communications psychology

Article

A Nature Portfolio journal

https://doi.org/10.1038/s44271-025-00282-x

The temporal dynamics of metacognitive experiences track rational adaptations in task performance

Check for updates

Luc Vermeylen[1,2] ✉, Senne Braem[1], Ivan I. Ivanchei[1,3], Kobe Desender[2], J. M. García-Román[4], Carlos González-García[4], María Ruz[4] & Wim Notebaert[1]

Human task performance elicits diverse subjective metacognitive experiences, such as boredom, effort, fatigue, and frustration, which are considered to play important roles in the monitoring and regulation of cognitive processes. Yet, their specific contributions to task performance remain poorly understood. Therefore, we investigated the temporal dynamics underlying these metacognitive experiences and latent cognitive processes supporting task performance. We used a time-on-task design using a conflict task and analyzed the data using a comprehensive approach encompassing behavioral, model-based, subjective, and neural measures (N = 111). Our results show that changes in cognitive processes can be understood as a rational attempt to optimize task performance and that distinct metacognitive experiences are related to different aspects of this rational endeavor. These findings suggest that metacognitive experiences act as tools for humans to gain insights into the optimality of their cognitive performance.

Humans tend to have several subjective metacognitive experiences when performing cognitive tasks[1,2]. For example, we often have an internal sense of certainty or confidence with respect to our performance—a well-studied feature of decision-making that facilitates performance in the absence of feedback[3–6]. However, task performance tends to be accompanied by a much broader spectrum of metacognitive experiences, including feelings of boredom, effort, fatigue, and frustration. Consider, for example, a young student immersed in an examination setting: the challenging nature of the exam may elicit a sense of frustration, prompting the student to pause, reevaluate, and adopt a more cautious strategy. Hence, these experiences are thought to provide valuable insights into task performance and serve as a guide for strategic adjustments[3–7]. However, the precise relationship between such experiences and cognitive task mechanisms remains unclear. Nevertheless, a deeper understanding of how (and which) metacognitive experiences influence task performance could pave the way for targeted training programs aimed at enhancing awareness of the experiences that exert the most substantial impact on performance[8,9].

One of the common challenges in maintaining optimal levels of task performance is the presence of irrelevant stimuli in the environment that capture our attention away from the primary task. For instance, imagine the young student distracted by subtle hallway noises, a flickering light, or a ticking clock. To regain their focus and align their behavior with their academic goal, they first need to register the distraction and then take steps to restore their concentration. This ability to adapt our information processing in response to environmental changes falls under the domain of cognitive control and encompasses the mental processes essential for managing goal-directed behavior in the presence of irrelevant stimuli or conflict, where mutually incompatible representations compete for control over action. While traditional accounts described conflict as the trigger for behavioral adaptations[8], more recent accounts have argued that the negative metacognitive experiences associated with cognitive control and conflict might be the trigger instead[9–11]. These experiences include aversive reactions towards the conflicting stimuli themselves (what we will refer to as "conflict aversiveness"[12]), but also other feelings of effort, boredom, frustration, and fatigue[13,14]. Therefore, contemporary accounts of cognitive control take into account that there are experiential costs (and benefits) associated with investing cognitive control, and propose that these subjective experiences may serve as influential forces driving behavioral regulation[9,10,15–17].

While it has become evident that cognitive control and conflict elicit a broad palette of negative metacognitive experiences, it remains unclear which specific experiences offer insights into the underlying cognitive processes of task performance. Equally uncertain is the way these experiences act as sources for making adaptive changes to guide and optimize performance. While attempts are being made to construct normative

[1]Department of Experimental Psychology, Ghent University, Ghent, Belgium. [2]Brain and Cognition, KU Leuven, Leuven, Belgium. [3]Centre for Research in Cognition and Neurosciences, Université Libre de Bruxelles, Brussels, Belgium. [4]Mind, Brain, and Behavior Research Center, University of Granada, Granada, Spain. ✉e-mail: vermeylenl@gmail.com

computational accounts of how such subjective experiences could arise and what their exact signaling and regulatory function could be[18], there is a scarcity of empirical data that simultaneously tracks the evolution of task performance and the evolution of multiple subjective experiences over time, which is crucial for advancing theoretical understanding. The dimension of time is important here, as these metacognitive experiences and task performance are not static but constantly evolving. Therefore, in the present study, we aim to address this gap by investigating which type of metacognitive experiences were related to changes in well-studied parameters of cognitive task performance. Across two studies (behavioral study, $N = 66$; EEG study, $N = 45$), human participants engaged in a standard conflict Flanker task over 18 blocks of time-on-task (TOT), spanning nearly 2 h. After each block, participants reported their metacognitive experiences of conflict aversiveness, boredom, effort, fatigue and frustration (Fig. 1). We analyzed this TOT design using a comprehensive approach encompassing behavioral, subjective, and neural measures, as well as several theory-driven computational and analytic approaches, with the aim of providing insight into the dynamic interplay between metacognitive experiences and task performance.

Methods

Participants

Seventy-one participants participated in the online behavioral experiment, which was conducted on Prolific Academic and took on average 117.7 min to complete (participants were compensated at 15.12 pounds in total or 7.56 per hour). All participants performed the experiment in English (only participants whose first language was English were allowed to participate). Five participants were excluded from the analyses due to not adhering to the instructed break time (i.e., 30 s, more than 2.5 SD above the mean break time), missing too many responses (more than 2.5 SD above the mean number of missed responses), responding too slow (higher than 2.5 SD above the mean reaction time (RT)) or making too many errors (higher than 2.5 SD above the mean error rate). The average age of the remaining 66 participants (31 male and 33 female participants, 2 unknowns, assessed by self-report) was 29.73 years (SD = 5.26, min = 18, max = 36). Participants had to agree with a consent form before starting the experiment. Forty-nine participants participated in the EEG experiment, which was conducted at the Research Center for Mind, Brain and Behavior at the University of Granada and took on average 116.8 min to complete (participants were compensated 35 EUR for the whole session, including preparation time). Four participants were excluded from the analyses due to not adhering to the exclusion criteria specified above. The average age of the remaining 45 participants (16 male and 29 female participants, assessed by self-report) was 21.42 years (SD = 2.75, min = 18, max = 31). Thirty-nine participants performed the experiment in Spanish, and 6 participants performed the experiment in English. Participants signed an informed consent form before starting the experiment. All experiments were approved by the Ethical Committee of Ghent University Psychology and Educational Sciences or the University of Granada. This study was not preregistered.

Design, paradigm, and procedure

The experiment was programmed using jsPsych[19] in the online setting and PsychoPy 3[20] in the EEG setting. Participants were instructed that they were going to perform a simple decision-making task almost continuously over a period of 2 h. In the online setting, they were encouraged to turn off notifications from their smartphone and computers. Participants were instructed that they should not miss more than 5% of the trials, make more than 30% errors, or respond too fast (response < 100 ms) on more than 5% of the trials.

Next, participants were instructed on the Flanker task (based on ref. 21) which stated that they had to press S (online) or left control (EEG) when the central target arrow pointed to the left, and press L (online) or right control (EEG) when the central target arrow pointed to the right. Participants started each trial with a fixation display (mean = 700 ms, randomly drawn from a uniform distribution between 600 and 800 ms). Next, the horizontal flankers appeared for a short duration (83 ms), followed by the central target (33 ms). After this, the stimulus disappeared from the screen and participants were required to respond within a 1200 ms window (see Fig. 1 for a visual depiction of the timing and sequence of events in the Flanker task). They were instructed to respond as fast and accurately as possible. The next trial started immediately after a response was made. This was followed by a short practice version of the Flanker task (60 trials) where trial-by-trial feedback was provided ("Too slow," "Too fast," "Correct!," "Wrong!") ending in a display of feedback showing their average RT and accuracy, the percentage of missed responses and the percentage of impossibly fast responses. Participants needed to achieve an accuracy of 70% in this practice phase to continue the experiment.

Next, participants were told that they would be questioned about their mental state after each block of the Flanker task. Participants were told that their feelings towards the task and stimuli might change over the course of the experiment, but they do not necessarily have to, and that it is simply important to provide an accurate and reliable report of their subjective experience at each point in time. Next, participants were explained the difference between congruent and incongruent stimuli, so they could give their informed preference about whether they preferred being presented with congruent ("same") or incongruent ("different") stimuli (i.e., conflict aversiveness). A visual analog scale (VAS) was presented that started in the center (which was labeled as indifferent) and had "extreme preference for SAME" and "extreme preference for DIFFERENT" on each end of the VAS (left and right, respectively, see Fig. 1 for a simplified visual depiction). Next, they received VAS polling their levels of effort ("How hard did you work in the task?"; not hard—very hard), frustration ("How annoyed are you with the task?"; not annoyed—very annoyed), boredom ("How boring do you find task?"; not boring—very boring) and fatigue ("How mentally fatigued

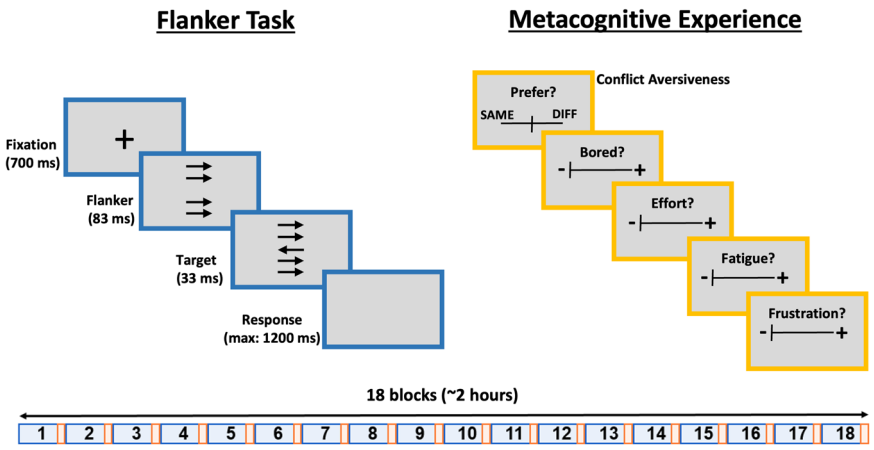

Fig. 1 | Task design. Participants performed a Flanker task (adapted from Fischer et al.[21]) for 18 blocks, which lasted roughly 2 h. After each block, participants reported on their metacognitive experiences (conflict aversiveness, boredom, effort, fatigue, frustration) using visual analog scales. To measure conflict aversiveness, participants reported whether they preferred being presented with congruent ("SAME") or incongruent ("DIFF") trials by moving the marker away from the middle (which is the indifference point, i.e., no preference). For the other measures of metacognitive experience, the scale always started at the left-most position, denoting that the experience was not present at all (whereas the right-most position denoted that the experience was highly present).

are you now?"; not fatigued—very fatigued). These questions were based on previous work[14].

Finally, participants started the actual experiment, which consisted of 18 blocks of 260 trials each in which congruency (congruent, incongruent) and target direction (left, right) were always balanced. After each block, participants received VAS polling their metacognitive experience, and participants were allowed to take a short break of 30 s (and were instructed to strictly adhere to this break time).

Time-on-task analyses

All behavioral and modeling analyses were performed in R (4.1.3). For the raw RT analyses, we removed the first trial of each block and errors. We also removed trials with a RT below 200 ms (note that RT measurement was locked to the Flanker onset, which preceded target onset) and only kept trials between the 1st and 99th quantile (this rather minimal data trimming procedure was applied because of the subsequent diffusion modeling which models the whole RT distribution). For the metacognitive experience measures, we recoded the conflict aversiveness measure to increase interpretability. The responses to this scale were originally recorded from 0 to 100, with 50 being the indifference point. We rescaled this to vary from −50 to 50 so that zero became the indifference point, and positive values align with a more negative evaluation of conflict (a preference for congruent trials), and negative values align with a less negative evaluation of conflict (a preference for incongruent trials). In this way, higher values of conflict aversiveness denote a stronger dislike of incongruent stimuli. All other VAS were recorded from 0 to 100.

The behavioral data were modeled using a variant of the drift diffusion model (DDM). In the vanilla DDM, decisions are formed through the accumulation of evidence over time, which is governed by three parameters: First, the rate or efficiency of the accumulation is determined by the drift rate. Second, accumulation stops at a certain decision boundary, which quantifies one's level of cautiousness and accounts for the speed-accuracy tradeoff. Third, additional processes that contribute to the RT outside the decision itself are captured by the non-decision time. However, the standard DDM fails to fully account for the patterns observed in RT distributions of conflict tasks[22]. To account for such patterns, instead of capturing the process of evidence accumulation with a single parameter, it can be further decomposed into contributions from the relevant and irrelevant stimulus dimensions (Diffusion Model for Conflict Tasks[23]). The relevant dimension, such as the target in a Flanker task, is assumed to have a constant drift rate. On the other hand, the irrelevant dimension, such as the distractors in a Flanker task, has a time-varying drift rate modeled using a gamma impulse response, which varies in terms of its peak amplitude ("of the automatic activation function"). The latter parameter represents how much the irrelevant dimension influences decision formation in a short-lived manner, i.e., the degree of "irrelevant capture" or the performance costs due to conflict (often termed "interference effects" in the context of RT or accuracy measures). However, when the decision boundary changes with TOT, peak amplitude is no longer a pure measure of irrelevant capture, as its influence on decision formation depends on the height of the decision boundary. For instance, a peak amplitude value of 20 evidence units indicates more interference when the boundary is set at 30 compared to when it is set at 60. Therefore, to provide a more reliable assessment of irrelevant capture, it is crucial to consider changes in the decision boundary alongside peak amplitude. Thus, we introduced a novel metric, the peak-to-bound ratio (PTB-ratio). By normalizing peak amplitude by the decision boundary (dividing peak amplitude by the boundary), we obtain a measure that reflects the impact of the irrelevant dimension independently of the decision boundary setting. Consequently, this approach ensures that the parameters uniquely align with their proposed underlying mechanisms.

The model was fit using the DMCfun package[24] on each block of data (260 trials) using the Root Mean Square Error cost function that quantifies the discrepancy between the model predictions and observed data on 19 bins (percentiles) of the correct RT distributions for congruent and incongruent trials and 5 bins of the Conditional Accuracy Function (see refs. 23,24 for more details on the cost function). As there is no analytic solution for this diffusion model, predictions of the model were retrieved using a Monte Carlo simulation and were based on 20000 trials for each congruency condition. The cost function was minimized using the Differential Evolution algorithm, which was run for 500 iterations (fitting one dataset, i.e., one block of one subject, took roughly 45 min, resulting in a total of 61 days of serial computing time). We used the following boundaries for the parameter space: drift rate (0.2–1.2), decision boundary (30–110), non-decision time (200–600), peak amplitude (10–110). Peak latency was fixed to 28, and variability in non-decision time was fixed to 25. These fixed values are based on grand average values that were received after fitting the model per three blocks rather than on every block. We fixed these parameters as they were not our main interest (peak latency is mostly related to differences between conflict tasks, not within a given task) and lead to lower parameter recovery performance when fitting the model per block (rather than per three blocks, see Supplementary Fig. S1 for parameter recovery and fit assessment). The diffusion constant was fixed to 4 (ref. 23). Note that the plausible parameter ranges for this model differ from the standard DDM. PTB-ratio was calculated by dividing the peak amplitude by the decision boundary.

To investigate the effect of TOT, we ran a Linear Mixed Effects model predicting the DDM parameters, the metacognitive experiences, and the EEG indices (see below) by the effect of TOT (i.e., block: 1–18). We included random intercepts and random slopes for the effect of TOT up to the first degree (either linear or logarithmic, depending on the best model). We tested multiple models of TOT (linear, logarithmic, quadratic, and cubic) and chose the best model based on the lowest Bayesian Information Criterion (BIC). All performed statistical tests are two-sided.

Intra-individual relationships

To investigate the intra-individual relationships between the time series of the DDM parameters and the metacognitive measures, we used Multivariate Bayesian Linear Mixed Effects models (using the brms package[25], MCMC settings: 8 chains of 4000 iterations with 1000 as warmup) to predict all the DDM parameters' time series by all the metacognitive experience' time series while controlling for Experiment (Behavioral, EEG). We included random slopes for all predictors but did not need random intercepts, as the means of all time series were constrained to be zero due to the standardization of each subject's time series. This allowed us to exclude inter-individual differences from this analysis (see Supplementary Figs. S2 and S3) and conveniently returns standardized beta weights, which allows for a comparison of the relative magnitude of the different predictors. We also modeled the correlational structure between the dependent variables by including estimations of the correlations between the random slopes of the different dependent variables as well as correlations between the residuals of the multiple dependent variables. Weakly informative priors were used for all estimated beta weights (normal distribution with mean zero and a standard deviation of 2.5). We report the mean of the posterior distribution as the estimate for the standardized beta weight (which also acts as an effect size due to its standardized nature), together with its standard error (SE) and its 95% Highest Density Interval (HDI). In addition, we report the probability of direction (pd), which is considered an index of effect existence and represents the certainty with which an effect goes in a particular direction[26]. A *pd* value of 97.5% corresponds approximately to the two-sided *p* value of 0.05.

These models were conducted on the detrended time series, or what we coin "local fluctuations." The observed data was detrended by subtracting the best linear or logarithmic trend (depending on the best fit, as discussed above) from the observed data, i.e., leaving us with the residual time series. Thus, the beta weights reflect similarities in terms of sub-trend fluctuations between two given time courses—showing relationships that cannot be distorted by similarities in the global trend. Data distributions for the regressions were assumed to be normal, but this was not formally tested.

Optimality simulations and path analyses

The reward rate landscapes for the optimality analyses were performed by calculating the pseudo-reward rate for different combinations of peak amplitude (20–120, in increasing steps of 0.1) and decision boundary (20–120) given the average values for the other DDM parameters or a range of decision boundary (20–120) for each block given the other DDM parameters for that block. The pseudo-reward rate was calculated in the following manner (refs. 27,28):

$$Reward\ Rate = \frac{1 - ER}{DT + NDT + ITI} \quad (1)$$

Where ER refers to the error rate, DT to the decision time, NDT to the non-decision time, and ITI to the inter-trial interval. This reward rate was calculated for congruent and incongruent trials separately and then summed. The calculation of the reward rate for each combination was based on 20,000 simulations from the DDM. The optimal trajectory was defined as the maximum trajectory in the resulting reward rate landscape.

The path analyses were performed by Multivariate Bayesian Linear Mixed Effects modeling using the brms package. The different paths are described by different regression equations, which are estimated jointly. We included random slopes for all predictors but did not need random intercepts (for the same reason as above), along with correlation among the random slopes of independent variables between the multiple regression equations. Again, we used weakly informative priors for all betas (normal distribution with mean zero and a standard deviation of 2.5). Below, we present the model equations (in the lme4/brms style syntax) for the final path model:

$$\begin{aligned} Conflict\ Aversiveness = {} & 0 + Peak\ Amplitude \\ & + (0 + Peak\ Amplitude | p | Subject) \end{aligned} \quad (2)$$

$$\begin{aligned} Frustration = {} & 0 + Peak\ Amplitude + Conflict\ Aversiveness \\ & + (0 + Peak\ Amplitude + Conflict\ Aversiveness | p | Subject) \end{aligned} \quad (3)$$

$$\begin{aligned} Fatigue = {} & 0 + Boundary + Frustration \\ & + (0 + Boundary + Frustration | p | Subject) \end{aligned} \quad (4)$$

$$\begin{aligned} Boundary = {} & 0 + Peak\ Amplitude + Conflict\ Aversiveness + Frustration \\ & + (0 + Peak\ Amplitude + Conflict\ Aversiveness \\ & + Frustration | p | Subject) \end{aligned} \quad (5)$$

The part of the equation in parentheses is related to the random effects structure. The notation (X|p|Y) denotes that there are random slopes of independent variable *X* for each grouping of *Y* (i.e., subject) that are allowed to be correlated (|p|) across the multiple dependent variables. The zeros denote the absence of fixed and random intercepts. The indirect effects are calculated by multiplying the posteriors of the betas from the two respective paths.

EEG data acquisition and preprocessing

EEG was recorded from 64 active electrodes (actiCap snap, Brain Products). Impedances were brought below 10 kΩ before starting the experiment. EEG activity was referenced online to the C1 electrode, and signals were digitized at a sampling rate of 500 Hz. The EEG recordings were analyzed with MNE (ref. 29, version 0.24.1) in Python (version 3.9.7). The signal was band-pass filtered between 0.1 and 40 Hz, downsampled to 250 Hz, and epoched from −1000 to 1500 ms around Flanker onset. Independent Component Analysis (ICA) was estimated on a more highly filtered version of the data (1 Hz high-pass filter), as this leads to more stable solutions of the ICA algorithm and is recommended by MNE. The blink component was automatically identified (and subtracted from the data) as the component with the highest correlation with the two most frontal electrodes (Fp1, Fp2) in which blink activity is most prominent. The automatic selection was manually inspected for correctness (in four subjects, we manually identified a second blink component). Next, an automatic trial rejection procedure (autoreject, see ref. 30 for details of the algorithm, version 0.2.2) and a fixed peak-to-peak threshold (150 muV) were used to remove trials with artifacts ($M = 3.70\%$, $SD = 4.21\%$). The resulting epochs were visually inspected (raw epochs and power spectra) to confirm that the procedure worked as expected. Next, we interpolated bad channels (0 channels: $N = 27$, 1 channel: $N = 8$, 2 channels: $N = 4$, 3 channels: $N = 6$, 4 channels: $N = 4$) and re-referenced the data to the average of all electrodes.

The Lateralized Readiness Potential (LRP) was calculated for both congruent and incongruent trials using the double difference method ($[C3_{left} - C4_{left}] - [C3_{right} - C4_{right}]$), which results in a positive polarity when the correct response hand is activated and a negative polarity when the incorrect response hand is activated. The signal was baseline corrected from −200 to 0 ms and spatially filtered by computing the Laplacian, which decreases the effect of volume conductivity and produces a more distinct topography. Next, we extracted features from the LRP that could be considered neural surrogates for the DDM parameters. First, we computed the slope of the average LRP between the time of the first incongruent negative peak and the second incongruent peak ("LRP slope"). This timing was chosen to avoid contamination of the signal by impact from the irrelevant dimension (which mainly impacts motor preparation before the first incongruent negative peak). Second, we extracted the positive peak amplitude of the LRPs ("LRP amplitude") as an approximation to the decision boundary. All neural peak amplitude measures were based on averaging across a window of 50 ms before and after the grand average peak amplitude. Third, the onset latency of the LRPs ("LRP latency") was estimated using segmented (piecewise) regression. Based on visual inspection of the grand average data, we restricted the segmented regression to look for one breakpoint between 0 and 276 ms. Visual inspection indeed revealed that the segmented regression was successful at estimating the onset of motor preparation. Last, we used the (absolute) negative peak amplitude on incongruent trials ("LRP dip") as a neural proxy for irrelevant capture.

Finally, we used a representational similarity analysis logic to evaluate whether the overlap between the model parameters and the EEG indices across time is larger than what could be expected based on chance. To this end, we constructed a null distribution by simulating fifty thousand random beta matrices, and we computed the Spearman correlation between the random matrix and the observed beta matrix.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Results

Diffusion modeling of task performance

To increase power, we collapsed the sample across the behavioral and EEG experiment where possible ($N = 111$). We observed clear behavioral changes with TOT (Fig. 2A). Participants became faster while maintaining a stable error rate and suffered less irrelevant capture (i.e., reductions in both RT and error rate congruency effects, see Supplementary Note 1 for details). While previous TOT studies typically relied on analyses of average RTs and error rates (e.g., refs. 31–33), sequential sampling models of decision formation, such as extensions of the DDM[23], allow us to model task performance of cognitive control tasks in terms of theoretically grounded cognitive processes[34]. Here, decision formation is governed by four main parameters: drift rate, decision boundary, non-decision time, and peak amplitude of the automatic activation function (see "Methods"). To investigate the effects of TOT, we examined the impact of TOT on these parameters by fitting separate Linear Mixed Effect Models for linear, logarithmic, quadratic, or cubic effects of TOT. We chose the model with the lowest BIC (see "Methods").

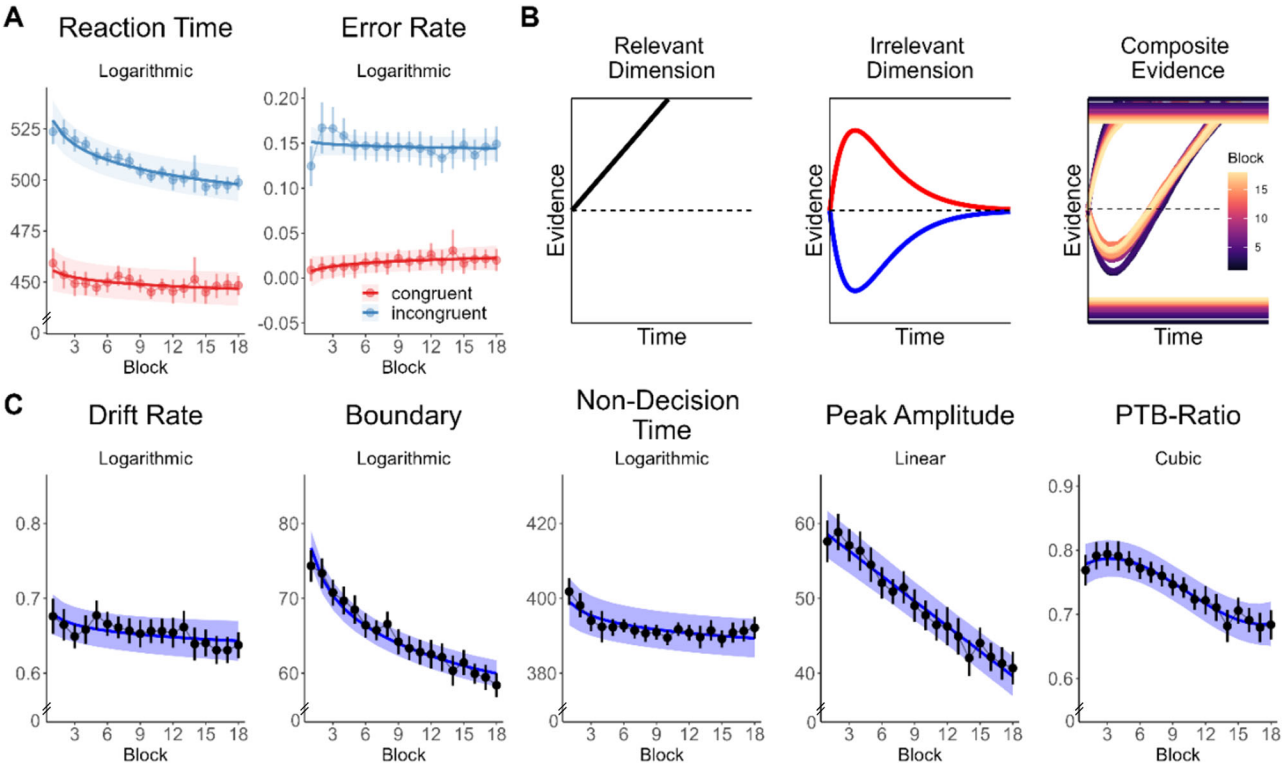

Fig. 2 | Task performance. A Time-on-task effects in behavior (RT and accuracy). **B** Functional form of the latent decision variable for the relevant (constant drift rate) and irrelevant dimension (time-varying drift rate following a gamma impulse function with its main parameter peak amplitude) and evolution of the composite latent decision variable and decision boundary over time-on-task. The upper boundary represents the correct response, while the lower boundary represents the error response. Red denotes activation of the irrelevant dimension on congruent trials, while blue denotes activation of the incongruent trials. **C** Time-on-task effects on the DDM parameters. The subtitle denotes the best model for time-on-task (linear, logarithmic, quadratic, or cubic) based on BIC. PTB-Ratio peak-to-bound ratio. Error bars represent the 95% within-subject CI. Shading represents the 95% CI for the predicted effect of time. $N = 111$ participants.

The results revealed slight logarithmic decreases in drift rate, β (standardized Beta) = −0.058, SE = 0.023, 95% confidence interval (CI) [−0.103, −0.014], $t(110) = -2.57$, $P = 0.012$, and non-decision time, $\beta = -0.084$, SE = 0.022, 95% CI [−0.128, −0.041], $t(110) = -3.80$, $P < 0.001$, showing that the rate of evidence accumulation decreased, and less time was spent on processes outside decision formation over time (Fig. 2C). Decision boundary showed a strong logarithmic decrease, $\beta = -0.325$, SE = 0.023, 95% CI [−0.371, −0.279], $t(110) = -13.88$, $P < 0.001$, suggesting participants became less cautious. Furthermore, peak amplitude exhibited a strong linear decrease, $\beta = -0.300$, SE = 0.023, 95% CI [−0.345, −0.255], $t(110) = -13.09$, $P < 0.001$. To more correctly assess the impact of the irrelevant dimension on decision formation in the context of changing decision boundaries, we also calculated the PTB-ratio (see "Methods"). This metric showed a cubic pattern: an early increase followed by a decrease, plateauing towards the end ($\beta_{linear} = -0.305$, SE = 0.037, 95% CI [−0.392, −0.217], $t(427.33) = -8.20$, $P < 0.001$; $\beta_{quadratic} = -0.022$, SE = 0.015, 95% CI [−0.049, 0.006], $t(1774) = -1.90$, $P = 0.058$; $\beta_{cubic} = -0.111$, SE = 0.029, 95% CI [0.042, 0.180], $t(1774) = 3.85$, $P < 0.001$).

Together, these diffusion modeling results show that participants mainly became less cautious while the degree of irrelevant capture decreased. For more details on the diffusion model, fit assessment, and parameter recovery, see "Methods" and Supplementary Fig. S1.

A neural proxy for the model-based latent decision variable

Next, we leveraged human EEG recordings to establish a neural proxy for the model-based latent decision variable of the DDM. Previous research has shown that preparatory motor signals, such as the LRP, encode this latent decision variable[21,35–40]. This means that we can use the LRP to explore the neural plausibility of the extended DDM.

In our EEG sample ($N = 45$), we observed a clear correspondence between the LRPs and latent decision variables of the DDM. Specifically, motor preparation demonstrated an immediate build-up towards the correct response on congruent trials, while exhibiting a "dip" towards the incorrect response (i.e., irrelevant capture) before favoring the correct response on incongruent trials (Fig. 3A). Next, we extracted features from these LRP waveforms (see "Methods" for details) to test whether they function as neural proxies for the DDM parameters. First, we extracted the average "LRP slope" as a proxy for the drift rate. Second, the positive peak amplitude of the LRPs, referred to as "LRP amplitude," was extracted to approximate the decision boundary. Third, the onset latency of the LRPs ("LRP latency") can function as a surrogate for the non-decision time. Last, the early negative peak amplitude on incongruent trials, denoted as "LRP dip," was chosen as a neural proxy for the degree of model-based irrelevant capture. As a first indication that these neural indices matched their model-based counterparts, we showed that these neural indices showed similar logarithmic (LRP slope: $\beta = -0.139$, SE = 0.031, 95% CI [−0.200, −0.078], $t(44) = -4.48$, $P < 0.001$; LRP amplitude: $\beta = -0.107$, SE = 0.029, 95% CI [−0.164, −0.050], $t(44) = -3.69$, $P < 0.001$; LRP latency: $\beta = -0.101$, SE = 0.044, 95% CI [−0.188, −0.014], $t(44) = -2.27$, $P = 0.029$) or linear (LRP dip: $\beta = -0.227$, SE = 0.035, 95% CI [−0.297, −0.157], $t(44) = -6.35$, $P < 0.001$) decreases over the course of the experiment, mirroring their model-based counterparts (Fig. 3B).

While the similarity between the model parameters, neural indices, and their underlying latent decision variables can be appreciated qualitatively, we sought to quantitatively evaluate this. To this end, we examined the intra-individual similarity between the time series of the DDM parameters and the neural indices. Specifically, we used a Bayesian Multivariate Linear Mixed Effects Model to predict local fluctuations (i.e., the detrended or residual

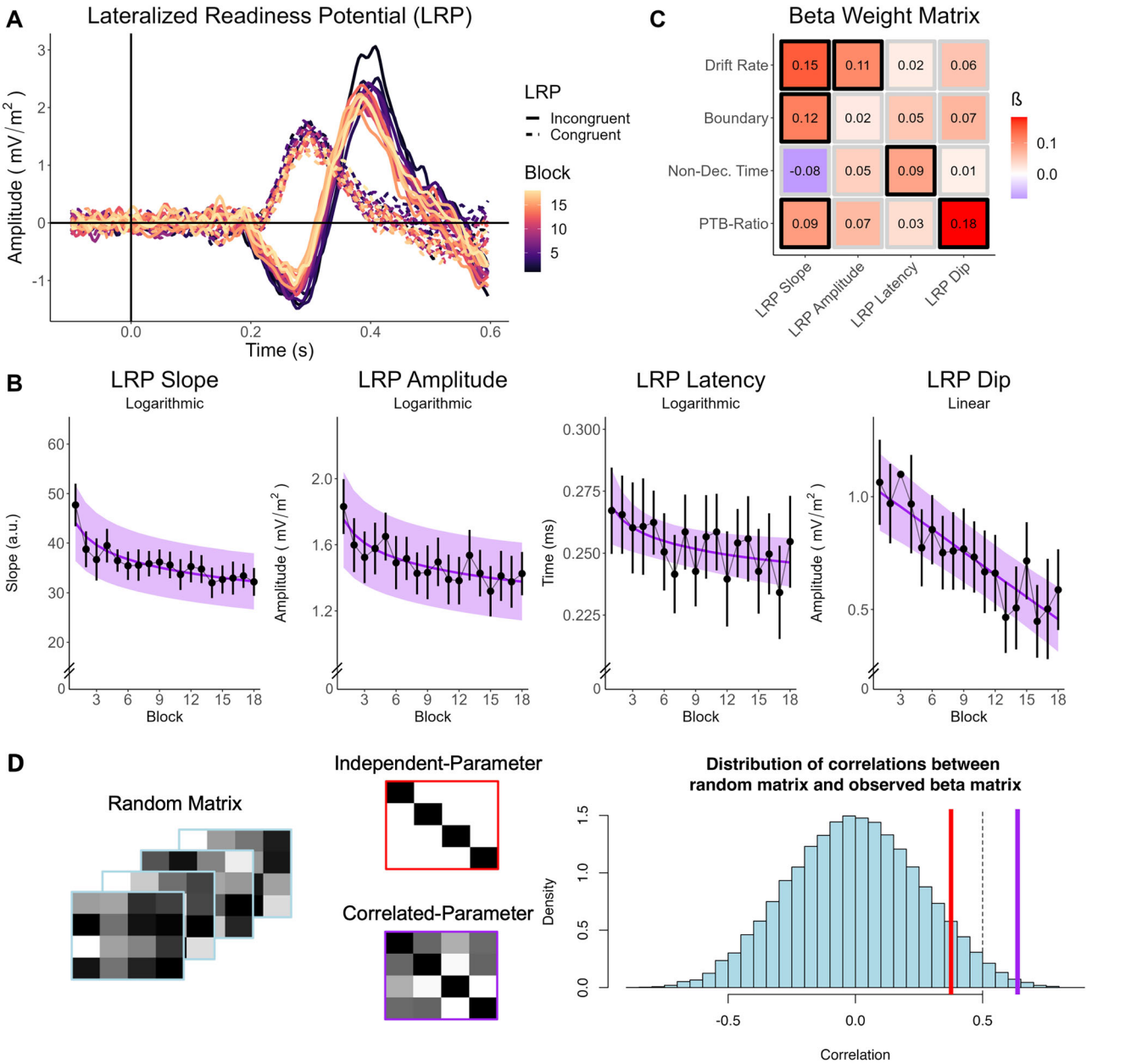

Fig. 3 | Preparatory motor signals and their relationship to the diffusion model. **A** Lateralized Readiness Potential (LRP) for congruent trials (dashed lines) and incongruent trials (solid lines), and each block. **B** Time courses of the extracted LRP features. The subtitle denotes the best model for time-on-task (linear, logarithmic, quadratic, or cubic) based on BIC. **C** The (standardized) beta weights from the multivariate mixed effect model quantifying the intra-individual relationships between local fluctuations in LRP features and local fluctuations in DDM parameters. Black boxes denote coefficients where the 95% HDI does not contain zero. **D** Examples of random model matrices used for the representational similarity analysis (blue outline), the independent-parameter model matrix (red outline), the correlated-parameter model matrix (purple outline) and the null distribution based on the random model matrices (blue) and the observed correlations for the specific (red line) and overlap (purple line) model. The dashed line refers to $P = 0.05$. PTB-Ratio peak-to-bound ratio. Error bars represent the 95% within-subject CI. Shading represents the 95% CI for the predicted effect of time. $N = 45$ participants.

time series, see "Methods") of the DDM parameters by local fluctuations of the neural indices. This model returns a beta matrix, which should not be confused with a correlation matrix, as it quantifies the hierarchically estimated within-subject relationships with the effect of the other independent variables removed and the correlation across dependent variables taken into account. As expected, we found positive relationships between LRP slope and drift rate (Fig. 3C), $\beta = 0.145$, 95% HDI [0.053, 0.231], *pd* (probability of direction, i.e., proportion of the posterior distribution that is of its median sign) > 99.99%, LRP latency and non-decision time, $\beta = 0.085$, 95% HDI [0.009, 0.159], $pd = 98.77\%$, and LRP dip and PTB-ratio, $\beta = 0.183$, 95% HDI [0.092, 0.268], $pd > 99.99\%$, showing that fluctuations in model parameters closely tracked fluctuations in the neural indices. We did not find a relationship between LRP amplitude and boundary, $\beta = 0.022$, 95% HDI [−0.068, 0.115], $pd = 67.79\%$. Yet, boundary was related to the LRP slope, $\beta = 0.116$, 95% HDI [0.017, 0.208], $pd = 99.09\%$, as was PTB-ratio, $\beta = 0.092$, 95% HDI [0.005, 0.178], $pd = 98.28\%$.

While this beta matrix shows the relationships that are present between the DDM parameters and the neural indices, it does not necessarily show whether these results adequately portray the expected overlap between the two domains. Therefore, we also adapted a model comparison approach

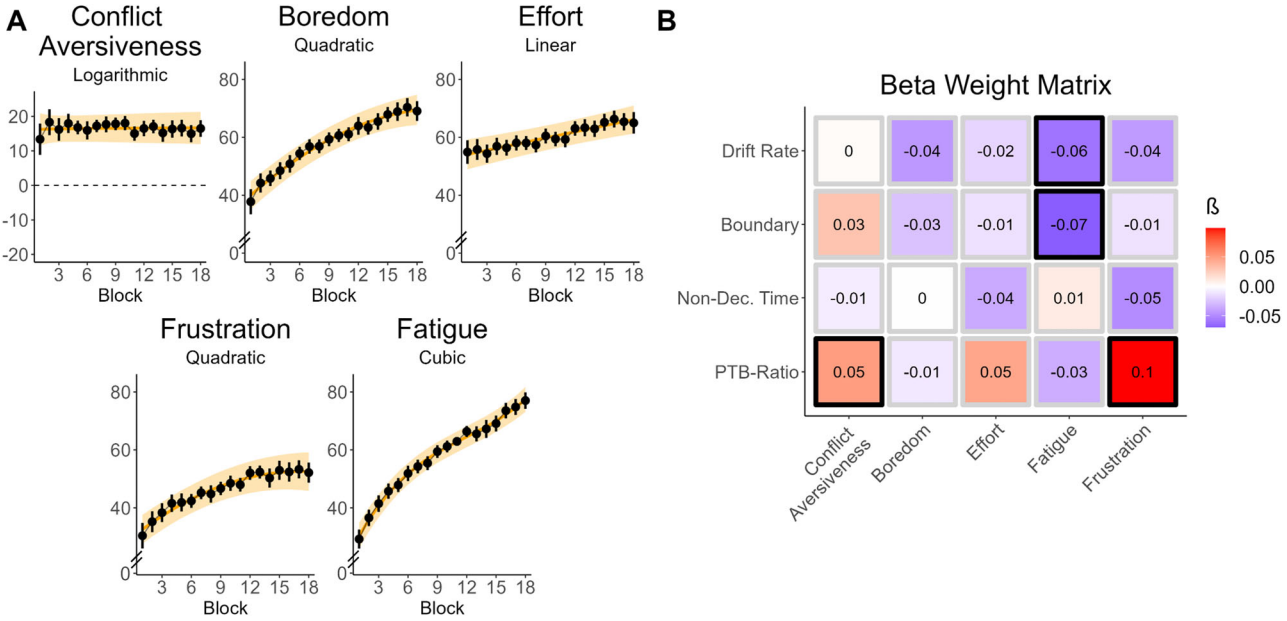

Fig. 4 | Metacognitive experiences and their relationship with DDM parameters. **A** Time-on-task effects for the metacognitive experiences. **B** Intra-individual relationships between the DDM parameters and Metacognitive Experiences. Black boxes denote coefficients where the 95% HDI does not contain zero. The subtitle denotes the best model for time-on-task (linear, logarithmic, quadratic, or cubic) based on BIC. PTB-Ratio peak-to-bound ratio. Error bars represent the 95% within-subject CI. Shading represents the 95% CI for the predicted effect of time. $N = 111$ participants.

using a representational similarity analysis logic. First, we created two theoretical models, which represent the degree of overlap one would expect between the DDM and neural domain. In a first model, the "independent-parameter model," we assume that the drift rate is specifically related to LRP Slope, decision boundary only to the LRP amplitude, non-decision time only to the LRP latency, and PTB-Ratio only to the LRP Dip. This model can thus be represented as the diagonal (Fig. 3D). However, we know that this assumption of independence is not realistic because of the existing correlations between the DDM parameters themselves, which can be expected to manifest in the neural data as well (e.g., with an established correlation between drift rate and boundary, one also expects a correlation between LRP slope and LRP amplitude). Therefore, in a second model, the "correlated-parameter model," we included the empirically observed correlations between the DDM parameters in the representation model (Fig. 3D). We evaluated the representational similarity between the observed beta matrix and our two model matrices through Spearman correlation. Our findings indicate that while the correspondence between the observed data and the independent-parameter model did not demonstrate a significant improvement over the null (random) model ($R_s(14) = 0.376$, 95% CI [−0.147, 0.730], $P = 0.152$, p value for the permutation test based on the null distribution: $P_{permutation} = 0.076$), the correlated-parameter model significantly outperformed the null model ($R_s(14) = 0.637$, 95% CI [0.206, 0.860], $P = 0.008$, $P_{permutation} = 0.005$).

Together, this indicates that there is a compelling pattern of convergence between the neural indices and DDM parameters, even though we did not find a direct relationship between LRP amplitude and decision boundary. Nevertheless, these results contribute to the investigation of the neural plausibility of this DDM for the Flanker task[23]. Analyses on the inter-individual level can be found in Supplementary Fig. S3.

Temporal dynamics of metacognitive experiences

Having established a computational basis for task performance, validated by our neural proxy, we now aimed to answer our main question: which subjective experiences provide metacognitive insight into which cognitive processes? First, we established the TOT effects for the metacognitive experiences. The average level of conflict aversiveness was found to be significantly above zero, b_0 (intercept) = 16.52, SE = 2.11, 95% CI [20.65, 12.38], $t(110) = -7.84$, $P < 0.001$, indicating that participants generally found conflict aversive, consistent with previous findings[12]. Although the best TOT model for conflict aversiveness was logarithmic (Fig. 4A), there was no statistically significant evidence for a change in conflict aversiveness with TOT, $\beta = -0.008$, SE = 0.028, 95% CI [0.004, 0.029], $t(110) = -0.20$, $P = 0.839$. However, a more detailed exploration of inter-individual differences, as presented in Supplementary Fig. S3, reveals a more nuanced picture. These analyses revealed the presence of two subgroups: one showing an increase in conflict aversiveness over time and another showing a decrease. Boredom increased in a quadratic fashion, $\beta_{linear} = 0.300$, SE = 0.025, 95% CI [0.250, 0.350], $t(110) = 11.82$, $P < 0.001$, $\beta_{quadratic} = -0.059$, SE = 0.010, 95% CI [−0.093, −0.024], $t(1775) = -5.89$, $P < 0.001$, showing an initial steady increase that slowed down towards the end of the experiment. Effort showed a small, but constant linear increase, $\beta = 0.128$, SE = 0.029, 95% CI [0.071, 0.185], $t(110) = 4.39$, $P < 0.001$. Fatigue showed a cubic pattern: the initial increase in the beginning of the experiment slowed down in the middle of the experiment but then increased again more steeply towards the end of the experiment, $\beta_{linear} = 0.361$, SE = 0.035, 95% CI [0.270, 0.451], $t(264.20) = 10.27$, $P < 0.001$, $\beta_{quadratic} = -0.074$, SE = 0.009, 95% CI [−0.106, −0.043], $t(1774) = -8.21$, $P < 0.001$, $\beta_{cubic} = 0.118$, SE = 0.023, 95% CI [0.058, 0.178], $t(1774) = 6.90$, $P < 0.001$. Frustration showed a quadratic pattern, $\beta_{linear} = 0.184$, SE = 0.027, 95% CI [0.132, 0.237], $t(110) = 13.88$, $P < 0.001$, $\beta_{quadratic} = -0.056$, SE = 0.009, 95% CI [−0.087, −0.024], $t(1775) = -6.22$, $P < 0.001$, with an initial steady increase that plateaued and eventually even seemed to reverse towards the end of the experiment. Together, these TOT effects show that the different metacognitive experiences had distinct temporal dynamics.

The relation between different metacognitive experiences and cognitive processes

Next, we quantified the relationships between metacognitive experiences and the DDM parameters, within individuals, by conducting a Bayesian Multivariate Mixed Effects Model predicting local fluctuations in DDM parameters by local fluctuations in metacognitive experiences (for the inter-individual level analyses, see Supplementary Fig. S3, for the intra-individual level analyses between the behavioral indices and metacognitive experiences, see Supplementary Fig. S4). Our analysis yielded two primary

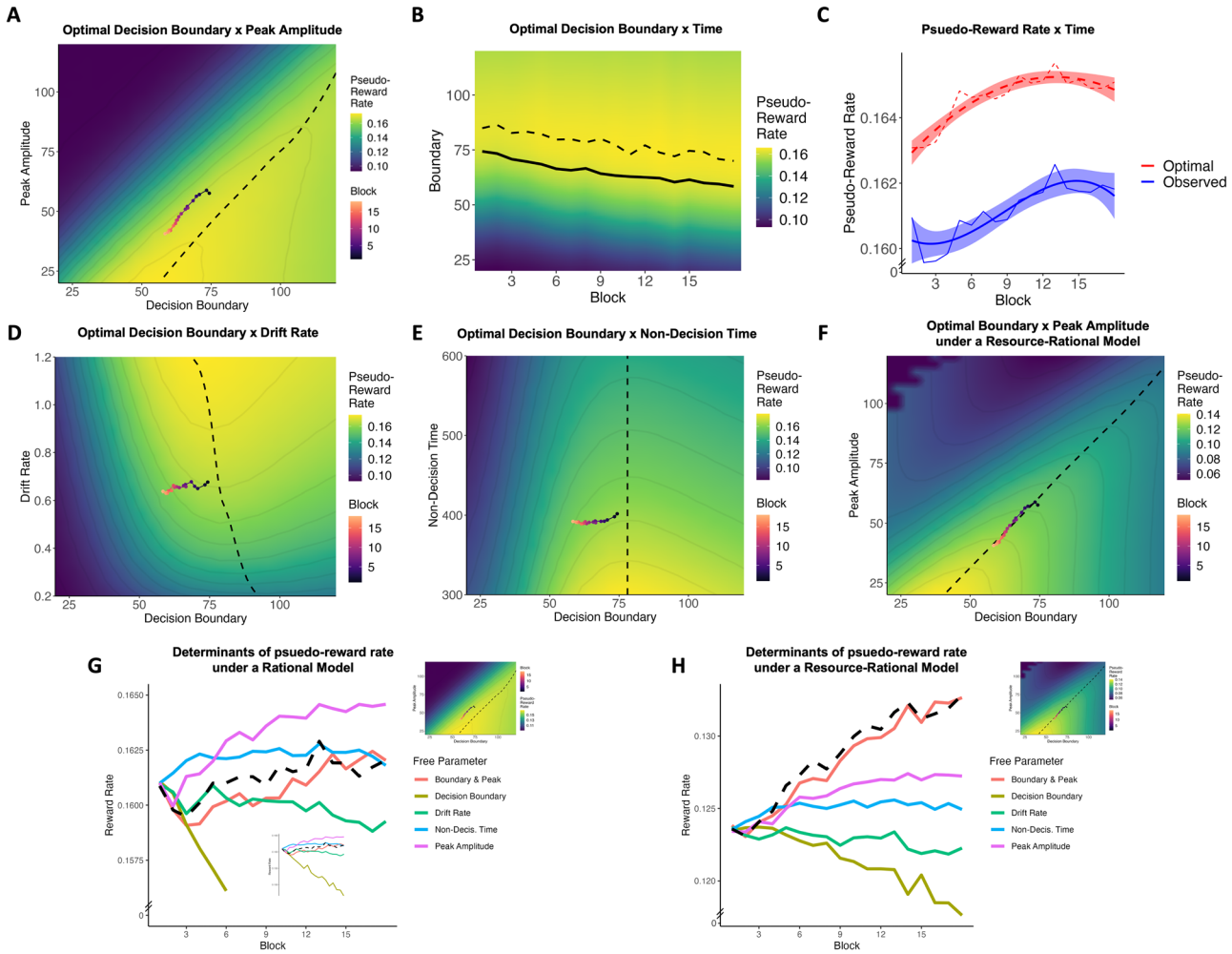

Fig. 5 | Pseudo-reward rate and the time course of DDM parameters. A Pseudo-reward rate landscape for peak amplitude and boundary, revealing the optimal setting of the decision boundary for each level of peak amplitude (dashed line). The colored line depicts the observed time-on-task trajectory of peak amplitude and boundary within this space. **B** The pseudo-reward rate landscape for the boundary across blocks—given the trajectory of the other parameters. The dashed line shows the optimal boundary trajectory, while the solid line shows the observed boundary trajectory. **C** Optimal versus observed changes in pseudo-reward rate with time-on-task. **D** Pseudo-reward rate landscape for drift rate and boundary (**E**) and non-decision time and boundary. **F** Pseudo-reward rate landscape for peak amplitude and boundary within a resource-rational framework. **G** Determinants of reward rate in the rational model. Each line shows the trajectory of pseudo-reward rate if only the free parameter was allowed to vary over the experiment, while the others are fixed to their initial value. **H** Determinants of pseudo-reward rate in the resource-rational model.

findings. First, PTB-ratio was positively related to conflict aversiveness, $\beta = 0.050$, 95% HDI [0.004, 0.092], $pd = 98.54\%$, and frustration, $\beta = 0.099$, 95% HDI [0.048, 0.149], $pd > 99.99\%$, showing that people evaluated conflict as less negative and were less frustrated when the irrelevant stimulus dimension captured their attention less (Fig. 4B). Second, drift rate displayed a negative relationship with fatigue, $\beta = -0.059$, 95% HDI [−0.108, −0.011], $pd = 99.23\%$, as did boundary, $\beta = -0.069$, 95% HDI [−0.121, −0.018], $pd = 99.58\%$: higher levels of fatigue were associated with less efficient evidence accumulation and lower levels of cautiousness, indicating more general changes in performance. Together, our findings show that frustration and conflict aversiveness may provide metacognitive insights specifically into the extent of irrelevant capture, while fatigue may offer a more general indicator of performance efficiency, along with one's level of response caution.

A rational perspective on task performance

To explore potential mechanisms underlying the observed dynamics of task performance, we turned to a rational framework. The strongest TOT effect in performance was found on the decision boundary, which strongly decreased over time. Notably, within sequential sampling frameworks, the decision boundary stands out as a distinctive parameter intimately linked to strategic control, whereas the other parameters are predominantly thought to reflect ability[41] (but see ref. 42). Importantly, setting the decision boundary, and thus strategic adjustments in task performance, directly impacts the optimality of decision-making[27,43–45]. Assuming that participants aim to optimize their performance, there usually is a single setting for the decision boundary that will maximize an objective criterion such as the reward rate[27]. Setting the decision boundary too low results in faster sampling of rewards but at a lower success rate, while setting it too high will lead to more successes at the cost of time. The optimal setting of the decision boundary, therefore, aims to maximize reward in the least amount of time.

However, in the current conflict task and model, the maximization of an objective criterion such as the (pseudo-)reward rate (in the absence of rewards, as in our task, correct responses can be considered pseudo-rewards[46]) will also depend on the strength of the irrelevant dimension, as modeled by the peak amplitude parameter (which showed the second largest TOT effect). This dependence of optimality on irrelevant capture is demonstrated by the simulations performed in Fig. 5A. This suggests that optimizing the reward rate requires setting a decision boundary relative to the strength of the irrelevant dimension. Consequently, when the sensitivity

to irrelevant capture decreases over time, as we observed, it can be considered rational to lower the decision boundary to a similar degree. Failing to adapt the decision boundary as irrelevant capture decreases would result in unnecessary time wasted. In line with this, we found a very strong positive relationship between local fluctuations in peak amplitude and local fluctuations in boundary, $\beta = 0.829$, 95% HDI [0.796, 0.862], $pd = 100\%$, showing that changes in peak amplitude were generally accompanied by similar changes in boundary.

Additional simulations on the optimal decision boundary (Fig. 5B), given the trajectory of all other estimated parameters, provided further evidence that the optimal decision boundary in this task was to lower it with TOT (dashed line), and so did our subjects (solid line). Similarly, the optimal and observed trajectory in the pseudo-reward rate landscape of boundary and peak amplitude (Fig. 5A) further confirmed that similar decreases in peak amplitude should be and are followed by similar decreases in boundary. Finally, we show that lowering the decision boundary was indeed beneficial, as it was associated to an increase in the pseudo-reward rate with TOT (Fig. 5C) for both the optimal trajectory (best model: logarithmic), $\beta = 0.903$, SE = 0.107, 95% CI [0.676, 1.130], $t(16) = 8.41$, $P < 0.001$, and the observed trajectory (best model: cubic), $\beta_{linear} = 0.818$, SE = 0.112, 95% CI [0.578, 1.058], $t(14) = 7.31$, $P < .001$; $\beta_{quadratic} = -0.222$, SE = 0.112, 95% CI [−0.462, 0.018], $t(14) = -1.98$, $P = 0.067$; $\beta_{cubic} = -0.326$, SE = 0.112, 95% CI [−0.566, −0.086], $t(14) = -2.91$, $P = 0.011$.

To test whether changes in peak amplitude might trigger adjustments in decision boundary, we conducted a Granger causality analysis to examine the temporal flow of information between these variables. We found that past values of peak amplitude from the previous block predicted the decision boundary above and beyond past values of boundary ($\bar{z} = 3.53$, $P < 0.001$). The analysis for the reverse direction (past boundary predicting current peak amplitude) did not yield a statistically significant result ($\bar{z} = 1.08$, $P = 0.281$). It is important to note that while Granger causality analysis provides insights into the temporal associations between variables, it does not establish causal relationships definitively. Parameter recovery analyses, conducted to check if the observed relationship between peak amplitude and boundary could be a modeling artifact, showed no significant correlation between these parameters in the recovered data (Pearson $R(94) = -0.040$, 95% CI [−0.238, 0.162], $P = 0.700$). We also ruled out alternative models of rationality. We show that the dynamics of the decision boundary can only be considered rational with respect to the dynamics of peak amplitude (Fig. 5A). With respect to drift rate and non-decision time, the dynamics of the boundary do not follow the optimal trajectory, even exhibiting orthogonality (Fig. 5D, E).

We further show that the observed increase in pseudo-reward rate primarily arises from the interplay between peak amplitude and decision boundary (Fig. 5G). By simulating various pseudo-reward rate trajectories, each time isolating the effect of an individual parameter while keeping the others fixed to their initial value, we found that the decrease in peak amplitude significantly boosts the reward rate, while the observed changes in drift rate and non-decision time had smaller effects. Notably, reducing the decision boundary in isolation decreases the reward rate, underscoring that its adaptive value is best understood in the context of peak amplitude. When viewed together, they show a reasonable alignment with the observed reward rate trajectory.

Interestingly, these simulations also reveal the potential suboptimality of our subjects, as they consistently fall below the optimal solution (Fig. 5C). This behavior could be interpreted as a form of resource-rational behavior, where participants demonstrate rationality to a certain extent, avoiding excessively high decision boundaries that only marginally increase the reward rate. This suggests the presence of an additional cost currently not reflected in the pseudo-reward rate. One way to address this behavior is to introduce an additional cost to the model, which we hypothesized to be equivalent to the height of the decision boundary (scaled by a weight). This effectively creates a resource-rational model where the unnecessary slowing down associated with high boundaries is avoided (cf. participants' behavior). Integrating this cost in the rational models revealed its ability to account for the perceived suboptimality (Fig. 5F). Moreover, this model further highlighted the collaborative impact of peak amplitude and decision boundary on pseudo-reward rate. Assuming that participants were optimizing within such a resource-rational framework, the observed increase in pseudo-reward rate can be primarily attributed to the joint dynamics between peak amplitude and boundary (Fig. 5H).

Together, these optimality analyses suggest that the main temporal dynamics of task performance may reflect an adaptive interplay between the ability to filter out irrelevant information (indexed by decreasing peak amplitude) and strategic adjustments (reflected in a lowering of the decision boundary) that serve to optimize the pseudo-reward rate.

Metacognitive experiences track distinct rational adjustments in task performance

In a final set of exploratory analyses, we investigated how the various metacognitive experiences are related to the main performance optimization dynamics identified in the previous section. To investigate this question, we conducted multilevel path analyses using Bayesian Multivariate Linear Mixed Effects Modeling, where we tested whether the relationship between peak amplitude and boundary is mediated by different metacognitive experiences.

In a first path model, we performed parallel mediation to test whether any of the metacognitive experiences mediated the relationship between peak amplitude and boundary. Here, we found that peak amplitude predicted conflict aversiveness, $\beta_{direct} = 0.052$, 95% HDI [0.007, 0.096], $pd = 100\%$, while effort and frustration predicted the decision boundary ($\beta_{direct} = -0.030$, 95% HDI [−0.058, −0.001], $pd = 97.7\%$; $\beta_{direct} = -0.052$, 95% HDI [−0.082, −0.021], $pd = 100\%$). However, we did not find evidence for a mediator. Given the unexpected outcome that fatigue did not predict boundary (in contrast to our prior findings), we constructed a second model to investigate whether boundary predicted fatigue instead. Our analysis revealed that boundary did indeed predict fatigue, $\beta_{direct} = -0.082$, 95% HDI [−0.132, −0.031], $pd = 99.9\%$. In a third model, we introduced serial mediation to investigate whether peak amplitude's prediction of conflict aversiveness, along with effort and frustration's prediction of boundary, created an indirect pathway through these variables. Although we did not find evidence for an indirect path, we did identify a relationship between conflict aversiveness and effort, $\beta_{direct} = 0.064$, 95% HDI [0.011, 0.115], $pd = 99.2\%$. This led us to explore a fourth model, incorporating serial mediation from conflict aversiveness through effort and frustration. In this model, we discovered a significant indirect effect over these three metacognitive experiences, $\beta_{indirect} < -0.001$, 95% HDI [$< -0.001$, < -0.001], $pd = 97.97\%$. In a fifth and final model, we examined whether there was a relationship between frustration and fatigue and whether this relationship was mediated by the decision boundary. We observed a direct relationship between frustration and fatigue, $\beta_{direct} = 0.257$, 95% HDI [0.189, 0.326], $pd = 100\%$, and we found that boundary mediates this relationship, $\beta_{indirect} = 0.003$, 95% HDI [0.001, 0.008], $pd = 99.70\%$. Leave-one-out cross-validation revealed that the final model (Model 5) outperformed the other models in terms of expected log pointwise predictive density (ELPD), which serves as a measure of predictive accuracy or fit. The differences in ELPD (and SE of the differences) between Model 5 and the other models were as follows: Model 1: −142.3 (22.4), Model 2: −174.9 (24.9), Model 3: −136.1 (19.8), Model 4: −108.0 (16.0). A visual representation of the final path model is depicted in Fig. 6, and the other path models can be found in Supplementary Note 2 and Supplementary Fig. S5.

Discussion

The current work investigated how metacognitive experiences dynamically track cognitive processes underlying task performance. To this end, we employed a TOT design during which participants performed a conflict Flanker task for 2 h, where irrelevant information had to be ignored. This task was structured in short blocks, and after each block, we probed participants' metacognitive experiences of conflict aversiveness, boredom, effort, fatigue, and frustration.

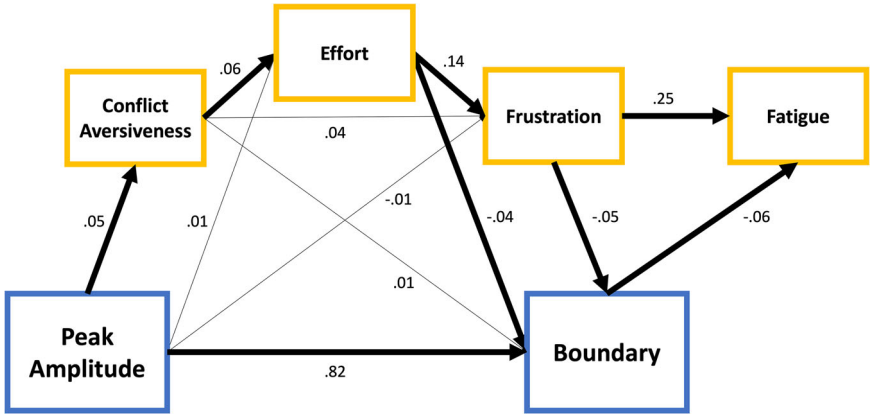

Fig. 6 | Path model. Significant relationships (for which the 95% CI does not include zero) have dark arrows. Blue relates to task performance parameters, yellow to metacognitive experiences.

We showed that the temporal dynamics of task performance were dominated by two shifts: a reduction in irrelevant capture and a decrease in decision boundary as TOT progressed. Next, we explored whether the model-based dynamics were reflected in the LRP and found significant overlap. Subsequently, we found that the main TOT dynamics in performance were related to each other and appeared consistent with a rational attempt to optimize the (pseudo-)reward rate. Metacognitive experiences were associated with those adaptive shifts in task performance: conflict aversiveness was related to changes in irrelevant capture, while frustration and effort were associated with adjustments to the decision boundary, which appeared to be aimed at optimizing reward rate. The degree to which participants adjusted their boundary based on the influence of the irrelevant dimension appeared to be partly driven by the reported levels of effort invested in the task. Feelings of fatigue may arise from this ongoing, effortful process of constant adaptation. Collectively, these findings highlight the links between metacognitive experiences and the regulation of cognitive processes, suggesting potential pathways through which these experiences might be related to adaptive behavior by reflecting (and potentially influencing) the optimality of decision-making. However, further research is needed to establish the causal nature of these relationships.

One of our main findings was that conflict aversiveness seemed to track the objective level of interference due to conflict ("irrelevant capture") both at the intra- and inter-individual level (see Supplementary Fig. S3): higher levels of conflict aversiveness were associated with higher levels of model- and neural-based irrelevant capture and vice versa. This finding is consistent with the notion that the aversive experience of conflict mirrors the cost of exerting control, and these costs are proportional to the intensity of required control. In other words, higher control intensities incur greater costs[9–12,15,16]. This aligns with previous empirical investigations into the subjective nature of control processes, demonstrating relationships between objective performance costs due to conflict, subjective estimates of this cost, and subsequent avoidance behavior[47–50]. However, some studies have failed to establish such connections (e.g., refs. 51,52), potentially because they examined interference effects in RT and error rates separately, without employing a comprehensive model-based estimate of interference that accounts for the positioning of one's decision boundary (see below for more on the importance of incorporating measures of decision boundary for the computation of irrelevant capture).

Moreover, combining RT and accuracy naturally lends itself to a normative interpretation rooted in reinforcement learning and neuroeconomic frameworks[15,53]. That is, the degree of irrelevant capture can be conceptualized as the loss in reward rate due to conflict. In essence, when irrelevant capture is substantial, it imposes a significant cost on the achievable reward rate, as incongruent trials carry both a time penalty and an increased risk of failing to attain the (pseudo-)reward. This perspective aligns with reinforcement learning theories, which have posited that the aversive evaluation of conflict can be inferred from task performance based on the lowered reward expectancy for incongruent compared to congruent trials[54,55]. Furthermore, our finding that individuals started to perceive conflict more positively as irrelevant capture diminished over time aligns with the idea that successfully resolving conflict can be experienced as rewarding, or at least more positive, and this is likely related to changes in the underlying reward expectancies[55–62].

Having accurate subjective estimates of irrelevant capture probably plays a crucial role in guiding behavioral adjustments. In fact, some have argued that such subjective estimates may hold even greater importance in driving adaptations than the objective level of interference alone[9,63,64]. As demonstrated by our simulations, when levels of irrelevant capture shift over time, it is desirable to adapt the decision boundary proportionally in order to maintain the reward rate at a consistent level. However, our results indicate the subjective estimates of irrelevant capture are not directly associated with strategic adaptations in our task, aligning with mixed findings regarding the role of subjective estimates in guiding trial-by-trial adaptations[65–67]. Instead, we found that the metacognitive experiences of effort, frustration, and fatigue were more closely related to strategic changes in the decision boundary. Specifically, our findings suggest that while effort and frustration might signal the need for adjustments in the decision boundary, the actual adjustments might promote feelings of fatigue.

The results regarding frustration and fatigue are consistent with prior research, indicating that both increased levels of frustration and fatigue are associated with a reduction in the decision boundary, leading to "more reckless" responses[14,68]. For example, Lin et al.[14] found that an ego depletion manipulation resulted in lower decision boundaries, which were linked to more pronounced feelings of fatigue. While our findings are in line with these results, we propose that the direction of this association might vary depending on the specific task and context. From a normative perspective, the optimal decision boundary is determined by factors such as the reward structure and the level of irrelevant capture. Therefore, when these factors change, adjustments to the decision boundary, either upwards or downwards, may be necessary to maintain optimal performance. An interesting avenue for further research would therefore be to investigate whether any substantial deviation from the optimal boundary would be associated with increased feelings of fatigue and frustration. This could involve manipulating the optimal boundary experimentally through the reward structure and measuring the resulting changes in metacognitive experiences.

We propose that fatigue, in this context, may arise as a consequence of the effortful process of cognitive adaptation. Although participants appear successful at dealing with the task at hand (they were able to increase the pseudo-reward rate), their consistent high effort ratings suggest that this improvement did not necessarily make the task subjectively easier. Rather, their enhanced performance appears to be linked to an ongoing, effortful process of adaptation, requiring continuous monitoring and adjustments to maintain optimality. The observation that participants began to exhibit performance decrements towards the end of the experiment, indicated by a decrease in pseudo-reward rate, further suggests they may have been approaching their capacity limit.

When viewed through a normative lens, our participants' performance appeared suboptimal. Intriguingly, while previous research suggests that humans often err on the side of over-cautiousness (with decision boundary above the optimal boundary, e.g., refs. 27,69,70), our simulations indicated that participants leaned towards an overly urgent strategy (with decision boundary below the optimal boundary[71]). This tendency to minimize time investment, at the cost of reduced reward, is consistent with resource-rational frameworks[16,72–77]. In these frameworks, the perceived cost of time associated with allocating cognitive control resources, which we modeled as a cost associated with conservative decision policies (high decision boundary), can lead individuals to accept a lower level of accuracy to conserve effort, leading to a tradeoff between effort and accuracy. This tradeoff phenomenon has been observed beyond our study in various cognitive domains such as visual cognition[78] and language[79]. Although other factors might contribute to this pattern, such as individual differences in risk aversion or miscalibration of the objective criterion, the resource-rational perspective offers a plausible explanation for the observed suboptimality.

Our investigation also sheds light on a crucial, yet often overlooked, aspect concerning the measurement of conflict. Our findings highlight that the extent of irrelevant capture not only depends on the strength of evidence from the irrelevant dimension but also on the height of one's decision boundary. This means that researchers may have inadvertently confounded measures of irrelevant capture with variability in the level of cautiousness, potentially explaining the low correlations of irrelevant capture measures in individual difference and reliability studies across various conflict tasks[80–82]. Therefore, we emphasize the importance of using measures that account for both the impact of the irrelevant dimension and express this impact relative to the height of one's decision boundary, such as the PTB-ratio, or examine the interplay between peak amplitude and decision boundary in greater detail. For instance, simply correlating peak amplitude parameter values from multiple conflict tasks to assess a general sensitivity to irrelevant information, disregarding the height of the decision boundary could lead to misleading results, as the decision boundary is likely to vary between tasks, and it co-determines the final impact of the peak amplitude parameter on decision formation.

Our findings show a relationship between peak amplitude and decision boundary, raising the question of whether changes in peak amplitude precede adjustments in the decision boundary. While our data do not allow for definitive causal claims, there are theoretical and empirical reasons to consider the possibility that people adapt their decision boundary in response to changes in irrelevant capture. First, the decision boundary is typically considered a strategic parameter that can be manipulated via instructions[83] and in response to conflict[42,84–89]. In contrast, measures for the impact of irrelevant information (which are here quantified by peak amplitude) are often utilized to measure individual differences in inhibitory capacity or executive functioning, implying a closer association with relatively stable abilities rather than dynamic strategic adjustments[82,90–93]. Second, within a sequential sampling framework, the concept of optimality typically applies solely to the decision boundary[27]. When evaluating an objective criterion like reward rate, an optimal decision boundary can be identified with respect to the other parameters, whereas there is not an optimal drift rate, peak amplitude, or non-decision time with respect to the other parameters: higher drift rate and lower peak amplitude and non-decision time are always more advantageous. Therefore, while further research is needed to definitively establish the directionality of this relationship, these theoretical considerations, combined with results from our Granger causality analyses, suggest that the temporal precedence of peak amplitude changes is a plausible hypothesis.

Limitations

Certain limitations of the current study should be acknowledged. First, the correlational nature of our findings, particularly the path analyses, precludes definitive causal inferences regarding the link between metacognitive experiences and model parameters. Thus, experimental manipulations are needed to verify the proposed pathways. Second, while our interpretation is grounded in normative, (resource-)rational, and evidence accumulation frameworks, we acknowledge that this represents one specific theoretical perspective. Alternative models, potentially incorporating factors beyond optimal decision-making and standard evidence accumulation processes, might turn out to provide a more comprehensive account of the observed behavior. Third, while the overall correspondence between model parameters and LRP indices was significant, the decision boundary parameter, specifically, lacked a direct correlate in the LRP amplitude, suggesting that this abstract model parameter may be represented in a more distributed or complex manner in the brain, warranting further investigation.

Conclusions

To conclude, our findings suggest that metacognitive experiences are related to adaptive adjustments in cognitive control and highlight potential pathways through which metacognitive experiences may trigger rational cognitive adaptations. These findings advance our understanding of adaptive decision-making and offer valuable insights for developing computational models of metacognitive experiences.

Data availability

The data to replicate the results are available in the following repository: https://osf.io/c4d7q/.

Code availability

The code to replicate the results are available in the following repository: https://osf.io/c4d7q/.

Received: 19 January 2025; Accepted: 20 June 2025;
Published online: 03 July 2025

References

1. Efklides, A. The Systemic Nature of Metacognitive Experiences. In *Metacognition*. (eds Chambres, P., Izaute, M. & Marescaux, P. J.) https://doi.org/10.1007/978-1-4615-1099-4_2 (Springer, Boston, MA, 2002).
2. Flavell, J. H. Metacognition and cognitive monitoring: a new area of cognitive–developmental inquiry. *Am. Psychol.* **34**, 906 (1979).
3. Fleming, S. M., Dolan, R. J. & Frith, C. D. Metacognition: computation, biology and function. *Philos. Trans. R. Soc. B Biol. Sci.* **367**, 1280–1286 (2012).
4. Efklides, A. Metacognition and affect: What can metacognitive experiences tell us about the learning process? *Educ. Res. Rev.* **1**, 3–14 (2006).
5. Efklides, A. Interactions of metacognition with motivation and affect in self-regulated learning: the MASRL model. *Educ. Psychol.* **46**, 6–25 (2011).
6. Katyal, S. & Fleming, S. M. The future of metacognition research: Balancing construct breadth with measurement rigor. *Cortex* **171**, 223–234 (2024).
7. Frith, C. D. Consciousness,(meta) cognition, and culture. *Q. J. Exp. Psychol.* **76**, 1711–1723 (2023).
8. Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S. & Cohen, J. D. Conflict monitoring and cognitive control. *Psychol. Rev.* **108**, 624 (2001).
9. Dignath, D., Eder, A. B., Steinhauser, M. & Kiesel, A. Conflict monitoring and the affective-signaling hypothesis—an integrative review. *Psychon. Bull. Rev.* **27**, 193–216 (2020).
10. Inzlicht, M., Bartholow, B. D. & Hirsh, J. B. Emotional foundations of cognitive control. *Trends Cogn. Sci.* **19**, 126–132 (2015).
11. Botvinick, M. M. Conflict monitoring and decision making: reconciling two perspectives on anterior cingulate function. *Cogn. Affect. Behav. Neurosci.* **7**, 356–366 (2007).
12. Dreisbach, G. & Fischer, R. Conflicts as aversive signals for control adaptation. *Curr. Dir. Psychol. Sci.* **24**, 255–260 (2015).

13. Saunders, B., Milyavskaya, M. & Inzlicht, M. Variation in cognitive control as emotion regulation. *Psychol. Inq.* **26**, 108–115 (2015).
14. Lin, H., Saunders, B., Friese, M., Evans, N. J. & Inzlicht, M. Strong effort manipulations reduce response caution: a preregistered reinvention of the ego-depletion paradigm. *Psychol. Sci.* **31**, 531–547 (2020).
15. Shenhav, A., Botvinick, M. M. & Cohen, J. D. The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron* **79**, 217–240 (2013).
16. Shenhav, A. et al. Toward a rational and mechanistic account of mental effort. *Annu. Rev. Neurosci.* **40**, 99–124 (2017).
17. Kurzban, R., Duckworth, A., Kable, J. W. & Myers, J. An opportunity cost model of subjective effort and task performance. *Behav. Brain Sci*. **36**, 661–679 (2013).
18. Agrawal, M., Mattar, M. G., Cohen, J. D. & Daw, N. D. The temporal dynamics of opportunity costs: a normative account of cognitive fatigue and boredom. *Psychol. Rev*. https://doi.org/10.1037/rev0000309 (2021).
19. De Leeuw, J. R. jsPsych: a JavaScript library for creating behavioral experiments in a Web browser. *Behav. Res. Methods* **47**, 1–12 (2015).
20. Peirce, J. W. PsychoPy—psychophysics software in Python. *J. Neurosci. Methods* **162**, 8–13 (2007).
21. Fischer, A. G., Nigbur, R., Klein, T. A., Danielmeier, C. & Ullsperger, M. Cortical beta power reflects decision dynamics and uncovers multiple facets of post-error adaptation. *Nat. Commun.* **9**, 1–14 (2018).
22. Servant, M., Montagnini, A. & Burle, B. Conflict tasks and the diffusion framework: Insight in model constraints based on psychological laws. *Cognit. Psychol.* **72**, 162–195 (2014).
23. Ulrich, R., Schröter, H., Leuthold, H. & Birngruber, T. Automatic and controlled stimulus processing in conflict tasks: Superimposed diffusion processes and delta functions. *Cognit. Psychol.* **78**, 148–174 (2015).
24. Mackenzie, I. G. & Dudschig, C. DMCfun: an R package for fitting Diffusion Model of Conflict (DMC) to reaction time and error rate data. *Methods Psychol.* **5**, 100074 (2021).
25. Bürkner, P.-C. brms: an R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* **80**, 1–28 (2017).
26. Makowski, D., Ben-Shachar, M. S. & Lüdecke, D. bayestestR: describing effects and their uncertainty, existence and significance within the Bayesian framework. *J. Open Source Softw.* **4**, 1541 (2019).
27. Bogacz, R., Brown, E., Moehlis, J., Holmes, P. & Cohen, J. D. The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol. Rev.* **113**, 700–765 (2006).
28. Gold, J. I. & Shadlen, M. N. Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. *Neuron* **36**, 299–308 (2002).
29. Gramfort, A. et al. MEG and EEG data analysis with MNE-Python. *Front. Neurosci*. **7**, 267 (2013).
30. Jas, M., Engemann, D. A., Bekhti, Y., Raimondo, F. & Gramfort, A. Autoreject: automated artifact rejection for MEG and EEG data. *NeuroImage* **159**, 417–429 (2017).
31. Boksem, M. A. S., Meijman, T. F. & Lorist, M. M. Effects of mental fatigue on attention: an ERP study. *Cogn. Brain Res.* **25**, 107–116 (2005).
32. Möckel, T., Beste, C. & Wascher, E. The effects of time on task in response selection—an ERP study of mental fatigue. *Sci. Rep.* **5**, 10113 (2015).
33. Arnau, S., Brümmer, T., Liegel, N. & Wascher, E. Inverse effects of time-on-task in task-related and task-unrelated theta activity. *Psychophysiology* **58**, e13805 (2021).
34. Ratcliff, R., Smith, P. L., Brown, S. D. & McKoon, G. Diffusion decision model: current issues and history. *Trends Cogn. Sci.* **20**, 260–281 (2016).
35. Gluth, S., Rieskamp, J. & Büchel, C. Classic EEG motor potentials track the emergence of value-based decisions. *Neuroimage* **79**, 394–403 (2013).
36. Lui, K. K. et al. Timing of readiness potentials reflect a decision-making process in the human brain. *Comput. Brain Behav.* **4**, 264–283 (2021).
37. Nunez, M. D., Vandekerckhove, J. & Srinivasan, R. How attention influences perceptual decision making: single-trial EEG correlates of drift-diffusion model parameters. *J. Math. Psychol.* **76**, 117–130 (2017).
38. Nunez, M. D., Fernandez, K., Srinivasan, R. & Vandekerckhove, J. A tutorial on fitting joint models of M/EEG and behavior to understand cognition. *Behav. Res. Methods.* **56**, 6020–6050 (2024).
39. O'connell, R. G., Dockree, P. M. & Kelly, S. P. A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nat. Neurosci.* **15**, 1729–1735 (2012).
40. van Ravenzwaaij, D., Provost, A. & Brown, S. D. A confirmatory approach for integrating neural and behavioral data into a single model. *J. Math. Psychol.* **76**, 131–141 (2017).
41. Forstmann, B. U., Ratcliff, R. & Wagenmakers, E.-J. Sequential sampling models in cognitive neuroscience: advantages, applications, and extensions. *Annu. Rev. Psychol.* **67**, 641 (2016).
42. Ritz, H., Leng, X. & Shenhav, A. Cognitive control as a multivariate optimization problem. *J. Cogn. Neurosci.* **34**, 569–591 (2022).
43. Bogacz, R., Hu, P. T., Holmes, P. J. & Cohen, J. D. Do humans produce the speed–accuracy trade-off that maximizes reward rate? *Q. J. Exp. Psychol.* **63**, 863–891 (2010).
44. Thura, D., Beauregard-Racine, J., Fradet, C.-W. & Cisek, P. Decision making by urgency gating: theory and experimental support. *J. Neurophysiol.* **108**, 2912–2930 (2012).
45. Starns, J. J. & Ratcliff, R. The effects of aging on the speed–accuracy compromise: boundary optimality in the diffusion model. *Psychol. Aging* **25**, 377 (2010).
46. Botvinick, M. M., Niv, Y. & Barto, A. G. Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective. *Cognition* **113**, 262–280 (2009).
47. Vermeylen, L., Braem, S. & Notebaert, W. The affective twitches of task switches: task switch cues are evaluated as negative. *Cognition* **183**, 124–130 (2019).
48. Vermeylen, L., Braem, S., Notebaert, W. & Ruitenberg, M. F. L. The subjective evaluation of task switch cues is related to voluntary task switching. *Cognition* **224**, 105063 (2022).
49. Van Dessel, P., Liefooghe, B. & De Houwer, J. The instructed task-switch evaluation effect: Is the instruction to switch tasks sufficient to dislike task switch cues? *J. Cogn*. **3**, 1 (2020).
50. Fröber, K., Stürmer, B., Frömer, R. & Dreisbach, G. The role of affective evaluation in conflict adaptation: an LRP study. *Brain Cogn.* **116**, 9–16 (2017).
51. Dreisbach, G. & Fischer, R. Conflicts as aversive signals. *Brain Cogn.* **78**, 94–98 (2012).
52. Fritz, J. & Dreisbach, G. Conflicts as aversive signals: conflict priming increases negative judgments for neutral stimuli. *Cogn. Affect. Behav. Neurosci.* **13**, 311–317 (2013).
53. Abrahamse, E., Braem, S., Notebaert, W. & Verguts, T. Grounding cognitive control in associative learning. *Psychol. Bull.* **142**, 693 (2016).
54. Alexander, W. H. & Brown, J. W. Medial prefrontal cortex as an action-outcome predictor. *Nat. Neurosci.* **14**, 1338–1344 (2011).
55. Silvetti, M., Seurinck, R. & Verguts, T. Value and prediction error in medial frontal cortex: integrating the single-unit and systems levels of analysis. *Front. Hum. Neurosci.* **5**, 75 (2011).
56. Braem, S., Coenen, E., Bombeke, K., Van Bochove, M. E. & Notebaert, W. Open your eyes for prediction errors. *Cogn. Affect. Behav. Neurosci.* **15**, 374–380 (2015).

57. Inzlicht, M., Shenhav, A. & Olivola, C. Y. The effort paradox: effort is both costly and valued. *Trends Cogn. Sci.* **22**, 337–349 (2018).
58. Ivanchei, I. et al. A different kind of pain: affective valence of errors and incongruence. *Cogn. Emot.* **33**, 1051–1058 (2019).
59. Ivanchei, I. I., Braem, S., Vermeylen, L. & Notebaert, W. Correct responses alleviate the negative evaluation of conflict. *Q. J. Exp. Psychol.* **74**, 1083–1095 (2021).
60. Schouppe, N. et al. No pain, no gain: the affective valence of congruency conditions changes following a successful response. *Cogn. Affect. Behav. Neurosci.* **15**, 251–261 (2015).
61. Sayalı, C., Heling, E. & Cools, R. Learning progress mediates the link between cognitive effort and task engagement. *Cognition* **236**, 105418 (2023).
62. Sayalı, C. & Badre, D. Neural systems underlying the learning of cognitive effort costs. *Cogn. Affect. Behav. Neurosci.* **21**, 698–716 (2021).
63. Desender, K., Van Opstal, F. & Van den Bussche, E. Feeling the conflict: the crucial role of conflict experience in adaptation. *Psychol. Sci.* **25**, 675–683 (2014).
64. Corlazzoli, G., Desender, K. & Gevers, W. Feeling and deciding: subjective experiences rather than objective factors drive the decision to invest cognitive control. *Cognition* **240**, 105587 (2023).
65. Schuch, S. & Pütz, S. Mood state and conflict adaptation: an update and a diffusion model analysis. *Psychol. Res.* **85**, 322–344 (2021).
66. Zeng, Q. et al. Enhanced conflict-driven cognitive control by emotional arousal, not by valence. *Cogn. Emot.* **31**, 1083–1096 (2017).
67. Ivanchei, I. I., Braem, S., Vermeylen, L. & Notebaert, W. Evaluative conditioning of conflict aversiveness and its effects on adaptive control. *Motiv. Emot.* **48**, 832–844 (2024).
68. Saunders, B., Milyavskaya, M. & Inzlicht, M. What does cognitive control feel like? Effective and ineffective cognitive control is associated with divergent phenomenology: What does control feel like? *Psychophysiology* **52**, 1205–1217 (2015).
69. Holmes, P. & Cohen, J. D. Optimality and some of its discontents: successes and shortcomings of existing models for binary decisions. *Top. Cogn. Sci.* **6**, 258–278 (2014).
70. Starns, J. J. & Ratcliff, R. Age-related differences in diffusion model boundary optimality with both trial-limited and time-limited tasks. *Psychon. Bull. Rev.* **19**, 139–145 (2012).
71. Evans, N. J., Bennett, A. J. & Brown, S. D. Optimal or not; depends on the task. *Psychon. Bull. Rev.* **26**, 1027–1034 (2019).
72. Bhui, R., Lai, L. & Gershman, S. J. Resource-rational decision making. *Curr. Opin. Behav. Sci.* **41**, 15–21 (2021).
73. Boureau, Y.-L., Sokol-Hessner, P. & Daw, N. D. Deciding how to decide: self-control and meta-decision making. *Trends Cogn. Sci.* **19**, 700–710 (2015).
74. Gershman, S. J., Horvitz, E. J. & Tenenbaum, J. B. Computational rationality: a converging paradigm for intelligence in brains, minds, and machines. *Science* **349**, 273–278 (2015).
75. Griffiths, T. L., Lieder, F. & Goodman, N. D. Rational use of cognitive resources: levels of analysis between the computational and the algorithmic. *Top. Cogn. Sci.* **7**, 217–229 (2015).
76. Griffiths, T. L. Understanding human intelligence through human limitations. *Trends Cogn. Sci.* **24**, 873–883 (2020).
77. Lieder, F. & Griffiths, T. L. Resource-rational analysis: understanding human cognition as the optimal use of limited computational resources. *Behav. Brain Sci*. **43**, e1 (2020).
78. Regier, T., Kay, P. & Khetarpal, N. Color naming reflects optimal partitions of color space. *Proc. Natl. Acad. Sci. USA* **104**, 1436–1441 (2007).
79. Mahowald, K., Fedorenko, E., Piantadosi, S. T. & Gibson, E. Info/information theory: speakers choose shorter words in predictive contexts. *Cognition* **126**, 313–318 (2013).
80. Whitehead, P. S., Brewer, G. A. & Blais, C. Are cognitive control processes reliable? *J. Exp. Psychol. Learn. Mem. Cogn.* **45**, 765 (2019).
81. Schuch, S., Philipp, A. M., Maulitz, L. & Koch, I. On the reliability of behavioral measures of cognitive control: retest reliability of task-inhibition effect, task-preparation effect, Stroop-like interference, and conflict adaptation effect. *Psychol. Res*. **86**, 1–27 (2021).
82. Hedge, C., Powell, G., Bompas, A. & Sumner, P. Strategy and processing speed eclipse individual differences in control ability in conflict tasks. *J. Exp. Psychol. Learn. Mem. Cogn*. **48**, 1448–1469 (2021).
83. Katsimpokis, D., Hawkins, G. E. & Van Maanen, L. Not all speed-accuracy trade-off manipulations have the same psychological effect. *Comput. Brain Behav*. **3**, 252–268 (2020).
84. Held, L. K. et al. Reinforcement learning of adaptive control strategies. *Commun. psychol.* **2**, 8 (2024).
85. Cavanagh, J. F. et al. Subthalamic nucleus stimulation reverses mediofrontal influence over decision threshold. *Nat. Neurosci.* **14**, 1462–1467 (2011).
86. Frömer, R. & Shenhav, A. Filling the gaps: cognitive control as a critical lens for understanding mechanisms of value-based decision-making. *Neurosci. Biobehav. Rev.* **134**, 104483 (2022).
87. Danielmeier, C. & Ullsperger, M. Post-error adjustments. *Front. Psychol*. **2**, 233 (2011).
88. Grahek, I., Leng, X., Musslick, S. & Shenhav, A. Control adjustment costs limit goal flexibility: empirical evidence and a theoretical account. Preprint at *bioRxiv* https://www.biorxiv.org/content/10.1101/2023.08.22.554296v1 (2023).
89. Fontanesi, L., Palminteri, S. & Lebreton, M. Decomposing the effects of context valence and feedback information on speed and accuracy during reinforcement learning: a meta-analytical approach using diffusion decision modeling. *Cogn. Affect. Behav. Neurosci.* **19**, 490–502 (2019).
90. Friedman, N. P. & Miyake, A. The relations among inhibition and interference control functions: a latent-variable analysis. *J. Exp. Psychol. Gen.* **133**, 101 (2004).
91. Schmidt, J. R. Is conflict adaptation adaptive? An introduction to conflict monitoring theory and the ecological problems it faces. *Q. J. Exp. Psychol*. https://doi.org/10.1177/17470218231161555 (2023).
92. Jiménez, L., Méndez, C., Abrahamse, E. & Braem, S. It is harder than you think: on the boundary conditions of exploiting congruency cues. *J. Exp. Psychol. Learn. Mem. Cogn.* **47**, 1686 (2021).
93. Jiménez, L. & Méndez, A. It is not what you expect: dissociating conflict adaptation from expectancies in a Stroop task. *J. Exp. Psychol. Hum. Percept. Perform.* **39**, 271 (2013).

Acknowledgements

We would like to thank Amitai Shenhav and Alan Voodla for valuable comments on a previous draft of the manuscript. W.N., S.B. (G.0660.17N) and L.V. (11H5619N and 1242924N) were supported by the FWO—Research Foundation Flanders. S.B. was supported by an ERC Starting grant (European Union's Horizon 2020 research and innovation program, Grant agreement 852570). C.G.G. was supported by the Special Research Fund of Ghent University (BOF.GOA.2017.0002.03) and M.R. by the Spanish Ministry of Science and Innovation (PID2019-111187GB-I00). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

L.V.: conceptualization, methodology, data collection, formal analysis, writing—original draft. S.B.: conceptualization, methodology, supervision, writing—review and editing. K.D.: conceptualization, methodology, writing—review and editing. I.I.: conceptualization, methodology, writing—review and editing. J.M.G: data collection, writing—review and editing. C.G.G: supervision, resources—equipment, writing—review and editing. M.R.:

supervision, resources—equipment, writing—review and editing. W.N: conceptualization, methodology, supervision, writing—review and editing.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44271-025-00282-x.

Correspondence and requests for materials should be addressed to Luc Vermeylen.

Peer review information *Communications Psychology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: Marike Schiffer. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025