, Paris, France

[2] ETH, Zurich, Switzerland

***** These authors contributed equallly to the work

Address for correspondence:

Jean Daunizeau

Motivation, Brain and Behavior Group

Paris Brain Institute (ICM)

47, bd de l'Hôpital, 75013, Paris, France.

Tel: +33 1 57 27 43 26

Mail: jean.daunizeau@icm-institute.org

Web: http://sites.google.com/site/jeandaunizeauswebsite

Abstract

Decisions permeate every aspect of our lives - what to eat, where to live, 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. Yet, we don't always follow the fast and negligent lead of habits or intuitions. So what determines how much time and effort we invest when making decisions? A possibility is that the amount of cognitive resources the brain allocates to a given task (in particular, a decision task) optimizes a cost-benefit tradeoff. Our working hypothesis is that decision confidence serves as the benefit term of this resource allocation problem, hence the “metacognitive” nature of decision control. Here, we provide behavioural evidence that mice regulate their allocation of cognitive resources according to a confidence/effort a tradeoff. The importance of this work is twofold. On the computational side of things, we show whether and how this eventually enables them to override their default/automatic policy. On the behavioural side of things, we validate all the ensuing model predictions, in the context of an experimental setup with high ecological validity.

Introduction

What makes you ponder some decisions, but not others? Decisions permeate every aspect of our lives - what to eat, where to live, 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, Slovic, and Tversky 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 and negligent lead of habits or intuitions. So what determines how much time and effort we invest when making decisions?

Heedless decisions can be triggered by psychobiological determinants such as stress *(Porcelli and Delgado 2009; Porcelli, Lewis, and Delgado 2012)*, emotions *(Harlé and Sanfey 2007; Martino et al. 2006; Sokol-Hessner, Camerer, and Phelps 2013)* or fatigue *(Blain, Hollard, and Pessiglione 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, O'Doherty, and Rangel 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, Huk, and Shadlen 2005)*.

Over the past decade, computational accounts of decision making have formalized a simple and elegant mechanistic idea that relates decision time and accuracy. They assert that choices are only triggered once the accumulation of "evidence" in favour of alternative choice options crosses a critical decision threshold *(Ratcliff, Van Zandt, and McKoon 1999; Usher and Mcclelland 2001)*, which is modulated by external incentives *(Frank 2006; Simen, Cohen, and Holmes 2006)*. In turn, response times increase when the reward attached to correct decisions increases. Further computational work demonstrated that accumulation-to-bound policies, whose threshold is optimally set in the aim of maximizing decision

efficiency (i.e. expected reward rate), behave similarly *(Ditterich 2006; Drugowitsch et al. 2012)*. The idea is that spending time waiting for the evidence to identify what to do to reach a small reward eventually induces an *opportunity cost*: namely, that of foregoing higher rewards (had the invested decision time been spent on trying to obtain those). In turn, the system should spend most of its available time deliberating on decisions that yield high rewards. From a theoretical perspective, this implies that heedless decisions can be viewed as an evolutionary adaptation, in response to selective pressure that favors efficiency *(Pirrone, Stafford, and Marshall 2014)*.

But time may not be the only cost to pay when pondering decisions. Recall that most decision contexts trigger (quick and effortless) automatic behavioural tendencies, which originate from learned habits or intuitions *(Schneider and Shiffrin 1977)*. Typically, overriding such automatic behavioural tendencies involve some non-negligible amount of mental effort *(Massar, Csathó, and Van der Linden 2018; Sullivan-Toole et al. 2019)*. The latter can be thought of as the subjective cost of engaging "executive" resources (e.g., attention and working memory) in order to process decision-relevant information *(Birnberg and Shields 1984; Bordalo, Gennaioli, and Shleifer 2017)*. Importantly, it was recently suggested that executive resources may be strategically deployed, in the aim of maximizing the so-called *expected value of control* or EVC *(Musslick et al. 2015; Ozcimder et al. 2017; Shenhav, Botvinick, and Cohen 2013)*. Here, the EVC measures how worth the investment of executive resources is expected to be. More precisely, the EVC discounts the anticipated benefits of allocating resources with its ensuing anticipated effort cost. This implies that the amount of executive resources the brain allocates to a given task (in particular, a decision task) optimizes a reward-effort tradeoff. In conjunction with simple optimality principles, this idea has proven fruitful for understanding the relationship between mental effort and peoples' performance in various cognitive tasks, in particular those that involve executive resources *(Griffiths, Lieder, and Goodman 2015; Lieder et al. 2018)*. Recently, it was adapted to the specific case of value-based decision making, and framed as a self-contained computational model: namely, the Metacognitive Control of Decisions or MCD *(Lee and Daunizeau 2021)*. Importantly, online implementations of MCD *(Lee et al. 2023)* generalize accumulation-to-bound models, including those that optimize decision efficiency *(Tajima et al. 2019; Tajima, Drugowitsch, and Pouget 2016)*. The reason is twofold. First, effort can unfold along two

dimensions: intensity and duration. In particular, information processing speed may be accelerated by allocating more executive resources to the task. Under this perspective, both speed and accuracy may increase at the cost of intensifying effort, which is neglected in current accumulation-to-bound models of decision making. Second, the same decision control architecture can provide a quasi-optimal solution to the resource allocation problem, irrespective of the inner workings of the underlying decision system.

Central to this work is the notion that decision confidence serves as the benefit term of the resource allocation problem, hence the "metacognitive" nature of decision control. Formally, confidence derives from the uncertainty of subjective value representations, which evolve over decision time as the brain processes more value-relevant information. In brief, low confidence induces a latent demand for mental effort: the brain refines uncertain value representations by deploying cognitive resources, until they reach the optimal confidence/effort trade-off. Interestingly, this mechanism was shown to explain the (otherwise surprising) phenomenon of choice-induced preference change in humans (Lee and Daunizeau, 2020). More importantly, the MCD model makes quantitative out-of-sample predictions about many features of value-based decisions, including decision time, subjective feeling of effort, choice confidence and changes of mind. These predictions have already been tested in humans, using a dedicated behavioral paradigm (Lee and Daunizeau, 2021). In this work, we test whether the MCD framework can explain the properties of decision control in mice. This is important, because (transgenic lines of) mice are the most frequent rodent model for translational neuropsychiatric approaches. This paves the way for chemo- and opto-genetic investigations of the causal role of prefrontal subsystems in mental effort regulation.

We thus developed a self-administered, automated, operant conditioning paradigm. During several weeks, mice stay in a customized conditioning box, in which they obtain food only when they successfully engage with a deterministic, binary, visual decision task. Critically, we systematically varied decision difficulty (i.e. visual cue discriminability) and reward (i.e. amount of food), and measure the ensuing performance and response time. The latter serves as an empirical proxy for the effort invested in each decision. This enabled us to track, over multiple days and in a pseudo-ecological environment, how mice regulate their effort, above and beyond variations in endogenous psycho-biological states (such as circadian cycle

phase, hunger state, stress, etc) that may otherwise confound effort regulation mechanisms. Our working hypothesis is twofold. First, it is only worth allocating executive resources to process the visual cue when default/automatic policies provide insufficient confidence. Here, automatic policies derive from implicit learning of action-reward contingencies, which are partially reinforced in the task. Second, executive resources are deployed in a progressive and flexible manner, until they reach a threshold that maximizes the EVC. Under the MCD framework, this threshold optimally balances decision difficulty (which induces a demand for executive resources), decision importance (which scales with the reward attached to accurate decisions) and effort costs (which increase with the amount of allocated resources). This provides quantitative predictions regarding the relationship between decision difficulty and reward, on the one hand, and performance and effort, on the other hand.

Materials and Methods

Animals

All experimental procedures followed national and European guidelines and were approved by the institutional review boards (French Ministry of Higher Education, Research and Innovation; APAFiS Protocols no. 1418-2015120217347265 and 2021042017105235). Animals were bred and group-housed in the animal facilities of the Paris Brain Institute in Tecniplast ventilated polycarbonate cages under positive pressure with hard-wood bedding in groups of up to six animals per cage with *ab libitum* food and water access. The temperature was maintained at 21–23 °C and the relative humidity at 55 ± 10% with a 12-h light/dark cycle (lights on/off at 8am and 8pm, respectively). The animals were maintained in a 12-hour light/dark cycle (lights on/off at 8:00am/8:00 pm, respectively), and had *ab libitum* food and water access. In this study, n = 23 adult male mice were tested and no data excluded.

Experimental setup

Operant conditioning was conducted in custom-modified experimental chambers (ENV-007CTX, Med Associates, Vermont, USA) as previously described (Benzina et al. 2021), in which each of the tested animals lives and performs the conditioning task 24/7 in an automated, self-initiated, self-paced manner. Seven such experimental operant chambers were operated in parallel in the same experimental room. Briefly, the rear wall of each chamber housed the feeder compartment (equipped with an infrared light beam to detect feeder entry crossing) and drinking water bottle holder, the front wall held two tactile screens, and a custom-adapted, homemade pair of black Plexiglas gates equipped with a pair of infrared (IR) beams. These beams allowed for tracking the locomotion of the mouse into or out of the area with the tactile screens. Precision reward pellets (20 mg LabTabTM AIN-76A rodent precision pellets, TestDiet, Richmond, USA) served as the sole nutrition during the entire experiment and all rewards had to be earned through successful performance of the task.

Experimental task

Once placed in the operant conditioning chambers, the mice start with a pre-training phase of five different stages. In a first stage, a total of five precision pellets drop automatically once every minute if the previous pellet had successfully been recovered. In a second stage, the mouse is required to nose poke into the feeder compartment to trigger individual pellet delivery once per minute until a total of ten pellets has been received. In a third stage, crossing of the IR photobeams prompts the illumination of both tactile screens; the mouse is required to touch one of them to trigger pellet delivery in the feeder. There was no time limit for recovering the pellets at this stage. However, the condition for the mice to pass into the following pre-training stage was the recovery of the pellets within less than 15 seconds

for at least 66% of the times. Then, in the consecutive pre-training stage, as before, both screens were illuminated with white light upon gate crossing. However, mice now had to nose poke into the feeder in less than 15 seconds after screen touch in order to trigger pellet delivery. Having successfully recovered at least 66% of the pellets in this manner, mice passed into the actual task. The last (fifth) pre-training stage consisted in a deterministic, binary visual discrimination task with two stimuli pairs. The presented stimulus pair consisted in ten, either vertical or horizontal, white bars on a black screen, each of which was presented in a random manner (with an overall 50% frequency). To earn a precision pellet, mice need to associate each of the two visual cues with one screen touch, e.g. stimulus of ten horizontal/ vertical bars associated with the action of left/right screen touch, accordingly. The cue was shown simultaneously on both touchscreens for two reasons: first, even if mice would only partially sample from one screen, they would have all information at their disposal to successfully solve the task; and, second, the mouse signals her decision (which screen she touches) independently from the way we present the visual cue. The cue-outcome contingencies were randomized across individuals and operant chambers. All trials were self-initiated and onsets of the visual cue pair on the tactile screens were triggered by the mouse crossing custom-designed, homemade photobeam-equipped gates. Mice were allowed a maximum of 60 seconds after stimulus onset to respond with a screen touch. Upon correct response, the screens started blinking to give a positive feedback for the mouse, which had 15 seconds to nose poke in the food dispenser in the rear wall for obtaining a food reward. Past 15 seconds after correct response, the trial became unrewarded. In case of incorrect screen touch, mice received an aversive white light for five seconds. Additionally, in order to decrease the tendency of mice to lateralize, each trial following an incorrect response required the mouse to switch to the other screen (so-called

“corrective trials”). Past 60 seconds in the chamber compartment with the screens, the screens would switch off again; in these “time out” trials, mice then had to cross out and back into the feeder compartment to trigger the next trial. Trials where mice crossed in and out of the screen compartment within 60 seconds and without screen touch, were labelled as “aborted trials”.

Having reached 80% correct responses over the past 40 trials, the mice transitioned into the main (decision making) task, where we systematically modulated reward amounts and difficulty levels. Upon a correct response, the mouse earned either one or three reward pellets, under low or high reward condition, respectively. Mice subsequently performed low and high reward trial blocks of each 60 trials, consisting of 10 trials per difficulty level. Difficulty levels consisted of variations of the visual cue, whose association with the outcome was previously learned during pre-training stage five. More precisely, these variations consisted of different ratios of intersecting vertical and horizontal bars: the easiest difficulty level consisted of the same 10 vertical or 10 horizontal bars as in pre-training stage five; the next difficulty level consisted of 9:1 intersecting vertical and horizontal bars, etc, until five and five intersecting horizontal and vertical bars, representing the highest and unsolvable difficulty level, which was rewarded randomly 50% on each side. Except for the control 5:5 difficulty level, the cue-outcome contingency was identical to that of pre-training stage five, having identified the side with the dominant orientation of bars.

Computational model

Our computational model relies upon three main working assumptions: (i) prior to processing the visual cue, mice possess a default/automatic action policy that derives from trial-by-trial reinforcement learning of action-outcome contingencies, (ii) mice may override

the default/automatic policy by engaging in a controlled/effortful processing of the visual cue, and (iii) mice regulate their effort investment on a decision-by-decision basis, by optimizing the balance between decision confidence and effort cost, hence the "metacognitive" nature of decision control (Lee and Daunizeau 2021).

Assumption (i) states that mice learn the value of candidate actions by updating a running average of the amount of reward that followed each candidate action. This yields a default/automatic preference, which is but the difference in action values. When the learning rate is high, this reduces to a very simple "win-stay/lose-switch" default policy, whereby the mouse tends to repeat (resp., switch from) a successful (resp. unsuccessful) action.

Although her default policy will yield some relative success (cf. corrective trials in the pre-training phase), the mouse also has learned the value of paying attention to the visual cue. Assumption (ii) states that, as the mouse spends more time sampling the visual cue, she progressively recognizes the dominant orientation of bars, hence potentially changing her mind about which action will be successful. Let $\mu(z)$ and z be the (signed) strength of evidence regarding which side is most likely to yield a reward, and the amount of attentional resources that the mouse invest for processing the visual cue, respectively. For example, the evidence signal $\mu(z)$ can be thought of as an uncertain estimate of the difference between the numbers of vertical and horizontal bars. After having invested an amount z of resources, the mouse's subjective confidence can be described as the probability of having identified the correct answer side $P_c(z)$ (Lee and Daunizeau 2021):

$$P_c\left(z\right) \approx s\left(\frac{\pi\left|\mu\left(z\right)\right|}{\sqrt{3\sigma\left(z\right)}}\right) \tag{1}$$

where $s(\square)$ is the standard sigmoid mapping. Here, $1/\sigma(z) = 1/\sigma^0 + \beta \times z$ measures the evidence reliability, which starts at σ^0 and increases linearly with the amount of invested attentional resources at a rate β. Similarly, $\mu(z) = \mu^0 + \delta(z)$ starts at some (default) estimate μ^0, which is progressively by a (random) perturbation term $\delta(z) \square N(0, \gamma \times z)$ that results from investing an amount z of attentional resources to the visual cue. Critically however, the direction of the perturbation is a priori unknown, i.e. the information that can be obtained from attending to the visual cue cannot be anticipated prior to processing the cue. A tenable assumption however, is that the expected magnitude of the perturbation increases linearly with the amount of invested attentional resources (at a rate γ).

Equation (1) effectively states how investing an amount of resources eventually modifies decision confidence, under relatively agnostic assumptions regarding the way the visual cue is processed. This confidence level can be greater or smaller than the initial (default) confidence level, because new information has been assimilated since then. Based upon Equation (1), a metacognitive system will anticipate that investing additional resources will increase its confidence on average (Lee and Daunizeau 2021). But this may not always overcompensate the cost of spending more resources on the decision. Thus, what should the decision to stop or continue look like, in order to maximize the expected cost-benefit tradeoff? It turns out that this problem is one of *optimal stopping*, which is a special case of Markov Decision Processes (Feinberg and Shwartz 2012; Papadimitriou and Tsitsiklis 1987). In fact, it can be shown that the optimal stopping policy derives from a comparison between the current confidence $P_c\left(z\right)$ and some threshold value $\omega(z)$, which depends upon the

amount of resources invested so far. Let $\pi_\omega(z)$ be the stopping policy that is induced by the threshold $\omega(z)$:

$$\pi_\omega(z) = \begin{cases} STOP & \text{if } P_c(z) \geq \omega(z) \\ CONTINUE & \text{otherwise} \end{cases} \tag{2}$$

Finding the optimal stopping policy π_ω^* thus reduces to finding the optimal threshold $\omega^*(z)$. Assumption (iii) states that the mouse stops allocating attention to the cue when her confidence level reaches a threshold $\omega^*(z)$ that maximizes the following cost-benefit tradeoff:

$$\omega^*(z) = \arg\max_{\omega(z)} E\left[\underbrace{R \times P_c(z)}_{\text{benefit}} - \underbrace{C(z)}_{\text{cost}} \middle| \pi_\omega(z)\right] \tag{3}$$

where the expectation is taken under the stopping policy in Equation 2, R is the reward at stake, and $C(z) = \alpha \times z^\nu$ is the subjective cost of resource investment (where α is the unitary cost of effort and ν is the cost power). Practically speaking, Equation (3) can be solved by trial and error or by backward induction (Lee et al. 2023).

It follows that average decision time and accuracy depend on default preferences and reward at stake, as well as on effort efficacy (β and γ) and cost (α and ν) parameters. Here, we assume that mice have learned the effort efficacy and cost parameters during the pre-training phase of the experiment. In addition, mice can't anticipate the difficulty of a given trial prior to sampling the cue. Therefore, their stopping policy can only adjust to the reward at stake (which is blocked). It turns out that the net effect of increasing R is to heighten the confidence threshold, therefore increasing both response time and decision accuracy. Note that, under the MCD model, the predicted impact of default preferences on response time and decision accuracy recapitulates previous empirical work (Lopez-Persem, Domenech, and

Pessiglione 2016). In brief, if the default preference aligns with the cue, then decision accuracy is higher, and response time is shorter (because confidence reaches the threshold sooner). However, this effect is most noticeable when the cue is poorly informative (high difficulty). The impact of decision difficulty however, is much less trivial. In our context, difficult decisions induce some sort of visual clutter that renders the evidence signal very noisy and unreliable. As a result, the first visual saccades of the mouse may induce strong random oscillations in the evidence signal, which will eventually dampen, should the mouse invest more attentional resources. Quantifying the predicted impact of decision difficulty thus requires numerical simulations of our empirical setup, including two reward levels (R=1 or 3) and five difficulty levels (analogous to the 0:10, 1:9, 2:8, 3:7, 4:6 and 5:5 stimuli ratios). We have simulated 1000 trials per combination of reward and difficulty level, under the above MCD model. Figure 1 below summarizes the properties of MCD's optimal stopping policy.

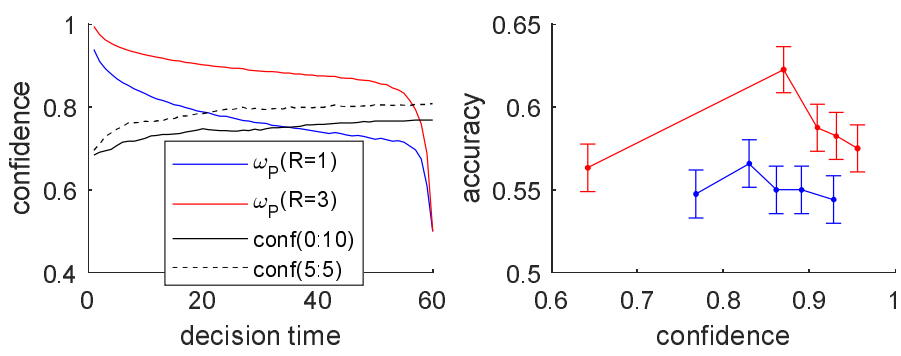

Figure 1: Basic properties of the metacognitive control of decisions. Left panel: The blue and red lines indicate the optimal confidence threshold (y-axis) as a function of decision time (x-axis), for low and high reward conditions, respectively. The black plain and dotted lines indicate the average confidence dynamics for the easiest (0:10) and hardest (5:5) condition, respectively. **Right panel**: Average decision accuracy (y-axis) is plotted against decision confidence (x-axis) for both low (blue) and high (red) reward conditions.

One can see that the optimal stopping policy adjusts to the reward at stake by effectively demanding a higher confidence to stop the deliberation. Also, increasing decision difficulty typically slows down the confidence gain that can be expected from investing more attentional resources. Taken together, these two mechanisms conspire to make subjective confidence a relatively unreliable proxy for decision accuracy (cf. upper-right panel in Fig. 1).

Now Figure 2 below summarizes the net effect of reward and difficulty onto decision accuracy and response time. To capture the impact of default preference, we also have regressed simulated decision accuracy and response time concurrently against reward, difficulty and a binary indicator of whether the default policy aligns with the cue (cf. lower panels of Fig. 2).

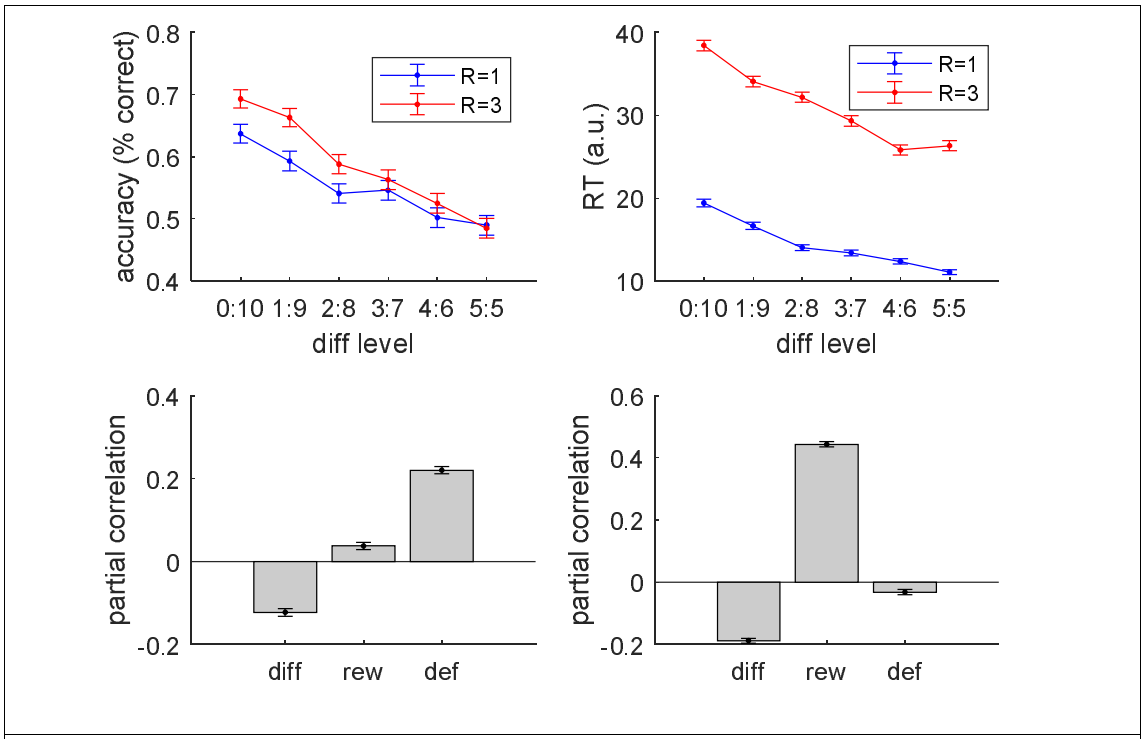

Figure 2: Impact of task conditions on decision accuracy and response time. Upper-left panel: Average decision accuracy (y-axis) is plotted against decision difficulty (x-axis) for both reward conditions. **Upper-right panel**: Average response time, same format. **Lower-left panel:** Multiple regression of decision accuracy: partial correlation coefficients (y-axis) are plotted for difficulty, reward and whether the default policy aligns with the cue. **Lower-right panel:** Multiple regression of response time, same format. In all panels,

errorbars show the standard error around the mean (s.e.m.).

One can see that the net effect of increasing decision difficulty is to decrease response time and (thus) decrease decision accuracy. However, increasing reward systematically increases decision time and (thus) decision accuracy. Note that the latter effect is relatively weak. This is because high resource investments effectively yield small performance gains.

Finally Figure 3 below summarizes the relationship between response time, on one hand, and decision accuracy or confidence, on the other hand (across difficulty levels).

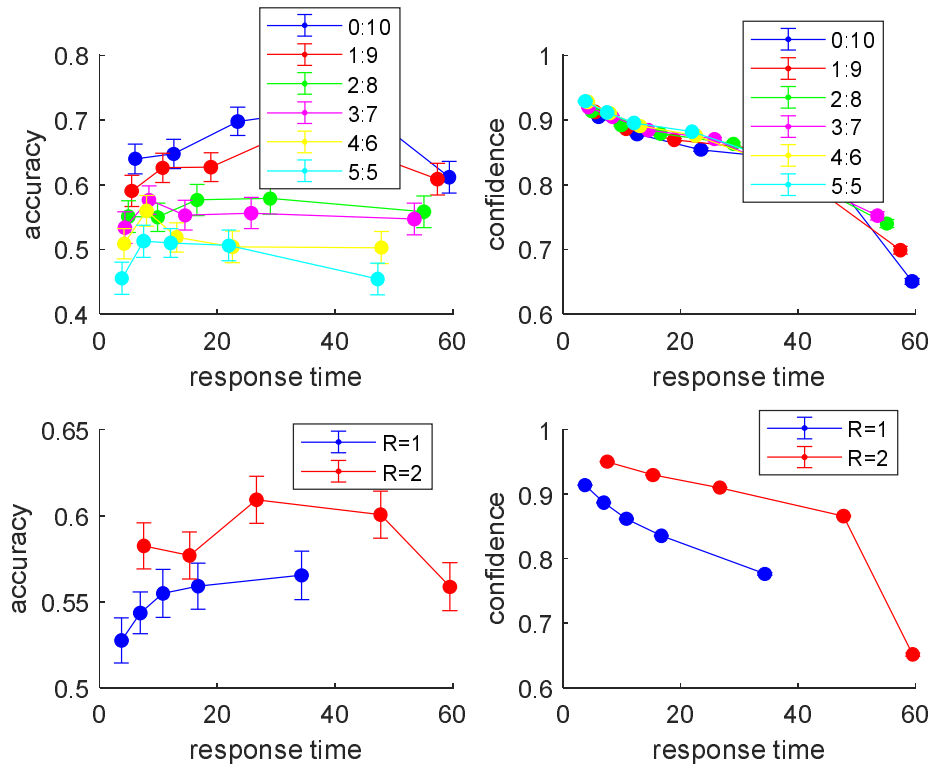

Figure 3: Relationship between response time and decision accuracy. Upper-left panel: Average decision accuracy (y-axis) is plotted against response time (x-axis) for all difficulty levels (color code). **Upper-right panel**: Subjective confidence, same format. **Lower-left panel**: Average decision accuracy (y-axis) is plotted against response time (x-axis) for both reward conditions (color code). **Lower-right panel**: Subjective confidence, same format. In all panels, errorbars show the standard error around the mean (s.e.m.).

One can see that the shape of the relationship between response time and decision accuracy slightly changes as difficulty increases. Although the overall relationship is slightly concave, the critical response time beyond which increasing response time *decreases* decision accuracy is smaller and smaller, as decision difficulty increases. In turn, the marginal utility of resource investments decreases with decision difficulty. This is not the case for subjective confidence, which tends to preserve its relationship with response time, across all difficulty levels. In fact, the difficulty-dependent relationships between confidence and response time are practically indistinguishable from each other. This derives from the fact that subjective confidence levels are fully determined by the thresholds dynamics (which are invariant across difficulty levels).

When splitting trials according to reward (lower panels of Fig. 3), the emerging pattern is slightly different. First, although the relationship between response times and decision accuracy is concave for the high reward condition, it is actually monotonically increasing for the low reward condition. This is simply because, under the low reward condition, response times only span a smaller range of variation, which lies before the average critical response time. Second, confidence monotonically decreases when response time increases (across its whole variation range). Third, increasing reward eventually increases both decision accuracy and confidence (this could already be seen in Fig. 1 and 2).

Behavioral results

Prior to conducting statistical analyses of choice and response time data during the main phase of the experiment, we performed a series of quality checks.

First, we checked the time it took for pre-training mice in stage five (until they reached the performance criterion) and the duration of the main experimental phase (after stage five). On average, mice took 1327 trials to complete the pre-training phase (standard deviation: 470), and they performed 3290 trials of the main task (standard deviation: 1340).

Second, we checked whether the circadian cycle influenced their behaviour. For each mouse, we binned the trials according to the hour of the day (from 12am to 12pm). We then estimated the percentage of trial occurrences, the rate of each trial type, as well as the mean response time and decision accuracy (for trials that are neither aborted nor unrewarded) at each hour of the day. This is summarized in Figure 4 below.

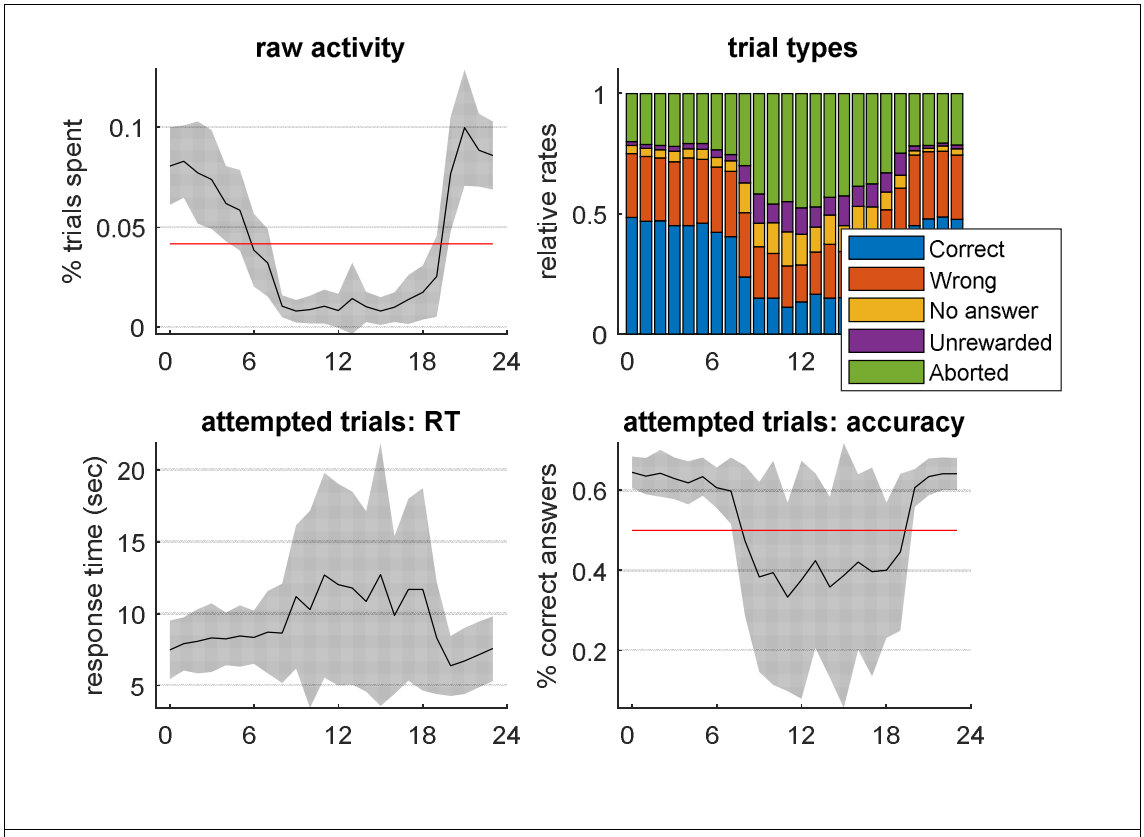

Figure 4: Effect of the circadian cycle. Upper-left panel: The mean percentage of trial occurrences (y-axis) is plotted against time of the day (from 12 am to 12 pm, x-axis). The red line indicates the equiproportional percentage of trial occurrences that would be observed if there was no circadian cycle (1/24). **Upper-right panel:** The relative rates of each trial type (y-axis) are plotted against time of the day (x-axis). **Lower-left panel:** The mean response time of correct/wrong trials (y-axis) is plotted against time of the day (x-

axis). **Lower-right panel:** The mean percentage of correct answers among correct/wrong trials (y-axis) is plotted against time of the day (x-axis). The red line indicates chance accuracy level (50%). In all panels, the grey shaded area indicates the standard deviation (across mice).

One can see that the circadian cycle has a profound effect on mice behaviour. In brief, during daytime, mice tend to disengage with the task (less than 18% of trials occur between 7am and 7pm), they often abort trials or fail to collect the reward, their average response time is longer (mean RT is 10.8 sec between 7am and 7pm, to be compared with 7.7 sec otherwise) and more variable (RT standard deviation is 1.5 sec between 7am and 7pm, to be compared with 0.7 sec otherwise), and their decision accuracy is lower (mean rate of correct answer is 41.5% between 7am and 7pm, to be compared with 63.1% otherwise). This is in fact expected, given the natural nocturnal rhythm of mice.

In what follows, we focus on correct and wrong trials of the main phase of the experiment, and ignore other types of trials. Note that the rate of aborted or unrewarded trials during nighttime is very low (about 20% on average). This is important, because this ensures that trials that eventually yield correct and wrong decisions dominate the experiment.

To begin with, we checked whether mice regulated their effort investment according to the model predictions. This reduces to analysing the impact of task factors (decision difficulty and reward at stake) as well as default/automatic preferences on RT data. To do so, we performed the following random-effect analysis, which reproduces the analysis of Figure 2 above (Holmes, Friston, and Friston 1998). First, we labelled each trial according to whether a win-stay/lose-switch policy would align with the cue or not. For each mouse, we then regressed RT data concurrently against reward trial difficulty, reward magnitude and the above default policy alignment indicator (all z-scored). We then reported the ensuing

regression coefficients at the group level, and performed one-sample t-tests. Next, we analysed the impact of task factors and default preference on decision accuracy. For each mouse, we then regressed RT data concurrently against reward trial difficulty, reward magnitude and the default policy alignment indicator (all z-scored). We then reported the ensuing regression coefficients at the group level, and performed one-sample t-tests. The results of these analyses are summarized in Figure 5 below.

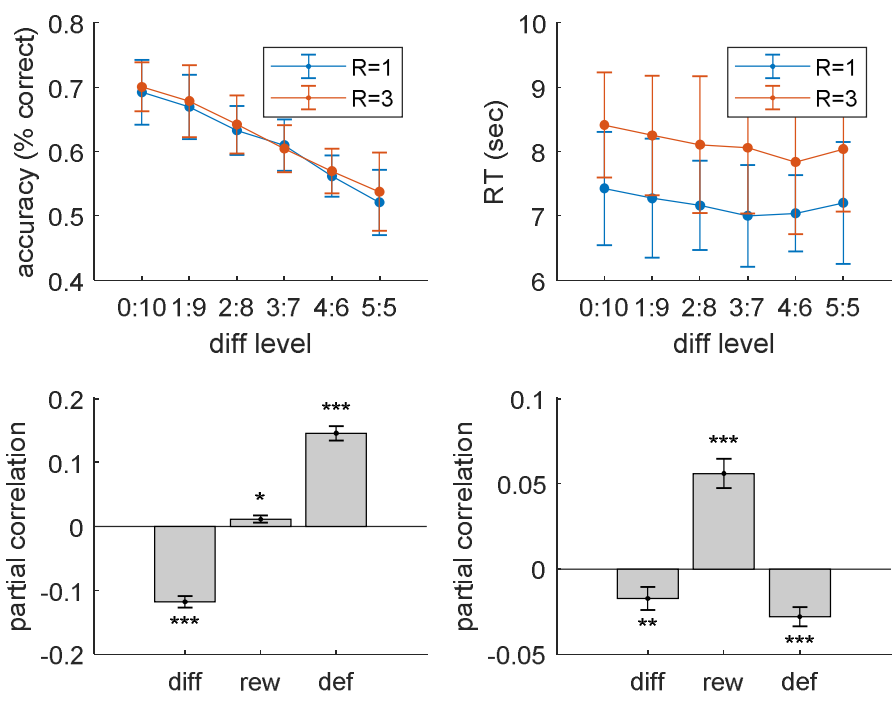

Figure 5: Effect of task factors and default policy on decision accuracy and response time. Upper-left panel: The mean percentage of correct responses (y-axis) is plotted against time decision difficulty (x-axis), for both reward levels (blue: one food pellet, red: three food pellets). **Upper-right panel:** Mean response time, same format. **Lower-left panel:** Multiple regression of decision accuracy: partial correlation coefficients (y-axis) are plotted for difficulty, reward and whether the default policy aligns with the cue. **Lower-right panel:** Multiple regression of response time, same format. In all panels, errorbars show the standard error around the mean (s.e.m.). In addition, one star, two stars and three stars indicate p-values that are below 0.05, 0.01 and 0.001, respectively.

As expected, response time significantly decreases with difficulty (partial correlation = -2.0% ± 0.7%, t_{25}= 2.53, p=0.01) and significantly increases with reward (partial correlation = 5.6% ± 0.9%, t_{25}= 6.55, $p<10^{-3}$). In other terms, mice spend most of their (deliberation) time on highly rewarded decisions, and shorten their deliberation time when decisions become difficult. In addition, response times are significantly shorter when the default/automatic policy aligns with the cue (partial correlation = -3.1% ± 0.7%, t_{25}= 4.23, $p<10^{-3}$). Under the MCD framework, this is because it takes more deliberation time to eventually override an inconsistent default/automatic policy.

Also, decision accuracy significantly decreases with difficulty (partial correlation = -11.8% ± 0.9%, t_{25}= 13. 3, $p<10^{-3}$) and significantly increases with reward (partial correlation = 1.1% ± 0.6%, t_{25}= 1.96, p=0.03). Note that although the reward effect is barely noticeable from the graphical summary of the raw performance data (cf. upper-left panel of Fig.3), it becomes much clearer when pre-processing performance data to remove artefactual confounds (such as, e.g., cue and response sides). This is important, because, under the MCD framework, the reward effect is the stigmata of the mechanism that adapts effort investments by adjusting the demand for decision confidence. In addition, decision accuracy is significantly higher when the default/automatic policy aligns with the cue (partial correlation = 14.6% ± 1.1%, t_{25}= 13.1, $p<10^{-3}$). All these analyses confirm our model predictions.

Under the MCD framework, mice may eventually override their default/automatic preference by investing attentional resources to the visual cue. More generally, we reasoned that the mouse's choice originates from a mixture of three components: a possible side bias, a default/automatic preference and a controlled/cue-related policy. The balance between these components is controlled by the amount of attentional resources that the mouse

invest to process the visual cue at each trial. This implies that the weight of each of these components on the mouse's choice should be modulated by response time, above and beyond task factors (reward and difficulty). To test this hypothesis, we performed a specific random effect analysis of the mice choice. For each mouse, we performed a logistic regression of her trial-by-trial binary response (1: left side, 0: right side) against three regressors: a constant (bias), a binary indicator of whether the win-stay/lose-switch policy would orient the choice towards the left (1) or the right (-1) side, and the cue orientation (1: left, -1: right). We also included the interaction of each of these regressors with reward, decision difficulty and response time (once orthogonalized with reward and decision difficulty). We then reported the ensuing regression coefficients at the group level, and test for statistical significance using one-sample t-tests. The results of this analysis are summarized in Figure 6 below, in terms of the mean regression coefficients.

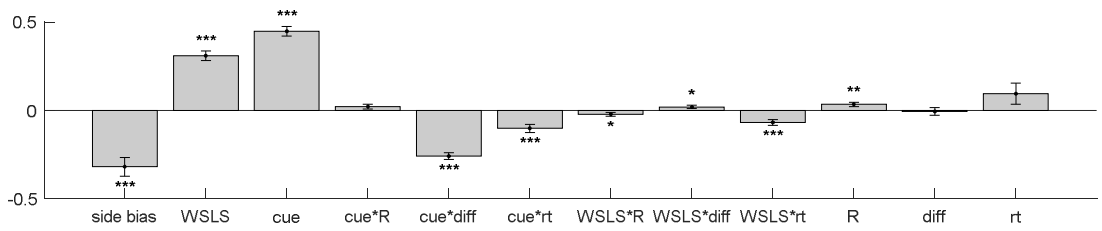

Figure 6: Logistic regression of trial-by-trial mouse answer side. See main text.

One can see that mice exhibit a significant bias towards the right side ($p<10^{-3}$), which is significantly reduced when reward increases ($p<10^{-3}$). The modulations of the side bias by difficulty and response time are not significant.

Also, mice answer side is significantly influenced by the default/automatic strategy ($p<10^{-3}$). Importantly, this influence is significantly reduced when reward increases ($p=0.03$), when

difficulty decreases (p=0.02), or when response time increase above and beyond task factors (p=0.002). In other terms, mice answers follow the default strategy mostly in trials where they invest few attentional resources to process the cue, which mostly happens when reward is low and/or difficulty is high.

Finally, mice answer side is significantly influenced by the information provided by the cue ($p<10^{-3}$), and significantly more so when difficulty decreases ($p<10^{-3}$) or when reward increases (p=0.05). Interestingly, the influence of the cue significantly *decreases* when response time increases above and beyond task factors ($p<10^{-3}$). Under the MCD framework, this is due to the paradoxical impact of response time, which may eventually impair decision accuracy (cf. Fig. 3).

To test this interpretation, we reproduced the analyses shown in Figure 3. For each mouse, we binned trials according to response time, separately for either difficulty or reward conditions, and measured the corresponding rate of correct answers. We then averaged these accuracy metrics across mice. Figure 7 below summarizes the results of this analysis.

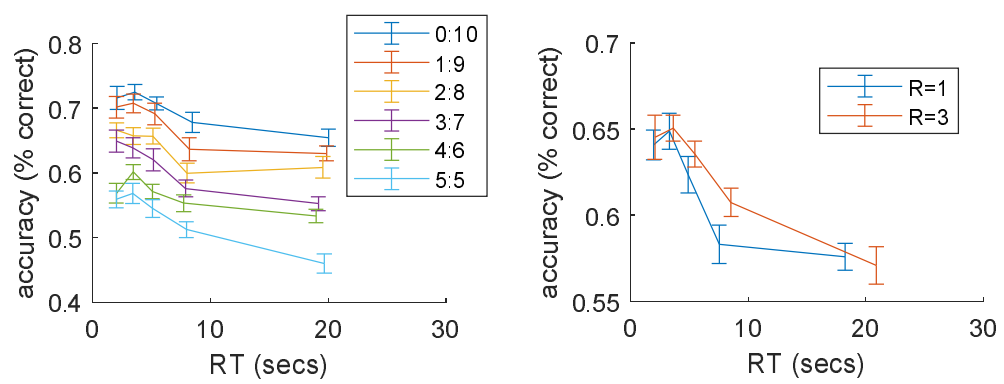

Figure 7: Relationship between response time and decision accuracy. Left panel: Average decision accuracy (y-axis) is plotted against response time (x-axis) for all difficulty levels (color code). **Right panel**: Average decision accuracy (y-axis) is plotted against response time (x-axis) for both reward conditions (color code). In all panels, errorbars show the standard error around the mean (s.e.m.).

As expected, decision accuracy globally decreases when response time increases, irrespective of the level of difficulty and/or reward. Interestingly, the reward effect on decision accuracy is now much clearer than when separating the trials according to decision difficulty (cf.upper-left panel of Fig. 5). More precisely, increasing reward seems to yield a gain in decision accuracy mostly for trials with intermediary response times. This concludes our data analysis.

Discussion

In this work, we have provided behavioural evidence that mice regulate their allocation of cognitive resources according to a confidence/effort a tradeoff. On the computational side of things, we have shown whether and how this eventually enables them to override their default/automatic policy. On the behavioural side of things, we have validated all the ensuing model predictions, in the context of an experimental setup with high ecological validity.

First, this work has implications for the neural architecture underlying the control of goal-directed behavior. Previous empirical studies relying on accumulation-to-bound models provided evidence that the adjustment of decision thresholds may be operated by the basal ganglia, in response to inputs from the orbitofrontal cortex or OFC *(Frank 2006)*. This is consistent with the fact that disruption of subthalamic nucleus or striatal activity may cause premature responding in humans *(Frank et al. 2007)* and in mice *(Burguière et al. 2013)*. In contrast, exhaustive neuroimaging meta-analyses indicated that cost/benefit computations involved in cognitive control very likely engage the anterior cingulate cortex or ACC *(Shenhav et al. 2013)*, which has known direct anatomical connections to the OFC and the basal ganglia *(Vogt and Paxinos 2014)*. This raises the questions of whether and how ACC

contributes to the control of decision making. We are currently addressing this issue using chemogenetic manipulations of ACC activity in mice.

Second, this work has implications for pathological neuropsychiatric conditions such as MDD[1], OCD[2] and ADHD[3], whose clinical symptoms can be understood in terms of specific forms of aberrant allocation of mental resources. For example, patients suffering from treatment-resistant depression typically exhibit variants of the so-called dysexecutive syndrome, i.e. co-occurring loss of cognitive/affective/behavioral control. Under the MCD framework, loss of control can have multiple causes, including: reduced sensitivity to reward, heightened sensitivity to effort costs, or distorted beliefs regarding the relationship between one's effort and one's resulting performance. A priori, these causes map to distinct symptomatic dimensions, i.e. anhedonia, low energy, and low self-esteem, respectively. Thus, should our neural investigations of MCD processes be successful, the dual computational/behavioral approach we present here may be fruitfully adapted to address outstanding issues in translational neuropsychiatry.

---

[1] *Major Depressive Disorder* is a mental disorder characterized by low mood, low self-esteem, anhedonia, low energy, and pain without a clear cause.

[2] *Obsessive-Compulsive Disorder* is a mental disorder characterized by the inability to prevent the occurrence of repetitive thoughts ("obsessions") and/or the urge to perform certain actions or routines ("compulsions"). Compulsive behavior may sometimes be aimed at obtaining relief from obsession-related anxiety.

[3] *Attention Deficit Hyperactivity Disorder* is a mental disorder characterized by reduced attentional control, excessive activity and impulsivity.

References

Benzina, Nabil, Karim N'Diaye, Antoine Pelissolo, Luc Mallet, and Eric Burguière. 2021. "A Cross-Species Assessment of Behavioral Flexibility in Compulsive Disorders." *Communications Biology* 4(1):1–12. doi: 10.1038/s42003-020-01611-y.

Birnberg, Jacob G., and Michael D. Shields. 1984. "The Role of Attention and Memory in Accounting Decisions." *Accounting, Organizations and Society* 9(3):365–82. doi: 10.1016/0361-3682(84)90020-5.

Blain, Bastien, Guillaume Hollard, and Mathias Pessiglione. 2016. "Neural Mechanisms Underlying the Impact of Daylong Cognitive Work on Economic Decisions." *Proceedings of the National Academy of Sciences* 113(25):6967–72. doi: 10.1073/pnas.1520527113.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2017. "Memory, Attention, and Choice." *Quarterly Journal of Economics*.

Burguière, Eric, Patrícia Monteiro, Guoping Feng, and Ann M. Graybiel. 2013. "Optogenetic Stimulation of Lateral Orbitofronto-Striatal Pathway Suppresses Compulsive Behaviors." *Science* 340(6137):1243–46. doi: 10.1126/science.1232380.

Ditterich, Jochen. 2006. "Evidence for Time-Variant Decision Making." *The European Journal of Neuroscience* 24(12):3628–41. doi: 10.1111/j.1460-9568.2006.05221.x.

Drugowitsch, Jan, Rubén Moreno-Bote, Anne K. Churchland, Michael N. Shadlen, and Alexandre Pouget. 2012. "The Cost of Accumulating Evidence in Perceptual Decision Making." *Journal of Neuroscience* 32(11):3612–28. doi: 10.1523/JNEUROSCI.4010-11.2012.

Drugowitsch, Jan, Valentin Wyart, Anne-Dominique Devauchelle, and Etienne Koechlin. 2016. "Computational Precision of Mental Inference as Critical Source of Human Choice Suboptimality." *Neuron* 92(6):1398–1411. doi: 10.1016/j.neuron.2016.11.005.

Feinberg, Eugene A., and Adam Shwartz. 2012. *Handbook of Markov Decision Processes: Methods and Applications*. Springer Science & Business Media.

Frank, Michael J. 2006. "Hold Your Horses: A Dynamic Computational Role for the Subthalamic Nucleus in Decision Making." *Neural Networks: The Official Journal of the International Neural Network Society* 19(8):1120–36. doi: 10.1016/j.neunet.2006.03.006.

Frank, Michael J., Johan Samanta, Ahmed A. Moustafa, and Scott J. Sherman. 2007. "Hold Your Horses: Impulsivity, Deep Brain Stimulation, and Medication in Parkinsonism." *Science* 318(5854):1309–12. doi: 10.1126/science.1146157.

Giguère, Gyslain, and Bradley C. Love. 2013. "Limits in Decision Making Arise from Limits in Memory Retrieval." *Proceedings of the National Academy of Sciences* 110(19):7613–18. doi: 10.1073/pnas.1219674110.

Greenwald, A. G., and M. R. Banaji. 1995. “Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes.” *Psychological Review* 102(1):4–27.

Griffiths, Thomas L., Falk Lieder, and Noah D. Goodman. 2015. “Rational Use of Cognitive Resources: Levels of Analysis Between the Computational and the Algorithmic.” *Topics in Cognitive Science* 7(2):217–229.

Harlé, Katia M., and Alan G. Sanfey. 2007. “Incidental Sadness Biases Social Economic Decisions in the Ultimatum Game.” *Emotion (Washington, D.C.)* 7(4):876–81. doi: 10.1037/1528-3542.7.4.876.

Heitz, Richard P. 2014. “The Speed-Accuracy Tradeoff: History, Physiology, Methodology, and Behavior.” *Frontiers in Neuroscience* 8. doi: 10.3389/fnins.2014.00150.

Holmes, A. P., K. J. Friston, and K. Friston. 1998. “Generalisability, Random Effects and Population Inference.”

Kahneman, Daniel, Paul Slovic, and Amos Tversky. 1982. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press.

Lee, Douglas G., Juliette Benon, William Hopper, Morgan Verdeil, Mathias Pessiglione, Fabien Vinckier, Sebastien Bouret, Marion Rouault, Raphael Lebouc, Giovanni Pezzulo, Christiane Schreiweis, Eric Burguiere, and Jean Daunizeau. 2023. “The Online Metacognitive Control of Decisions.” 2023.01.02.522463.

Lee, Douglas G., and Jean Daunizeau. 2021. “Trading Mental Effort for Confidence in the Metacognitive Control of Value-Based Decision-Making” edited by T. H. Donner, M. J. Frank, and A. Westbrook. *ELife* 10:e63282. doi: 10.7554/eLife.63282.

Lieder, Falk, Amitai Shenhav, Sebastian Musslick, and Thomas L. Griffiths. 2018. “Rational Metareasoning and the Plasticity of Cognitive Control.” *PLOS Computational Biology* 14(4):e1006043. doi: 10.1371/journal.pcbi.1006043.

Lim, S. L., J. P. O’Doherty, and A. Rangel. 2011. “The Decision Value Computations in the VmPFC and Striatum Use a Relative Value Code That Is Guided by Visual Attention.” *Journal of Neuroscience* 31(37):13214–23. doi: 10.1523/JNEUROSCI.1246-11.2011.

Lopez-Persem, Alizée, Philippe Domenech, and Mathias Pessiglione. 2016. “How Prior Preferences Determine Decision-Making Frames and Biases in the Human Brain.” *ELife* 5:e20317. doi: 10.7554/eLife.20317.

Marois, René, and Jason Ivanoff. 2005. “Capacity Limits of Information Processing in the Brain.” *Trends in Cognitive Sciences* 9(6):296–305. doi: 10.1016/j.tics.2005.04.010.

Martino, Benedetto De, Dharshan Kumaran, Ben Seymour, and Raymond J. Dolan. 2006. “Frames, Biases, and Rational Decision-Making in the Human Brain.” *Science* 313(5787):684–87. doi: 10.1126/science.1128356.

Massar, Stijn A. A., Árpád Csathó, and Dimitri Van der Linden. 2018. “Quantifying the Motivational Effects of Cognitive Fatigue Through Effort-Based Decision Making.” *Frontiers in Psychology* 9. doi: 10.3389/fpsyg.2018.00843.

Musslick, Sebastian, Amitai Shenhav, Matthew Botvinick, and Jonathan D Cohen. 2015. “A Computational Model of Control Allocation Based on the Expected Value of Control.”

Ozcimder, Kayhan, Biswadip Dey, Sebastian Musslick, Giovanni Petri, Nesreen K. Ahmed, Theodore L. Willke, and Jonathan D. Cohen. 2017. “A Formal Approach to Modeling the Cost of Cognitive Control.” *ArXiv:1706.00085 [q-Bio]*.

Palmer, John, Alexander C. Huk, and Michael N. Shadlen. 2005. “The Effect of Stimulus Strength on the Speed and Accuracy of a Perceptual Decision.” *Journal of Vision* 5(5):376–404. doi: 10.1167/5.5.1.

Papadimitriou, Christos H., and John N. Tsitsiklis. 1987. “The Complexity of Markov Decision Processes.” *Mathematics of Operations Research* 12(3):441–50. doi: 10.1287/moor.12.3.441.

Pirrone, Angelo, Tom Stafford, and James A. R. Marshall. 2014. “When Natural Selection Should Optimize Speed-Accuracy Trade-Offs.” *Frontiers in Neuroscience* 8. doi: 10.3389/fnins.2014.00073.

Porcelli, Anthony J., and Mauricio R. Delgado. 2009. “Acute Stress Modulates Risk Taking in Financial Decision Making.” *Psychological Science* 20(3):278–83. doi: 10.1111/j.1467-9280.2009.02288.x.

Porcelli, Anthony John, Andrea H. Lewis, and Mauricio R. Delgado. 2012. “Acute Stress Influences Neural Circuits of Reward Processing.” *Frontiers in Neuroscience* 6. doi: 10.3389/fnins.2012.00157.

Ratcliff, Roger, Trisha Van Zandt, and Gail McKoon. 1999. “Connectionist and Diffusion Models of Reaction Time.” *Psychological Review* 106(2):261–300. doi: 10.1037/0033-295X.106.2.261.

Schneider, Walter, and Richard M. Shiffrin. 1977. “Controlled and Automatic Human Information Processing: I. Detection, Search, and Attention.” *Psychological Review* 84(1):1–66. doi: 10.1037/0033-295X.84.1.1.

Shenhav, Amitai, Matthew M. Botvinick, and Jonathan D. Cohen. 2013. “The Expected Value of Control: An Integrative Theory of Anterior Cingulate Cortex Function.” *Neuron* 79(2):217–40. doi: 10.1016/j.neuron.2013.07.007.

Simen, Patrick, Jonathan D. Cohen, and Philip Holmes. 2006. “Rapid Decision Threshold Modulation by Reward Rate in a Neural Network.” *Neural Networks* 19(8):1013–26. doi: 10.1016/j.neunet.2006.05.038.

Sokol-Hessner, Peter, Colin F. Camerer, and Elizabeth A. Phelps. 2013. "Emotion Regulation Reduces Loss Aversion and Decreases Amygdala Responses to Losses." *Social Cognitive and Affective Neuroscience* 8(3):341–50. doi: 10.1093/scan/nss002.

Sullivan-Toole, Holly, Samantha DePasque, Bailey Holt-Gosselin, and Adriana Galván. 2019. "Worth Working for: The Influence of Effort Costs on Teens' Choices during a Novel Decision Making Game." *Developmental Cognitive Neuroscience* 37:100652. doi: 10.1016/j.dcn.2019.100652.

Tajima, Satohiro, Jan Drugowitsch, Nisheet Patel, and Alexandre Pouget. 2019. "Optimal Policy for Multi-Alternative Decisions." *Nature Neuroscience* 22(9):1503–11. doi: 10.1038/s41593-019-0453-9.

Tajima, Satohiro, Jan Drugowitsch, and Alexandre Pouget. 2016. "Optimal Policy for Value-Based Decision-Making." *Nature Communications* 7:12400. doi: 10.1038/ncomms12400.

Thorngate, Warren. 1980. "Efficient Decision Heuristics." *Behavioral Science* 25(3):219–25. doi: 10.1002/bs.3830250306.

Usher, Marius, and James L. Mcclelland. 2001. "On the Time Course of Perceptual Choice: The Leaky Competing Accumulator Model." *Psychological Review* 550–592.

Vogt, Brent A., and George Paxinos. 2014. "Cytoarchitecture of Mouse and Rat Cingulate Cortex with Human Homologies." *Brain Structure & Function* 219(1):185–92. doi: 10.1007/s00429-012-0493-3.

Wang, Zheng, and Jerome R. Busemeyer. 2016. "Interference Effects of Categorization on Decision Making." *Cognition* 150:133–49. doi: 10.1016/j.cognition.2016.01.019.

Wyart, Valentin, and Etienne Koechlin. 2016. "Choice Variability and Suboptimality in Uncertain Environments." *Current Opinion in Behavioral Sciences* 11:109–15. doi: 10.1016/j.cobeha.2016.07.003.

Contributions

JD, EB conceptualised the research; CS conducted experiments; CS, ME, EB, JD analysed data; CS, EB, JD wrote the article; CS, EB, JD edited the manuscript.

Acknowledgements

All animal work was conducted at the PHENO-ICMice facility. This work furthermore benefitted from the equipment and services from the ICM FabLab (Pierre Tissier). This work

was realised with the following funding: ICM-IHU Big Brain Theory Project (BBT-ICM-IHU 2016 + 2018) (JD, EB), Agence Nationale de la Recherche (ANR-xx-ICM-DECIMOTIV) (JD, EB); and the L'Oréal-UNESCO Fellowship 2016 (CS). The core facilities were supported by "Investissements d'avenir" (ANR-10-IAIHU-06 and ANR-11-INBS-0011-NeurATRIS) and "Fondation pour la Recherche Médicale".

Conflict of interest.

The authors declare no conflict of interest.