Title:

Expected Costs of Mental Efforts are Updated When People Exert Effort, not by Prospective Information

Abbreviated Title:

Mental Effort Cost Prediction Errors are Signaled Following Effort Exertion, not Effort Cues

Authors:

Asako Mitsuto Nagase*[1,2,3,4,5], Andrew Westbrook[6], Keiichi Onoda[4,7], Kenji Morita[3,5,8], Toshikazu Kawagoe[4,9], Shuhei Yamaguchi[4], Rei Akaishi[10], and Ritsuko Hanajima[1]

[1] Division of Neurology, Department of Brain and Neurosciences, Faculty of Medicine, Tottori University, 86-1 Nishi-Cho, Yonago 683-8504, Japan

[2] Research Fellowship for Young Scientists, Japan Society for the Promotion of Science, Tokyo 102-0083, Japan

[3] Physical and Health Education, Graduate School of Education, The University of Tokyo, Hongo 7-3-1, Bunkyo-Ku, Tokyo 113-0033, Japan

[4] Department of Neurology, Faculty of Medicine, Shimane University, Enya-Cho 89-1, Izumo 693-8501, Japan

[5] Department of Cognitive Neuroscience, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-Ku, Tokyo 113-0033, Japan

[6] Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, RI, USA

[7] Faculty of Psychology, Otemon Gakuin University, Osaka, Japan

[8] International Research Center for Neurointelligence (WPI-IRCN), The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

[9] College of Contemporary Psychology, Rikkyo University, Saitama, Japan

[10] Center for Brain Science, RIKEN, Wako, Japan

*Correspondence: asako3210+academic@gmail.com

Keywords:

mental effort, fMRI, prediction error, frontal cortex, striatum, uncertainty, decision-making and learning, temporal-difference learning, reinforcement learning, model-based fMRI

Summary

To understand human behaviour, it is crucial to reveal the mechanisms by which we learn and decide about effort costs and benefits in an uncertain world. Whereas the mechanisms for reward learning are well-understood, the mechanisms for effort cost learning, and especially mental effort, remain elusive. Initially, we hypothesized that cost learning follows temporal-difference learning such that brains update expected costs when informed about upcoming effort levels. However, fMRI data revealed neural correlates of a cost-prediction error of mental effort during task execution and not in response to an effort cue about upcoming effort demands. These results imply that expected costs are updated during actual effort exertion, implying that the adaptive learning of mental effort cost does not follow traditional temporal-difference learning. This study contributes to the understanding of the neural mechanism underlying unwillingness to work, suggesting that humans learn mental effort costs only when they experience effort exertion.

Introduction

Humans and animals learn the effort that tasks require and avoid it or exert effort adaptively to achieve their goals. To understand our goal-directed and flexible behaviour, it is therefore crucial to reveal the mechanisms for effort-cost and benefit decision-making and learning in our uncertain world (Wallis and Rushworth, 2014). But where does the information about effort costs and benefits come from? Humans and animals are assumed to make decisions by integrating learned effort costs and benefits and evaluating net values. While the mechanisms for reward learning are well-understood and the mechanisms for effort cost-benefit evaluation have been investigated intensively (Doya, 2008; Glimcher and Fehr, 2014; Shenhav et al., 2013; Sutton and Barto, 2018; Walton and Bouret, 2019; Westbrook et al., 2021; Westbrook et al., 2020), the mechanisms for effort cost learning, and especially about mental effort costs, remain elusive. Striatal dopamine dynamics reflect a dissociation between the learning of reward values and physical effort costs (Gan et al., 2010). Although studies have consistently shown that humans avoid mental effort, few have investigated the neural mechanisms by which mental effort costs are learned (Cavanagh et al., 2014; Kool et al., 2010; Nagase et al., 2018; Sayali and Badre, 2019; Sayalı and Badre, 2021; Westbrook et al., 2013). Therefore, little is known about what mechanism underlies adaptive learning to avoid mental effort.

Theories about mental effort have emphasized the integrative functions of the dorsomedial frontal cortex (dmFC) and/or the dorsal anterior cingulate cortex (dACC) (Alexander and Brown, 2011; Holroyd and Verguts, 2021; Shenhav et al., 2013; Silvetti et al., 2018; Vassena et al., 2017). Yet, these models do not make explicit assumptions about the dynamics by which human brains learn mental effort costs. There is evidence that brains signal a cost prediction error (CPE), estimated by reinforcement learning (RL) algorithms, of mental effort demand (Nagase et al., 2018; Sayali and Badre, 2019; Sayalı and Badre, 2021). Previous evidence showed that the learning of negatively-valanced stimuli such as pain, air puff, or reward omission follows temporal-difference (TD) learning (Matsumoto and Hikosaka, 2009; Schultz et al., 1997; Seymour et al., 2004). The TD learning algorithm is that the estimated value of future rewards is updated with TD error based on the differences between estimates of value immediately before and after information about future rewards becomes available. It remains unclear, however, whether information on upcoming effort demand updates the expectation of mental effort costs based on TD reward learning-like prediction errors.

Initially, we assumed that mental effort cost learning would proceed in the same way as the reward learning, this is, TD learning. Namely, we first hypothesized that information in cues about upcoming effort demands updates the expected cost of mental effort. We predicted that CPEs would, at the time of the effort cues, correlate positively with the activity in the dmFC/dACC, and negatively with the activity in the striatum.

Our predictions were based on three lines of evidence. First, monkey electrophysiology data reveals that different neural populations of the mPFC encode the PE of rewards and punishments positively during learning, as increases in firing rates (Matsumoto et al., 2007; Rushworth and Behrens, 2008). Second, our previous functional magnetic resonance imaging (fMRI) study for humans showed that the activity in the dmFC/dACC and other regions correlates positively with CPE signals of mental effort (Nagase et al., 2018). Third, human fMRI data reveal a positive correlation of PE of physical effort with the activity in the dmFC (Hauser et al., 2017), though there is a possibility that the regions that represent PE of effort are not the same among mental and physical effort.

There is also reason to suspect that CPE correlate negatively with activity in the striatum. In terms of anatomy, parallel cortico-basal ganglia loops are a likely substrate for effort cost learning given their role in action selection via dopamine-mediated plasticity, regardless of whether the learning algorithm of effort cost is TD learning or not (Alexander et al., 1986; Reynolds and Wickens, 2002). Also, according to the hypothesis that reward PE reflects integrated utility, CPE signals should correlate negatively with striatal activity in contrast with the positive correlation for reward PE signals (McClure et al., 2003; Seymour et al., 2007). We note, in contrast, that PEs for positive punishment, such as pain and loss of financial reward, are opposite and positively correlated with striatal activity (Seymour *et al.*, 2007; Seymour *et al.*, 2004). As such, it remains unclear whether CPE signals correlate positively with dmFC/dACC activity and negatively with striatal activity at the time of the effort cues.

To test the first hypothesis, we designed two fMRI experiments of demand-selection tasks with effort cues, which let subjects predict the upcoming effort and generate CPE signals of TD error based on the differences between expected effort-cost before and after the time of the effort cues. We used mental arithmetic and spatial reasoning problems as demanding tasks. In a trial, participants chose an option among two options (Figure 1). The upcoming effort level associated with each option was high or low and the probabilities changed over time. In other words, high or low effort levels associated with two options transited probabilistically. We next presented effort cues which were 100% informative of upcoming effort levels. The first hypothesis predicted that, at the time of effort cues, brains represent CPE signals (Supplemental Figure 1A). Contrary to the first hypothesis, at the time of effort cues, CPE signals did not correlate with the activity in the dmFC/dACC positively, and striatum negatively. We speculated that the learning of mental effort cost does not follow TD learning and considered alternative hypotheses. Specifically, we next hypothesized that the update of the expected cost requires effort initiation and thus predicted the same neural activity pattern but at the time of effort initiation (Supplemental Figure 1B). Although we found that CPE signals correlated positively with the activity of the orbitofrontal cortex (OFC)/ventromedial prefrontal cortex (vmPFC)/ventral anterior cingulate cortex (vACC), this result did not confirm our second hypothesis about the anatomical location of the CPE. Our third hypothesis was that the update requires effort completion, and thus we predicted the same neural activity pattern with the first and second hypotheses but at the time of effort completion (Supplemental Figure 1C). Confirming our third hypothesis, CPE signals correlated with the activity of the dmFC/dACC and other regions positively, and the bilateral caudate nucleus (Cd) negatively, irrespective of demand types, at the time of effort completion. These results suggest that the expected cost of mental effort is updated during effort exertion, not at the time of upcoming effort cues. We conclude that the adaptive learning of mental effort cost does not follow TD learning.

Results

Summary of experimental designs

We conducted two fMRI experiments of demand-selection tasks with mental arithmetic and spatial reasoning problems ((Experiment (Exp.) 1: n=30 (males: females = 17:13), Exp. 2: n=28 (males: females = 12:16)) (Figure 1A). These two experiments had the same task structure but used different types of problems that require different types of cognitive demands. The flow of one trial of both Exp. 1 and Exp. 2 is described below. At the start of each trial, participants chose either a left or right arrow. After the arrow choice, one of two effort cues informing participants of upcoming effort demands was presented to participants, based on their choice. The effort cue was 100% associated with either low- or high-demand problems (Figure 1A). After the effort cue, a high-demand or low-demand problem was presented. The probability with which either arrow transitioned to a high- or low-effort (80% or 20%) cue changed dynamically over time (Figure 1B). Transition probabilities were reversed 5 times during a total of 180 trials. However, participants were instructed that "the probability that each type of problem is more likely to appear is set independently for the left and right arrows. These probabilities changed very slowly over time. The same situation will continue for at least 10 trials". After the effort cue, a problem was presented and participants answered it. Participants were instructed to "please solve the problem as fast and accurately as possible and redo the task if you think the answer may be wrong". Participants did not receive feedback on their response accuracy. There was no time limit either for arrow choices or answers to problems. Participants were instructed that the choices did not affect the overall duration of the session.

For both experiments, we prepared two sets of problems that required different levels of cognitive demand, i.e., low- and high-demand problems. The rule of mental arithmetic (division) problems of Exp. 1 was to divide a 5-digit number by 7 and report whether the remainder was small (<= 3) or large (>= 4). The dividend in low-demand problems (e.g., 35426) consisted of "two consecutive two-digit numbers that were multiples of 7" followed by a single one-digit number from 1 to 6, whereas the dividend in high-demand problems (e.g., 48106) did not contain any numbers that were multiples of 7. The rule of spatial reasoning (mental cube-folding) problems of Exp. 2 was to judge whether a 3D cube with three coloured faces and a simultaneously presented unfolded cube matched or not. The demand difference in spatial reasoning problems was whether the three faces shown on the 3D cube were adjacent on the unfolded cube (low-demand) or not (high-demand).

Answers of post-experimental questionnaire

Almost all of subjects answered yes to the question about whether subjects felt that the instruction is consistent with their experiences in the scanner. The 97% reported yes in Exp. 1 and 100% reported yes in Exp. 2. Only one subject of Exp. 1 answered no. These results suggest that most subjects thought that the probabilities of two problem types changed independently, though, in reality, the probability with which either arrow transitioned to each effort cue (80% or 20%) changed five times during 180 trials.

We also asked how much subjects liked to solve either one of two problems. The mean score on a 7-point scale was 1.00 ± 0.18 (mean ± SEM) in Exp. 1 and 0.93 ± 0.22 in Exp. 2 with the scale ranging from 0 (I preferred solving low-demand cued problems.) to 6 (I preferred solving high-demand cued problems). The individual differences of this avoidance scales did not correlate with individual differences of adjusted standardized residuals of a chi-square test which we use to index avoidance in either experiment ($r = 0.18$, $p = 0.34$ in Exp. 1 and $r = 0.35$, $p = 0.07$ in Exp. 2). These results suggest that most subjects preferred to solve low-demand problems but the subjective preferences and choice tendencies to avoid high-demand problems did not strongly relate..

Response time of arrow choices and both correct rate and solve time of demanding problems.

The response time for choosing an arrow cue (RT_{choice}) was 1.19 ± 0.12 sec (mean ± SEM) in Exp. 1 and 1.06 ± 0.08 in Exp. 2. The mean accuracies for high- and low-demand problems were, respectively, 0.95 ±

0.014 (mean ± SEM) and 0.98 ± 0.005 in Exp. 1, and 0.92 ± 0.014 and 0.99 ± 0.002 in Exp. 2. In both experiments, the rate of correct answers to the low-demand problems was on average higher than the rate of correct answers to the high-demand problems (paired t-test, $t(29) = -3.10$, $p = 4.3 \times 10^{-3}$, Exp. 1, $t(27) = -5.79$, $p = 3.7 \times 10^{-6}$, Exp. 2). The mean solution times for the high- and low-demand problems that participants answered correctly were, respectively, 14.72 ± 1.22 sec (mean ± SEM) and 1.91 ± 0.16 sec in Exp. 1, and 12.91 ± 0.84 sec and 4.12 ± 0.18 sec in Exp. 2. In both experiments, participants took more time for the high-demand problems (paired t-test, $t(29) = 11.02$, $p = 7.0 \times 10^{-12}$, in Exp. 1, $t(27) = 11.94$, $p = 2.8 \times 10^{-12}$, in Exp. 2). We set inclusion criteria (> 80% for all main experiment trials of both high- and low-demand problems) to ensure that participants solved problems diligently. By this criterion we excluded one and two participants in Exp. 1 and Exp. 2 for further analyses, respectively.

Adaptive learning to avoid mental effort with reflection on the experiences of two preceding trials

We assumed that participants who tried to avoid high-demand problems would stay on the same side (left or right) after they solved a low-demand problem in the previous trial, and would rather switch to the opposite side after they solved a high-demand problem. We tested this with a chi-square test (2 × 2 factors: problem type [high- or low-demand] in the previous trial × choice [same side or opposite side] in the current trial), as same as our previous study (Nagase *et al.*, 2018). We found that 27 (93.1 %, out of 29) and 24 (92.3%, out of 26) participants in Exp. 1 and Exp. 2, respectively, had a significant tendency of avoiding a high-demand problem ($p < 0.01$), indicating that most are demand avoiders. These results indicate that, reflecting on experiences one trial back, most participants learned to avoid high-demand problems in the uncertain environment where the upcoming effort level changed over time. Notably, in the new task design of the current study with the addition of effort cues (see the change I of STAR Methods), these results replicate our previous work in which most participants avoided high-demand problems (Chi-square test, $p < 0.01$) (Nagase *et al.*, 2018).

We also tested the effects of experiences in two trials back. We sorted the choice data into 4 pairs of bins (problem type [high- or low-demand] in the (k–2)-th trial × choice [same side or opposite side with (k–2)-th trial] in the (k-1)-th trial × problem type [high- or low-demand] in the (k–1)-th trial) and found that proportions of that k-th choice were the same as (k-2)-th choice (Supplemental Figure 2), replicating our previous study (Nagase *et al.*, 2018). There were some significant differences in the means: participants chose the same options with the (k–2)-th choice more frequently when the experienced demand of two trials-back was low versus high, with small to large effect sizes in Exp. 1 and 2, respectively (Supplemental Figure. 2Aa: $p = 1.7 \times 10^{-8}$, $d = -1.79$, $t(26) = -8.01$; Supplemental Figure. 2Ac: $p = 2.6 \times 10^{-2}$, $d = 0.74$, $t(22) = -2.40$; Supplemental Figure. 2Ad: $p = 4.7 \times 10^{-2}$, $d = 0.56$, $t(22) = -2.10$) and 2 pairs of bins in Exp. 2 (Supplemental Figure. 2Ba: $p = 2.6 \times 10^{-8}$, $d = -1.49$, $t(23) = -8.24$; Supplemental Figure. 2Bc: $p = 4.1 \times 10^{-4}$, $d = -1.08$, $t(22) = -4.15$). Most participants avoided high-demand problems significantly and they learned to avoid high-demand problems, reflecting their experiences of two preceding trials, adaptively under a dynamically-changing environment. These results all replicate our previous findings (Nagase *et al.*, 2018) that humans avoid mental effort based on a history of multiple-past experiences.

Model-based analysis of choice data

We performed model fitting and model comparison to discern which computational models explain the choice data when participants learned to avoid high-demand problems. Nearly the same models were used as in our prior study, though we cut off the five worse-performing models, for which the expected cost for the unchosen option on the 1st trial was an additional free parameter (Nagase *et al.*, 2018). We compared eleven models by which effort learning was influenced by permutations of solve times, incorrect rates, rest time and binarized demand levels (see Supplements and Nagase et al, 2018 for full details and winning

models described in detail in STAR methods). Bayesian Information Criterion (BIC) was used for model comparison, dividing eleven models into three groups: reinforcement learning (RL), rest-time RL, and probabilistic models. Among five models in the category of the RL models which learned each factor in a simple RL update algorithm, the RL model based on binarized effort demands was the best model in BIC scores for most or all of the participants in both experiments (25/27 (92.6%) in Exp. 1; 24/24 (100.0%) in Exp. 2; Supplemental Figure 5, Table 1). The result that the RL model outperformed the other RL models indicates that the cost of mental effort is learned as a nearly binary variable, rather than variables reflecting the solve time or errors. The RL model was better than the two rest-time RL models for all or most of the participants in BIC scores in both experiments (27/27 (100.0%) in Exp. 1; 21/24 (87.5%) in Exp. 2; Supplemental Figure 5, Supplemental Table 1). Among four models in the category of probabilistic models, the probabilistic selection model based on binarized effort demands was the best model in BIC scores for most participants (25/27 (92.6%) in Exp. 1; 23/24 (95.8%) in Exp. 2; Supplemental Figure 5, Supplemental Table 1). A direct comparison of BIC scores between the RL and probabilistic selection models, revealed that the probabilistic selection model was better than the RL model in both experiments (17/27 (63.0%) in Exp. 1; 14/24 (58.3%) in Exp. 2; Supplemental Figure 5, Supplemental Table 1). The adaptive avoidance of mental effort in an uncertain world was thus comparably explained by both probabilistic selection and reinforcement learning models based on mental effort costs although the probabilistic selection model explained the choice data better among a larger proportion of participants.

The probabilistic selection model makes a decision based on the experience of one preceding trials, while the RL model makes a decision reflecting the experiences of multiple preceding trials (see details in Behavioural Analysis of STAR Methods). We previously simulated the choice behaviour with these models and the effect of the two-preceding trials was captured by the RL, but not the probabilistic selection model, as expected (Nagase *et al.*, 2018). Simulation results mirror our previous work (Figure 8 of (Nagase *et al.*, 2018)). According to BIC values, the probabilistic selection model explained choices better than the RL model in both our previous and current studies. On the other hand, the RL model captured the sequential effect of past experienced demands better than the probabilistic selection model in simulation. Therefore, we judged that it was reasonable that we performed model-based fMRI analysis with the RL model in the current study as we did in our previous study. For comparison, we report the results of RL model fitting for all demand avoiders of Exp. 1 (Supplemental Figure 3) and Exp. 2 (Supplemental Figure 4).

Does the learning of mental effort cost follow the TD learning that updates the expected cost of future effort with TD error between temporally different estimates? To address this question, we analyzed the within-trial dynamics of blood oxygen level-dependent signals.

Null results for the initial hypothesis of cost prediction error signals at the time of effort cues

We initially hypothesized that the CPE learning algorithm is TD learning and expected costs are updated at the time of effort cues. We predicted that CPE signals correlate positively with activity in the dmFC/dACC, and negatively correlate with activity in the striatum, irrespective of demand type, at the time of effort cues (Supplemental Figure 1A). To test this hypothesis, we examined the activity of the whole brain to search for regions where the hemodynamic response at effort cue presentation was positively or negatively correlated with the CPE signals in a conjunction analysis between Exp. 1 and Exp. 2. The conjunction analysis involved a strict mask comprising common voxels from Exp. 1 and 2, with the threshold of cluster-level FWE corrected $p < 0.05$, and voxel-level uncorrected $p < 0.001$ from GLM 1 (Figure 2A). At the time of effort cues, activity in neither the dmFC/dACC nor the striatum was correlated with CPE signals significantly with the threshold of cluster-level, FWE corrected $p < 0.05$, and voxel-level, uncorrected $p < 0.001$ (see Figure 3A, Supplemental Figure 6, Supplemental Table 2) (the results of conjunction analyses will be reported with this threshold, unless otherwise stated). The conjunction analysis with the strict mask is a stringent to test of

whether there are reliable common signals, but may be overly conservative. To confirm the null results, we analyzed the conjunction using a relaxed mask that consisted of common voxels between the results of Exp. 1 and 2 with the threshold of voxel-level uncorrected $p < 0.01$ using GLM 1. Under relaxed criteria, activity in the vmPFC/vACC was positively correlated with the CPE signals at the time of effort cues (voxel-level uncorrected $p < 0.001$, see Supplemental Figure 7), but there was no super-threshold correlation with activity in the dmFC/dACC. These results thus fail to support the first hypothesis (Figure 5A).

In contrast to these null results, activity in the right intraparietal sulcus (IPS) correlates negatively with CPE signals at the time of effort cues, using the strict mask. This negative correlation may reflect the effect of the reset, which reflashes the past representation to promote cognitive shifts, in a part of the frontoparietal network (Bouret and Sara, 2005). Another interpretation is that the negative correlation in the right IPS represents the PE of subjective value. However, a fMRI meta-analysis reported that activity in the presupplementary motor area correlates positively with the domain-general effort expectation among physical and mental effort (Lopez-Gamundi et al., 2021). Indeed, CPE's appear to correlate positively with activity in the dmFC/dACC (Hauser *et al.*, 2017; Nagase *et al.*, 2018). Therefore, the location and correlation sign of the negative correlation in the right IPS is inconsistent with previous studies. Regarding the negative results, it is conceivable that participants ignored the effort cues. However, there was a significant neural response of typical task preparation in the dmFC/dACC, the right dlPFC, and the right IPS which was positively correlated with the parametric regressor of problem-demand as a result of conjunction analysis (Figure 4, Supplemental Table 3). This result indicates that most demand avoiders encoded information about upcoming effort demands at the time of effort cues, predicting the upcoming effort, and prepared for the task accordingly. Therefore, these null results imply that the adaptive mental effort cost learning does not follow traditional TD learning.

The neural correlates of cost prediction error signals at the time of effort initiation

In light of the results of GLM 1, we revised the first hypothesis and posited the second and third hypotheses that the update of expected cost requires effort initiation and completion, respectively (Supplemental Figure 1). We thus predicted that CPE correlates positively with the activity in the dmFC/dACC, and negatively with the activity in the striatum, at the time of effort initiation or completion (Supplemental Figure 1B and C). A conjunction analysis with the strict mask by using GLM 2 (Figure 2B) revealed that, at the time of problem presentation when effort is initiated, the activity in the orbitofrontal cortex (OFC)/ventromedial prefrontal cortex (vmPFC)/ventral anterior cingulate cortex (vACC) correlates positively with CPE signals (see Figure 3B, Supplemental Figure 8, Supplemental Table 4). Although we detected a positive correlation between the CPE and the activity of the large cluster including the dmFC/dACC in Exp. 1, the CPE was neither positively correlated with the activity in the dmFC/dACC in Exp. 2 nor negatively correlated with the activity in the striatum in either experiment. These results fail to support the second hypothesis (Figure 5B). However, they provide limited evidence that CPE's correlate with the OFC/vmPFC/vACC at the time of effort initiation, with the preceding information of upcoming effort levels.

The neural correlates of cost prediction error signals at the time of effort completion

We next tested the hypothesis that expected costs are updated at effort completion (Supplemental Figure 1C). A conjunction analysis with the strict mask (GLM 3; Figure 2C), revealed that activity in the dmFC/dACC, the left anterior insula (aINS), and the left primary motor cortex (M1)/ primary somatosensory cortex (S1)/IPS correlates positively, and activity in the bilateral caudate nucleus (Cd) correlates negatively with CPE signals at the time of effort completion (see Figure 3C, Supplemental Figure 9, Supplemental Table 5). Therefore, these results support the third hypothesis and suggest that the CPE signals are represented in the

dmFC/dACC and other regions positively, and bilateral Cd negatively at the time of effort completion (Figure 5C).

A potential confounding factor of CPE signals would be a confidence PE. Although prospective confidence for the upcoming problem at the time of choices and retrospective confidence for the solved problem at the time of effort completions would fluctuate per trial, participants' confidence at the time of each answer is unknown since we did not collect the trial-by-trial data of subjective confidence after the answers. Therefore, we cannot rule out encoding of retrospective confidence PE's as a confound. However, the confidence-related activation is usually detected in the frontopolar cortex, pregenual ACC, or ventrolateral PFC, not in the Cd (Bang and Fleming, 2018; Miyamoto et al., 2017; Miyamoto et al., 2021a). Also, it is hard to interpret that participants made a decision based solely on prospective confident without avoiding effort costs because current experimental designs did not give accuracy feedback, or performance-contingent reward or punishment. Another potential confounds of CPE signals is a satisfaction PE, following effort completion. Participants might feel greater satisfaction with completion of high-demand problems and learned trial-by-trial satisfaction instead of mental effort cost. However, participants tended to avoid higher demands, making it unlikely that they were driven to seek greater satisfaction. Hence, these confounding factors are unlikely.

Discussion

Summary of results

We first hypothesized that mental effort-cost learning follows TD learning and that the expected costs of mental effort are updated at the time of prospective effort cues. We therefore predicted that at the time of effort cues, CPEs correlate positively with activity in the dmFC/dACC and negatively with activity in the striatum. However, activity in neither the dmFC/dACC nor striatum correlate with CPEs at the time of effort cues, which was revealed by a conjunction analysis across Exp. 1 and 2. This result failed to support the first hypothesis. Therefore, we hypothesized that expected costs are updated either at effort initiation or completion. At effort initiation, the activity of the OFC/vmPFC/vACC correlates positively with CPEs. At effort completion the CPE correlates positively with the activity in the dmFC/dACC and other regions, and correlates negatively with activity in the bilateral Cd. These results suggest that human brains represent CPEs during effort exertion, especially at effort completion, and not at the time of effort cues. These results also suggest that expected cost updates require effort exertion and that adaptive learning of mental effort costs does not follow traditional TD learning.

Consideration of results and limitations

Behavioural data confirmed that humans tend to avoid mental effort, reflecting the experiences of multiple past events. The choice data of demand avoiders were approximated with a RL model based on mental effort levels, implying that humans update cached mental effort cost information. Questionnaire results suggested that most participants prefer low-demand problems but that retrospective subjective preferences have less relationship to behavioural choice tendencies.

Does the adaptive learning of mental effort costs follow TD learning? The activity in neither the dmFC/dACC nor striatum correlated with CPEs at the time of effort cues. This lack of a CPE correlation with the activity in the dmFC/dACC at the time of effort cues undermines this hypothesis (Figure 3A and 5A). The negative correlation in the right IPS activity with CPE might reflect the reset of frontoparietal network (Bouret and Sara, 2005), rather the PE of subjective values. These results imply that the expected

costs are not updated by the prospective information and that the adaptive learning of mental effort cost does not follow TD learning. These results also imply that the mental effort cost learning system dissociates phenomenologically from the reward learning system. One reason humans may not update the expected costs prior to task execution might be that humans can change their effort, on-the-fly by changing strategies to solve the task or encounter difficulties and solve tasks more slowly. It is also conceivable that people recognize that, with learning and practice, some mental tasks become less effortful than expected, and so they withhold mental effort cost updates until after effort is completed. Mismatches may be particularly significant in the case of mental effort which often involves unanticipated failures, e.g., in working memory operations, or unanticipated successes due to insight or skill-learning.

When is the expected cost of mental effort updated? The fact that activity in the OFC/vmPFC/vACC correlates positively with CPE signals when participants start exerting effort fails to support the second hypothesis (Figure 3B and 5B). Therefore, we conclude that the expected cost of mental effort is not updated at the time of effort initiation. However, we identified common activity patterns around the rostromedial prefrontal cortex with two experiments of our previous study. This means that these activity patterns were robust and the common clusters across four experiments were in the rostral gyrus (Supplemental Figure 10). These results stand in contrast with the general perspective in the field of value-based decision-making that the vmPFC positively encodes predicted reward value, discounted by mental effort cost, or reward PE signals (Hauser et al., 2015; Klein-Flügge et al., 2022; Lopez-Gamundi *et al.*, 2021; Westbrook et al., 2019). In contrast, it is hypothesized that the OFC tracks current state for elaborating a model of the external world (Sharpe et al., 2019; Wilson et al., 2014). Perhaps the mental action, triggered by effort initiation, recruits the OFC to update a model of the internal world, reflecting a transition into task-engaged states, which would differ between the states which participants solve high- and low-demand problems. Therefore, we speculate that the positive correlation of the OFC/vmPFC/vACC with CPE reflects an update to an internal model which maps mental actions and mental states, monitored by brains, and reported at the time of effort initiation. This update is important for credit assignment, because subsequent cognitive and motor) actions and outcomes should be evaluated with respect to progress towards the goal to which the effort is directed. The evaluative function may change both implicitly and even subjectively before versus after effort initiation.

So, when is the expected effort cost updated? We find that when participants complete effort exertion by answering the problems, the CPE correlates positively with the activity in the dmFC/dACC, left aINS, and left M1/S1/IPS and negatively with activity in the bilateral Cd, irrespective of demand type. These results support the third hypothesis and suggest that the update occurs at the effort completion (Figure 3C and 5C). A complementary effort completion signal may also be needed to trigger an update to the expected cost representations of the mapping between motor action and the outcome based on experienced effort. The possible learning algorithms, which update the cost expectations with the prediction errors, underlying the adaptive avoidance of mental effort may be based on Bayesian inference or non-TD reinforcement learning (Ma and Jazayeri, 2014; Sutton and Barto, 2018). While it remains unclear what the brain treats as costly, our data imply that the brain monitors neural dynamics associated with effort exertion and reports them as a CPE after exertion is completed. If an effort completion was the sub-goal of a trial for the brains, the brains might focus on solving a problem without monitoring. After achieving the goal, the brain might evaluate the objective or subjective effort costs.

Relation to previous studies

There is evidence of utility representations integrating reward value discounted by mental effort cost in the ventral striatum (VS), the vmPFC, the amygdala, and the dmFC/dACC (Botvinick et al., 2009; Westbrook *et al.*, 2019). The utility PE hypothesis of TD learning was posited to cover both reward and lack of reward,

thereby expanding the concept of TD learning to more general integrated subjective valuation (Schultz et al., 2017). Further, TD learning was reported as the learning algorithm of pain and loss of reward, just as for rewards, although the PE of these positive punishments is thought to be encoded with opposite signs to reward PE signals in the human striatum (Seymour *et al.*, 2007; Seymour *et al.*, 2004; Sutton and Barto, 2018). Modeling of behavioural and neural data indicate that mental effort costs respect RL algorithms (Nagase *et al.*, 2018; Sayali and Badre, 2019; Sayalı and Badre, 2021). Collectively, these prior studies imply that human brains learn mental effort cost as an integrated value in the manner of TD learning. Surprisingly, our findings revealed that learning of mental effort cost does not follow TD learning and expected costs are instead updated pursuant to actual effort exertion. These findings revise our understanding of the neural mechanisms of cost-benefit decision-making and learning, especially the learning mechanisms based on mental effort cost. We speculate that this might be because mental effort costs are uncertain until effort has been exerted and updates are thus withheld until effort has been experienced. As such, CPE updates do not conform to traditional TD learning algorithms applicable to other forms of punishments and rewards.

Our results revise the view about theories and hypotheses of mental effort and the function of the frontal cortex, especially the dmFC/dACC, and the striatum, especially the Cd. Our data further imply that the dmFC/dACC plays a critical role in mental-effort cost learning. Importantly, the negative correlation of the CPE with activity in the Cd was detected only at effort completion. Considering circuit anatomy (Alexander *et al.*, 1986), a negative correlation with activity in the Cd and a positive correlation activity in the dmFC/dACC imply a possibility that mental effort costs are learned in cortico-basal ganglia loops encompassing the dmFC/dACC and Cd.

Conclusion

Our data indicate that the learning algorithm of mental effort costs is not classical TD learning. We find that humans update expected costs by CPE when they finish exerting mental effort and not when they receive effort cues about upcoming task efforts. Thus, effort exertion appears to be required for humans to learn the costs of effortful tasks. These findings advance our understanding of the motivational mechanism underlying the human work and study in daily life. These findings revise our understanding of the mechanisms of mental effort-based decision-making and learning.

Figure legends

Figure 1. The behavioural paradigm of demand-selection tasks with effort cues.

(A) At the beginning of each trial, participants selected either a left or right arrow associated probabilistically with either low- or high-demand problems. After an effort cue, a problem was presented (high-demand or low-demand problem). Effort cues exactly indicated upcoming effort levels. Effort cues were either red or green (blue in the figure) circles in Exp. 1 and upward or downward triangles in Exp. 2, and cue types were counterbalanced across participants. Left and right arrows probabilistically determined which types of effort cues were presented. The probabilities of arrows associated with effort cues, which is as the same as upcoming effort levels, were changed dynamically over time (Figure 1B). The set of either 80 or 20% was exchanged 5 times during 180 trials. Following arrow cue selection, an unambiguous effort cue thus resolved uncertainty about upcoming effort demands. There was no time limit for response. (B) An example of the presentation rates of low-demand problems (moving average of latest 5 trials) associated with the left arrow-cue (light grey) and the right arrow-cue (dark grey).

Figure 2. GLMs and regressors in the fMRI analyses.

There were five events in a trial: presentation of arrow cues, arrow cue choice, presentation of an effort cue, problem presentation, and problem completion. We explored the neural correlates of the cost prediction error (CPE) estimated in the reinforcement learning model based on effort levels by using three general linear models (GLM1-3), according to three different hypotheses regarding the time of CPE generation/representation. All GLMs included regressors at the presentation of arrow cues with non-orthogonalized parametric modulations by $ExpectedCost_{chosen}$ and RT_{choice}, regressors at presentation of effort cues with non-orthogonalized parametric modulations by upcoming demand-level (1 and 0 for high- and low-demand problems, respectively), regressors at arrow cue choice, regressors problem presentation having the duration of solve-time with non-orthogonalized parametric modulations by demand-level (1 and 0 for high- and low-demand problems, respectively). Respectively, the GLMs also contained regressors for motor response at both arrow choice and answer (GLM1 and 2) or arrow choice (GLM3), regressors for head movements (not illustrated here), and CPE-modulated regressors at effort cue ((A) GLM1), at problem onset ((B) GLM2), or at answer time ((C) GLM3).

Figure 3. Neural correlates of the cost prediction error at the times of a presentation of a effort cue, a problem presentation, and an answer to the problem.

Conjunction analyses with a binary mask consisting of common voxels between the positive and/or negative correlations of the cost prediction error (CPE) at the time of effort cue presentation (A: GLM1 was used), problem presentation (C: GLM2 was used) or answer (D: GLM3 was used) in Exp. 1 and Exp. 2 at cluster-level FWE corrected $p < 0.05$ and voxel-level uncorrected $p < 0.001$. (A) No clusters were detected as a positive correlation. The right intraparietal sulcus (IPS) was detected as a negative correlation. (B) The orbitofrontal cortex (OFC)/vmPFC/vACC was detected as positive correlation. (C) Three clusters were detected in the dorsomedial frontal cortex/ dorsal anterior cingulate cortex (dmFC/dACC), left anterior insula (aINS), left primary motor cortex (M1)/ primary somatosensory cortex (S1)/IPS as a positive correlation. The bilateral caudate nucleus was detected as a negative correlation.

Figure 4. Neural correlates of the task preparation at the times of a presentation of a effort cue.

Conjunction analyses with a binary mask consisting of common voxels between the positive correlations of the task preparation at the time of effort cue presentation (A: GLM1 was used) in Exp. 1 and Exp. 2 at cluster-level FWE corrected $p < 0.05$ and voxel-level uncorrected $p < 0.001$. The right dorsolateral prefrontal cortex (dlPFC), right IPS, dmFC/dACC were detected as a positive correlation.

Figure 5. Hypotheses and Results.

Conjunction analyses revealed neither positive cost prediction error (CPE) correlations in the dmFC/dACC nor negative correlations in the striatum with (A) at the time of effort cues or (B) at the time of effort initiation. Neural results thus do not support the first and second hypotheses. (C) Conjunction analyses revealed positive CPE correlations in the dorsomedial frontal cortex/ dorsal anterior cingulate cortex (dmFC/dACC) and other regions (data not shown in the figure) and negative correlations in the bilateral caudate nucleus at the time of effort completion, supporting the third hypothesis.

Author Contributions

1. Conceptualization

AMN

2. Data curation

AMN

3. Formal analysis

AMN

4. Funding acquisition

SY and AMN

5. Investigation

AMN, TK, and KO

6. Methodology

AMN

7. Project administration

AMN, SY, and RH

8. Resources

AMN

9. Software

AMN

10. Supervision

AMN, SY, and RH

11. Validation

AMN, KM (the behavioural analysis with computational models), and KO (fMRI analysis)

12. Visualization

AMN

13. Writing – original draft

AMN

14. Writing – review & editing

AMN, AW (comments&edit), KM (comments), TK (comments), RA (comments), and RH (comments)

Declaration of interests

The authors declare no competing interests.

Acknowledgements

AMN was supported by Grant-in-Aid for Research Activity Start-up No. 18H06090, for JSPS Fellows No. 19J00964, for Early-Career Scientists No. 20K16475, Challenging Research (Exploratory) No. 19K21809, and for Scientific Research on Innovative Areas No. 15H05876 of JSPS. SY was supported by The Impulsing Paradigm Change through disruptive Technologies (ImPACT) programme in Japan. We are grateful to Katsuyuki Sakai, David L Aguilar, Ryan P Badman, Tetsuya Yamamoto, Kentaro Miyamoto, Mingbo Cai, Akihiro Funamizu, Taro Toyoizumi, Yosuke Morishima, Tomoki Haji, Atsushi Nagai, Takahiro Shimizu and the members of the Neurology Division of Tottori University, Akaishi Lab of RIKEN CBS, and Doya Lab of OIST for the advice, support, discussion, or feedback to this study.

STAR Methods

Participants

We recruited 30 participants (13 females; mean age, 23.0 ± 2.2) in Experiment (Exp.) 1 and 28 participants (16 females; mean age, 22.6 ± 1.6) in Exp. 2. Among these participants, 11 participants took part in both Exp. 1 and Exp. 2. We paid all participants 5,000 yen by bank transfer after their participation. No participants were taking any medicine or had a prior history of neuropsychiatric disorders. All participants were students of Shimane University, age 20-30 years, right-handed, native Japanese speakers, and received their compulsory education in Japan. Informed written consent was obtained from all participants before the experiment. The present study was approved by the medical ethics committee of Shimane University.

Behavioural tasks

We conducted two experiments, Exp. 1 and Exp. 2 (Figure 1). Except for changes I~VII as we will describe below, these experiments were almost similar behavioural protocols with our previous work (Nagase *et al.*, 2018). Consequently, we will explain all methods in the summary and list changes I~VII from our preceding procedure (please see Nagase *et al.*, 2018 for full details). The most important change was to add the effort cues that indicated the upcoming effort levels (i.e., high- or low-demand problems) between choices and problem presentations (please see the change I). There were two reasons why we added effort cues. First, if the first hypothesis was true and mental effort cost are learned via TD learning, this change should have allowed us to detect the CPE signals at the time of effort cues. This is because the PE signals generate to the earliest predictor, backpropagating from later predictors (Sutton and Barto, 2018). In our previous design, the timing of the CPE generation was not locked by experimental manipulations and we did not know when the CPE signals were generated, and thus explored the neural correlations at various timings. Second, the addition of effort cues allowed us to detect the signals of expected cost for the chosen option, reducing the confounding effect of task preparation at the time of choice-cues. At the time of choice-cues, the task preparation effects confound expected cost signals (expected effort demands) in both our previous and current designs. The effort cue reduced the necessity of the task preparation at the time of choice-cues in the current design because participants could get certain information of upcoming task type (i.e., high- or low-demand problems) at the time of effort cues, which was 2.5 seconds after the choice. We will report the results of the signals of expected cost of both Exp. 1 and Exp. 2 in another article in the near future.

We will describe the experimental paradigms in summary. The two experiments that we conducted had similar structures except for the following two points. The first point is that we used different types of problems in the two experiments. More specifically, we used mental arithmetic problems to divide a 5-digit number by 7 and report whether the remainder was small (<= 3) or large (>= 4) in Exp. 1, and spatial reasoning (mental cube-folding) problems to judge whether a 3D cube with three visible coloured faces and a concurrently presented unfolded cube matched or not in Exp. 2. The second point is that we used different features of effort cues that indicated upcoming effort levels; circle cues with red and green colors in Exp. 1 and upward and downward triangle cues in Exp. 2.

There were problems with two levels of cognitive demand, i.e., high-demand and low-demand problems in both experiments. We will describe how we made the difference in demand levels later. In Exp. 1, the dividend for the low-demand problem (e.g., 35426) was two consecutive two-digit numbers that were multiples of 7 followed by one single-digit number from 1 to 6, whereas the dividend for the high-demand problem (e.g., 48106) did not contain any multiples of 7. In Exp. 2, the difference between the low- and high-demand problems was whether or not the three faces shown on the 3D cube were adjacent to each other in the unfolded cube.

The flow of one trial in both experiments was described below (Figure 1A). There were 180 trials. At the start of each trial, participants choose either the left or right arrow. After the choice, an effort cue was presented. The effort cue was 100% associated with either low- or high-demand problems. After the effort cue, a problem was presented (high-demand or low-demand problem). The arrow determined the presentation probability of either low- or high-effort cues. The probabilities of arrows associated with upcoming effort cues, or effort levels, were alternated dynamically over time (Figure 1B). The set of either 80 or 20% was exchanged 5 times during 180 trials. After participants chose a cue with uncertainty, an effort cue was presented and participants knew which problem was to be presented with certainty, after that a problem was presented, and they were asked to answer it. There was no time limit for response. Participants' choices did not affect how fast they could finish the experiment and we instructed that.

In instruction, we did not explain to participants that there was a difference in the level of cognitive demands, between problem sets. Before the practice session(s), we gave verbal instructions and accepted questions from participants. After the practice session(s), participants were asked to explain what they thought the differences between the two types of problems were. If they answered that they thought the differences reflected differences in the actual content of the problems or differences in difficulty, we told them that we could not tell the answer was correct or not. If the difference reported by the participant was significantly wrong, we told them that the answer was wrong and asked them to practice. After the additional practice(s), if the participant subsequently reported things closer to reality, we told them that the answer was not incorrect. It should be noted that we never instructed the participants to avoid the high-demand problems to elicit voluntary choices. In addition, to minimize the influence of how fast the experiment could be completed, we instructed participants that their choices did not affect how fast they could finish the experiments.

Next, we will describe changes I ~ VII.

I: We added effort cues that corresponded exactly to the level of the upcoming effort to evoke cost PE signals. This is because the design of our previous experiments did not lock the time of cost PE generation. We used circle cues with red and green colors in Exp. 1. If participants were colour blind, we planned to use blue instead of green but there were no colour blindness participants. The colour combinations between upcoming difficulties and colours were counterbalanced across participants. We used upward and downward triangle

cues with grey colour in Exp. 2. The combination between upcoming difficulties and directions (upward or downward triangle) were also counterbalanced across participants.

II: We instructed that there were two types of problems and they are red and green problems in Exp. 1 and upward and downward problems in Exp. 2. In our previous study, we instructed just that there were two types of problems.

III: We changed the number of trials and some conditions in the practice block(s), comparing with our previous work. Participants executed practice block(s) for 15-75 trials (mean ± standard deviation (SD), 26.5 ± 16.1) in Exp. 1 and 15-105 trials (mean ± SD, 32.1 ± 20.1) in Exp. 2. We did not give oral instructions during the practice block(s). Interstimulus intervals from the choice timing of arrows to the presentation timing of effort cues were 0.25, 0.50, 0.65 seconds jittered in Exp. 1 and 2.50, 5.00, 6.50 seconds jittered in Exp. 2. The duration of effort cue presentation was 0.50 seconds in both Exp. 1 and 2. Interstimulus intervals from the presentation timing of effort cues to the presentation timing of problems were 0.25 seconds in Exp. 1 and 2.50 seconds in Exp. 2. Intertrial intervals from the timing of answer for problems to the presentation timing of arrows of a next trial were 0.25, 0.50, 0.65 seconds jittered in Exp. 1 and 2.50, 5.00, 6.50 seconds jittered in Exp. 2. The presentation probabilities of the two arrows, this is, probabilities of which levels of problems was presented according to the chosen arrows, were 0.50 and stable in both Exp. 1 and 2 to prevent making a choice tendency in main experiments. After or between the practice block(s), we asked participants how were red and green problems in Exp. 1 and how upward and downward triangle problems were to prevent that participant from misunderstanding the difference between the two kinds of problems (low- and high-demand problems). Notably, we did not explain the difference. If participants answered completely different, we said that your answer was different.

IV: The total trials of the main experiments in the scanner were 180 trials and the same as our previous design, but we changed the number of blocks from four to six. More specifically, there were 6 blocks multiplied by 30 trials in both Exp. 1 and 2.

V: There were 180 trials totally, which consisted of 6 blocks of 30 trials, in both Exp. 1 and 2 in the scanner. In each block, a rest period was imposed until 200, 400, and 600s in Exp. 1 and 2 after the completion of the 10th, 20th, and 30th trial in the block in <200, 400, and 600s in both Exp. 1 and 2, respectively.

VI: We checked whether participants are right-handed or not with the FLANDERS handedness questionnaire (Okubo et al., 2014).

VII: After the scan, participants answered questionnaires with some questions in both Exp. 1 and Exp. 2.

In Exp. 1, the first question was that "Did you feel that the following accurately describes the probabilities of red and green cues?"

After choosing an arrow, effort cues are presented that are 100% associated with the task types. The problems that followed the effort cues are always the division of 5-digit numbers, but there are two types of problems. The probability that each type of problem is more likely to appear is set independently for the left and right arrows. These probabilities changed very slowly over time. The same situation will continue for at least 10 trials. The presentation probabilities of which effort levels came after an arrow was chosen was maintained during the transition from block to block.

Subjects answered whether the answer is yes or no by drawing a circle as follows: Yes・No.

If the subject answered No, the subjects answered the second question. "If you answered No on the above-mentioned question, please explain what was the difference specifically."

The third question with a 7-point scale was that "How much did you like to solve either one of the red and green problems? Please draw a circle on the relevant vertical line on the horizontal bar below."

In Exp. 2, we changed the description of the first question of Exp. 1 from the division of 5-digit numbers to the spatial reasoning problems and from red and green cues to upward or downward triangles. The second and third questions were the same.

Image acquisition

We used a Phillips Ingenia 3.0T CX scanner with a 32-channel head coil (Shimane Institute of Health Science, Izumo, Shimane, Japan). Participants wore earplugs except for three participants of Exp. 1. Soft pads were used to reduce their head movements. The stimuli were presented via a mirror where the images of the monitor were reflected. Functional images were acquired in ascending order by T2-weighted echo planner imaging sequences [Flip angle: 80°; repetition time (TR): 2500ms; voxel size: 3.31mm × 3.37mm × 3.20mm; slice gap: no; field of view (FOV): 211.8 mm (Left-Right) × 212.3 mm (Anterior-Posterior) × 128.0 mm; slice number: 40; acquisition matrix: 64 × 63 × 40; volume number: 240 volumes × 6 blocks in both Exp. 1 and Exp. 2; parallel to the anterior and posterior commissures line]. We discarded the following 9 scans in each block to obtain the steady-state magnetization: the Phillips scanner deleted the first four scans automatically and we deleted the next five scans in each block for each participant.

Behavioural Analysis

Basic Analysis. Behavioural analyses were the same protocols as our previous work (please see Nagase et al., 2018). We set the same inclusion criteria of correct rate with our previous work (> 80% for both high- and low-demand problems) and one participant was excluded in both experiments. We conducted a χ^2 test and analyzed the effect of the demand that participants experienced two trials before.

Modelling, Model Fitting, and Comparisons. The models that we used were the almost same set as our previous work (please see Nagase et al., 2018), except for the following change. Concerning the following five reinforcement learning (RL) models, *ExpectedCost* for the unchosen option at the first trial [$ExpectedCost_{unchosen}(1)$] was assumed to be equal to $ExpectedCost_{Chosen}(1)$.

(1) solve time in individual trials (RL-Solve-Time-(ST) model)

(2) demand level of the problem (RL-High-Low (HL) model)

(3) incorrect solving (RL-Incorrect-Correct (IC) model)

(4) sum of (1) and (3), with a weighting parameter for (3) ($w_{incorrect}$) (RL-ST-IC model)

(5) sum of (2) and (3), with a weighting parameter for (3) ($w_{incorrect}$) (RL-HL-IC model)

This is because means of the Bayesian information criterion (BIC) of the models where $ExpectedCost_{unchosen}(1)$ was set to be a free parameter were worse than the BIC means of the above-mentioned models that had the same initial values in our previous results and the presentation probabilities of which effort levels came after choices were 0.5 in practice sessions (please see the change III). Finally, we tested the following eleven models that were used in our previous work: a RL model based on solve time (RL-ST model), a low/high-demand model (RL-HL), a model based on incorrect responses (RL-IC),a model that combines solve time and incorrect responses ST + IC (RL-ST-IC), a mode that combines low/high demands and incorrect choices (RL-HL-IC), a rest-time RL model (RL-Rest) and a rest-time model that also has

information about low/high demands (RL-Rest-HL). We also tested models based on Win-Stay-Lose-Shift probabilities (pWSLS), including information about low/high demands (pWSLS-HL), and incorrect responses (pWSLS-IC), and also a probabilistic selection (PS) model based on high/low demands (PS-HL) and incorrect responses (PS-IC). The reason we used multiple RL models was to test whether the choice behaviour was explained better by mental effort / demand rather than by other factors such as rest time or the feeling of mistakes. On the other hand, the reason we used several probability models was that the WSLS model is a widely-used model of choice behaviour and reward-seeking and we thought that the probabilistic selection model likely model based on observation of the choice data.

For convenience, we describe the RL-HL and PS-HL models below (which provided similarly good fits, according to model comparison, see Results), although the same description was also provided in our previous work (Nagase *et al.*, 2018). The RL-HL model was a RL model based on binarized effort cost, corresponding to low/high-demand. It was assumed that participants had the representation of the "Expected Cost (*ExpectedCost*)" of mental effort for options (i.e., arrows), this is, $ExpectedCost_{left}(k)$ and $ExpectedCost_{right}(k)$ (k = 1, 2, ..., 180: trials). At each trial k, either the left or right arrow was assumed to be chosen with the probabilities $P_{left}(k)$ and $P_{right}(k)$, respectively:

$$P_{left}(k) = \frac{\exp\left(-\beta \cdot ExpectedCost_{left}(k)\right)}{\exp\left(-\beta \cdot ExpectedCost_{left}(k)\right) + \exp\left(-\beta \cdot ExpectedCost_{right}(k)\right)}, \text{ and}$$

$$P_{right}(k) = \frac{\exp\left(-\beta \cdot ExpectedCost_{right}(k)\right)}{\exp\left(-\beta \cdot ExpectedCost_{left}(k)\right) + \exp\left(-\beta \cdot ExpectedCost_{right}(k)\right)}$$

$$= 1 - P_{left}(k),$$

where β was a free parameter was the inverse temperature. The "Cost Prediction Error (*CPE*)" was also assumed to be calculated:

$$CPE(k) = ActualCost(k) - ExpectedCost_{choice(k)}(k),$$

where Choice(k) was Left or Right depending on which was chosen. As $ActualCost(k)$, we considered demand level of the problem, more specifically, 1 and 0 for high- and low-demand problems, respectively.

ExpectedCost for the chosen option was then assumed to be updated as follows:

$$ExpectedCost_{Choice(k)}(k+1) = ExpectedCost_{choice(k)}(k) + \alpha \cdot CPE(k),$$

where α was a free parameter representing the learning rate. *ExpectedCost* for the unchosen option was assumed to be unchanged. *ExpectedCost* for the option chosen and unchosen at the 1st trial ($ExpectedCost_{chosen}(1)$ and $ExpectedCost_{unchosen}(1)$) was assumed 0.5 equally.

The PS-HL model was a probabilistic selection model based on the experiences of low/high-demand. Participants repeated their choice from the preceding trial with the probability a and b after solving low- and high-demand problems, respectively.

Please note that we referred to the "RL model" in this article as the "PE model" in our previous article. As a result of model comparisons (see Results), we judged that it was reasonable to use the RL-HL model in the model-based fMRI analyses described below.

We divided eleven models into three groups: RL, rest-time RL, and probabilistic models; RL-ST model, RL-HL model, RL-IC model, RL-ST-IC model, RL-HL-IC model, RL-Rest model and RL-Rest-HL model, and probabilistic models including pWSLS-HL model and pWSLS-IC model, and PS-HL model and PS-IC model.

Functional Imaging Analysis

We used SPM12 (Wellcome Trust Centre for Neuroimaging, UCL, London, https://www.fil.ion.ucl.ac.uk/spm/software/spm12/) for both preprocessing and analysis of functional imaging data. The preprocessing included the following steps: realigning to the first images using a six-parameter rigid body transformation, slice timing correction, coregistration, normalization, smoothing (with Gaussian kernel with a full-width at half-maximum of 8 mm). We set the inclusion criterion of both head movement (≤ 3mm) and behavioural performance (the above-mentioned high correct rate and the significant avoidance of high-demand problems on a χ^2 test ($p < 0.01$)) as same as our previous work. We excluded 2 and 1 (Exp. 1 and 2, respectively) participants with >3 mm head movements among the participants who met the performance criterion and showed significant avoidance of high-demanding problems (see the Results). We defined striatum as dorsal striatum (DS, putamen and caudate) and VS (nucleus accumbens and the surrounding regions) in this article.

We conducted general linear model (GLM) analyses of the BOLD data (Exp. 1, n=25; Exp. 2, n=. 23). To turn off the default orthogonalization for multiple parametric modulations as same as our previous work, we commented out the lines 115-117 of spm_get_ons.m and the lines 235-245 of spm_fMRI_design.m. The analyses of both individual and group levels were done at the whole-brain in each experiment.

We used three GLMs (see Figure 2) at the individual level analyses. In a trial, there were five events: arrow-cue presentation, the choice of an arrow, effort cue presentation, problem presentation, and the answer of problems.

First, we designed the GLM1 to test the first hypothesis that effort cues update the expected cost and induce the activity correlated with cost-PE in dmFC/dACC positively and in striatum negatively. Additionally, we also tested our a prior prediction that activity was correlated with *ExpectedCost* for the chosen option since the signals of the expected cost in our previous design could be confounded with the effect of task preparation.

Next, we designed GLM2 and GLM3 to test the revised hypothesis that the expected cost update requires actual task execution and the activity correlated with cost-PE in dmFC/dACC positively and in striatum is detected at the time of problem presentation or the answer of problems.

The trial-by-trial values of both *ExpectedCost* for the chosen option (referred to as $ExpectedCost_{chosen}$) and the Cost Prediction Error (*CPE*) for individual participants were estimated by the RL-HL model. The response time from the arrow-cue presentation to the choice of an arrow was referred to as RT_{choice}.

GLM1~3 were designed in the following manner. We set the regressors at the time of arrow-cue presentation with parametric modulations by $ExpectedCost_{chosen}$ and RT_{choice}, the regressors at the time of effort cue presentation with parametric modulations by problem-demand (0 and 1 for low- and high-demand problems, respectively) and CPE (this parametric modulation of CPE is only for GLM1), the regressor at the time of problem presentation with parametric modulation by CPE (this regressor with parametric modulation of CPE is only for GLM2), the regressors with the duration from problem presentation to the answer of problems with parametric modulations by problem-demand, the regressors at time of the answer of problems with parametric modulation by CPE (this regressor with parametric modulation of CPE is only for GLM3), the regressors for motor response at the time of the choice of an arrow and the answer of problems (GLM1 and GLM2 included them both timing and GLM3 included them only at the time of the choice of an arrow) and head movements. Each regressor in each GLM was convolved with the canonical hemodynamic response function, and one-sample t-tests were performed on individual maps of the regressors of interest for 25 participants in Exp. 1 and 23 participants in Exp. 2. The variance inflation factor (VIF) was calculated to

determine whether the collinearity of the regressors of interest was within an acceptable range where the cut-off value for the VIF used in collinearity issues was 5 or 10.

For GLM 1-3, the variance inflation factors (VIF) of CPE parametric regressors at three timings were smaller than 5 across sessions and participants, which was judged it to be acceptable for collinearity.

Our conjunction analyses with a binary mask were the same as our previous work (please see Nagase *et al.*, 2018). We used the strict mask of common voxels between the results of Exp. 1 and Exp. 2 exceeding the threshold of cluster-level FWE corrected $p < 0.05$ and voxel-level uncorrected $p < 0.001$. We then reported correlates detected in the masked conjunction analyses with a threshold of family wise error (FWE) corrected $p < 0.05$ and voxel-level uncorrected $p < 0.001$. The relaxed mask reflected common voxels between Exp. 1 and Exp. 2 with the threshold of voxel-level uncorrected $p < 0.01$. We reported correlates detected in the relaxed masked conjunction analyses with a threshold of voxel-level uncorrected $p < 0.001$.

Resource Availability

Upon request, the data of this study are available from the corresponding author, except a participant's MRI data of individual-level analysis and MRI structure data in Exp. 1 because we could not get the agreement of data share of the participant.

Supplemental information

Supplemental Figure 1. Hypotheses and the Predictions.

Supplemental Figure 2. The effect of experiences of two preceding trials.

Supplemental Figure 3. Results of model fitting for demand avoiders in Exp. 1.

Supplemental Figure 4. Results of model fitting for demand avoiders in Exp.2.

Supplemental Figure 5. BIC scores.

Supplemental Figure 6. Neural correlates of cost-prediction-error at the time of effort cues by using GLM1.

Supplemental Figure 7. Neural correlates of the cost prediction error at the times of a presentation of an effort cue with relaxed mask.

Supplemental Figure 8. Neural correlates of cost-prediction-error at the time of effort initiation by using GLM2.

Supplemental Figure 9. Neural correlates of cost-prediction-error at the time of effort completion by using GLM3.

Supplemental Figure 10. Overlap of cost-prediction-error at the time of effort initiation among previous and current results.

Supplemental Table 1. BIS scores and best-fit parameters.

Supplemental Table 2. Neural correlates of cost-prediction-error at the time of effort cues by using GLM1.

Supplemental Table 3. Neural correlates of Task Preparation at the time of effort cues by using GLM1.

740 **Supplemental Table 4. Neural correlates of cost-prediction-error at the time of effort initiation by**
741 **using GLM2.**

742 **Supplemental Table 5. Neural correlates of cost-prediction-error at the time of effort completion by**
743 **using GLM3.**

744

745 **References**

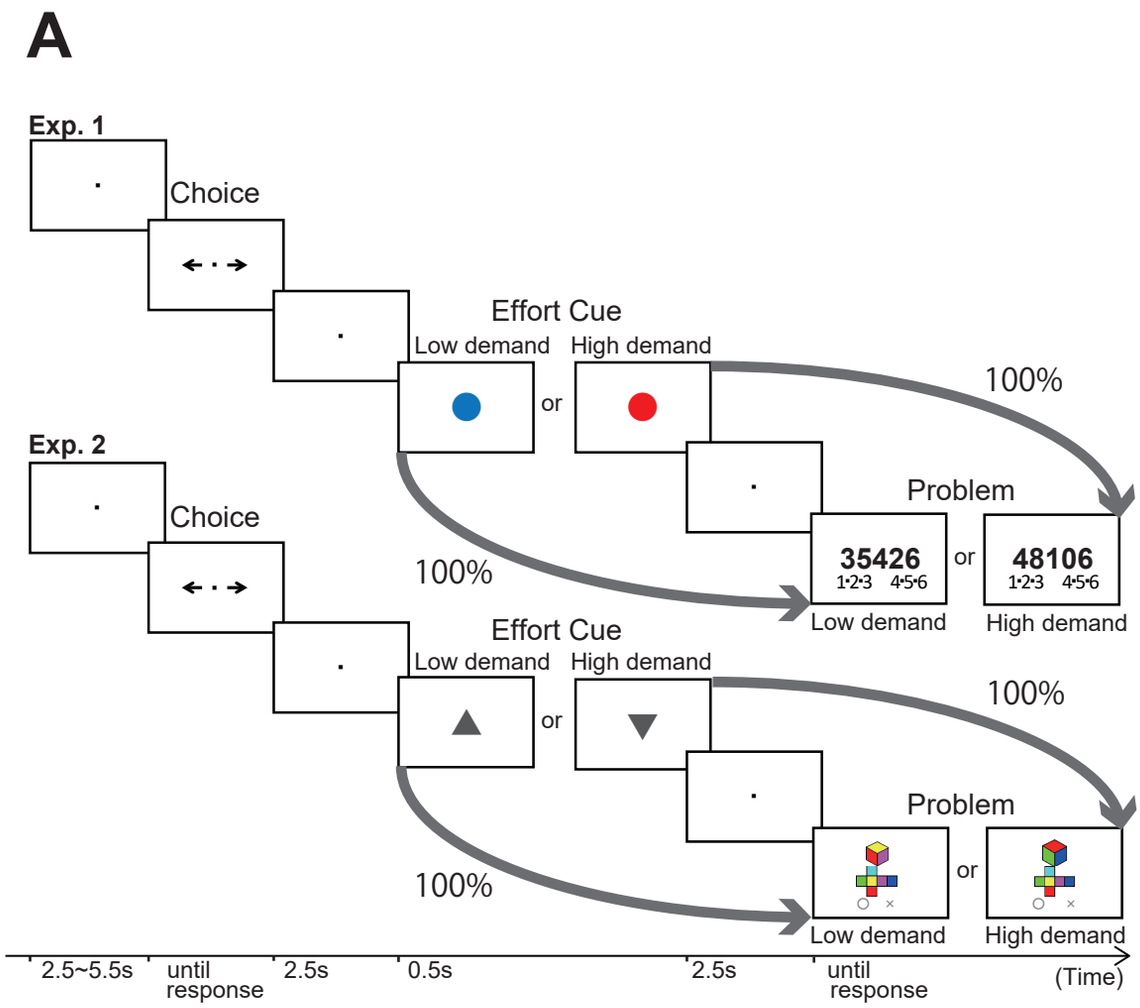
746 Alexander, G.E., Delong, M.R., and Strick, P.L. (1986). PARALLEL ORGANIZATION OF FUNCTIONALLY SEGREGATED
747 CIRCUITS LINKING BASAL GANGLIA AND CORTEX. Annual Review of Neuroscience *9*, 357-381.
748 10.1146/annurev.ne.09.030186.002041.
749 Alexander, W.H., and Brown, J.W. (2011). Medial prefrontal cortex as an action-outcome predictor. Nature
750 Neuroscience *14*, 1338-U1163. 10.1038/nn.2921.
751 Bang, D., and Fleming, S.M. (2018). Distinct encoding of decision confidence in human medial prefrontal cortex.
752 Proceedings of the National Academy of Sciences *115*, 6082-6087. 10.1073/pnas.1800795115.
753 Botvinick, M.M., Huffstetler, S., and McGuire, J.T. (2009). Effort discounting in human nucleus accumbens.
754 Cognitive Affective & Behavioral Neuroscience *9*, 16-27. 10.3758/cabn.9.1.16.
755 Bouret, S., and Sara, S.J. (2005). Network reset: a simplified overarching theory of locus coeruleus noradrenaline
756 function. Trends in Neurosciences *28*, 574-582. 10.1016/j.tins.2005.09.002.
757 Cavanagh, J.F., Masters, S.E., Bath, K., and Frank, M.J. (2014). Conflict acts as an implicit cost in reinforcement
758 learning. Nature Communications *5*, 5394. 10.1038/ncomms6394.
759 Doya, K. (2008). Modulators of decision making. Nature Neuroscience *11*, 410-416. 10.1038/nn2077.
760 Gan, J.O., Walton, M.E., and Phillips, P.E.M. (2010). Dissociable cost and benefit encoding of future rewards by
761 mesolimbic dopamine. Nature Neuroscience *13*, 25-27. 10.1038/nn.2460.
762 Glimcher, P.W., and Fehr, E. (2014). Introduction: A Brief History of Neuroeconomics. In Neuroeconomics (Second
763 Edition), P.W. Glimcher, and E. Fehr, eds. (Academic Press), pp. xvii-xxviii. https://doi.org/10.1016/B978-0-12-
764 416008-8.00035-8.
765 Haber, S.N., and Knutson, B. (2010). The Reward Circuit: Linking Primate Anatomy and Human Imaging.
766 Neuropsychopharmacology *35*, 4-26. 10.1038/npp.2009.129.
767 Hauser, T.U., Eldar, E., and Dolan, R.J. (2017). Separate mesocortical and mesolimbic pathways encode effort and
768 reward learning signals. Proceedings of the National Academy of Sciences of the United States of America *114*,
769 E7395-E7404. 10.1073/pnas.1705643114.
770 Hauser, T.U., Hunt, L.T., Iannaccone, R., Walitza, S., Brandeis, D., Brem, S., and Dolan, R.J. (2015). Temporally
771 Dissociable Contributions of Human Medial Prefrontal Subregions to Reward-Guided Learning. Journal of
772 Neuroscience *35*, 11209-11220. 10.1523/jneurosci.0560-15.2015.
773 Holroyd, C.B., and Verguts, T. (2021). The Best Laid Plans: Computational Principles of Anterior Cingulate Cortex.
774 Trends in Cognitive Sciences *25*, 316-329. 10.1016/j.tics.2021.01.008.
775 Klein-Flügge, M.C., Bongioanni, A., and Rushworth, M.F.S. (2022). Medial and orbital frontal cortex in decision-
776 making and flexible behavior. Neuron. 10.1016/j.neuron.2022.05.022.
777 Kool, W., McGuire, J.T., Rosen, Z.B., and Botvinick, M.M. (2010). Decision Making and the Avoidance of Cognitive
778 Demand. Journal of Experimental Psychology-General *139*, 665-682. 10.1037/a0020198.
779 Lopez-Gamundi, P., Yao, Y.-W., Chong, T.T.J., Heekeren, H.R., Herrero, E.M., and Pallares, J.M. (2021). The neural
780 basis of effort valuation: A meta-analysis of functional magnetic resonance imaging studies. Neuroscience &
781 Biobehavioral Reviews. 10.1016/j.neubiorev.2021.10.024.
782 Ma, W.J., and Jazayeri, M. (2014). Neural Coding of Uncertainty and Probability. In Annual Review of Neuroscience,
783 Vol 37, S.E. Hyman, ed. pp. 205-220. 10.1146/annurev-neuro-071013-014017.
784 Matsumoto, M., and Hikosaka, O. (2009). Representation of negative motivational value in the primate lateral
785 habenula. Nature Neuroscience *12*, 77-84. 10.1038/nn.2233.
786 Matsumoto, M., Matsumoto, K., Abe, H., and Tanaka, K. (2007). Medial prefrontal cell activity signaling prediction
787 errors of action values. Nature Neuroscience *10*, 647-656. 10.1038/nn1890.

McClure, S.M., Berns, G.S., and Montague, P.R. (2003). Temporal prediction errors in a passive learning task activate human striatum. Neuron *38*, 339-346. 10.1016/s0896-6273(03)00154-5.
Miyamoto, K., Osada, T., Setsuie, R., Takeda, M., Tamura, K., Adachi, Y., and Miyashita, Y. (2017). Causal neural network of metamemory for retrospection in primates. Science *355*, 188-193. 10.1126/science.aal0162.
Miyamoto, K., Trudel, N., Kamermans, K., Lim, M.C., Lazari, A., Verhagen, L., Wittmann, M.K., and Rushworth, M.F.S. (2021a). Identification and disruption of a neural mechanism for accumulating prospective metacognitive information prior to decision-making. Neuron. 10.1016/j.neuron.2021.02.024.
Miyamoto, K., Trudel, N., Kamermans, K., Lim, M.C., Lazari, A., Verhagen, L., Wittmann, M.K., and Rushworth, M.F.S. (2021b). Identification and disruption of a neural mechanism for accumulating prospective metacognitive information prior to decision-making. Neuron *109*, 1396-+. 10.1016/j.neuron.2021.02.024.
Nagase, A.M., Onoda, K., Foo, J.C., Haji, T., Akaishi, R., Yamaguchi, S., Sakai, K., and Morita, K. (2018). Neural Mechanisms for Adaptive Learned Avoidance of Mental Effort. Journal of Neuroscience *38*, 2631-2651. 10.1523/jneurosci.1995-17.2018.
Okubo, M., Suzuki, H., and Nicholls, M.E.R. (2014). A Japanese version of the FLANDERS handedness questionnaire. The Japanese journal of psychology *85*, 474-481. 10.4992/jjpsy.85.13235.
Reynolds, J.N.J., and Wickens, J.R. (2002). Dopamine-dependent plasticity of corticostriatal synapses. Neural Networks *15*, 507-521, Pii s0893-6080(02)00045-x. 10.1016/s0893-6080(02)00045-x.
Rushworth, M.F.S., and Behrens, T.E.J. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. Nature Neuroscience *11*, 389-397. 10.1038/nn2066.
Sayali, C., and Badre, D. (2019). Neural systems of cognitive demand avoidance. Neuropsychologia *123*, 41-54. 10.1016/j.neuropsychologia.2018.06.016.
Sayalı, C., and Badre, D. (2021). Neural systems underlying the learning of cognitive effort costs. Cogn Affect Behav Neurosci. 10.3758/s13415-021-00893-x.
Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and reward. Science *275*, 1593-1599. 10.1126/science.275.5306.1593.
Schultz, W., Stauffer, W.R., and Lak, A. (2017). The phasic dopamine signal maturing: from reward via behavioural activation to formal economic utility. Current Opinion in Neurobiology *43*, 139-148. 10.1016/j.conb.2017.03.013.
Seymour, B., Daw, N., Dayan, P., Singer, T., and Dolan, R. (2007). Differential encoding of losses and gains in the human striatum. Journal of Neuroscience *27*, 4826-4831. 10.1523/jneurosci.0400-07.2007.
Seymour, B., O'Doherty, J.P., Dayan, P., Koltzenburg, M., Jones, A.K., Dolan, R.J., Friston, K.J., and Frackowiak, R.S. (2004). Temporal difference models describe higher-order learning in humans. Nature *429*, 664-667. 10.1038/nature02581.
Sharpe, M.J., Stalnaker, T., Schuck, N.W., Killcross, S., Schoenbaum, G., and Niv, Y. (2019). An Integrated Model of Action Selection: Distinct Modes of Cortical Control of Striatal Decision Making. Annu Rev Psychol *70*, 53-76. 10.1146/annurev-psych-010418-102824.
Shenhav, A., Botvinick, M.M., and Cohen, J.D. (2013). The Expected Value of Control: An Integrative Theory of Anterior Cingulate Cortex Function. Neuron *79*, 217-240. 10.1016/j.neuron.2013.07.007.
Silvetti, M., Vassena, E., Abrahamse, E., and Verguts, T. (2018). Dorsal anterior cingulate-brainstem ensemble as a reinforcement meta-learner. Plos Computational Biology *14*, e1006370. 10.1371/journal.pcbi.1006370.
Sutton, R.S., and Barto, A.G. (2018). Reinforcement Learning: An Introduction second edition Introduction.
Vassena, E., Holroyd, C.B., and Alexander, W.H. (2017). Computational Models of Anterior Cingulate Cortex: At the Crossroads between Prediction and Effort. Frontiers in Neuroscience *11*, 316. 10.3389/fnins.2017.00316.
Wallis, J.D., and Rushworth, M.F.S. (2014). Chapter 22 - Integrating Benefits and Costs in Decision Making. In Neuroeconomics (Second Edition), P.W. Glimcher, and E. Fehr, eds. (Academic Press), pp. 411-433. https://doi.org/10.1016/B978-0-12-416008-8.00022-X.
Walton, M.E., and Bouret, S. (2019). What Is the Relationship between Dopamine and Effort? Trends in Neurosciences *42*, 79-91. 10.1016/j.tins.2018.10.001.
Westbrook, A., Frank, M.J., and Cools, R. (2021). A mosaic of cost–benefit control over cortico-striatal circuitry. Trends in Cognitive Sciences. 10.1016/j.tics.2021.04.007.
Westbrook, A., Kester, D., and Braver, T.S. (2013). What Is the Subjective Cost of Cognitive Effort? Load, Trait, and Aging Effects Revealed by Economic Preference. Plos One *8*, e68210. 10.1371/journal.pone.0068210.

Westbrook, A., Lamichhane, B., and Braver, T. (2019). The Subjective Value of Cognitive Effort is Encoded by a Domain-General Valuation Network. Journal of Neuroscience *39*, 3934-3947. 10.1523/jneurosci.3071-18.2019.

Westbrook, A., van den Bosch, R., Maatta, J.I., Hofmans, L., Papadopetraki, D., Cools, R., and Frank, M.J. (2020). Dopamine promotes cognitive effort by biasing the benefits versus costs of cognitive work. Science *367*, 1362-+. 10.1126/science.aaz5891.

Wilson, R.C., Takahashi, Y.K., Schoenbaum, G., and Niv, Y. (2014). Orbitofrontal Cortex as a Cognitive Map of Task Space. Neuron *81*, 267-279. 10.1016/j.neuron.2013.11.005.


A
Exp. 1
Choice
Effort Cue
Low demand
High demand
or
100%
Exp. 2
Choice
Problem
35426
1·2·3 4·5·6
or
48106
1·2·3 4·5·6
Low demand
High demand
100%
Effort Cue
Low demand
High demand
or
100%
Problem
or
Low demand
High demand
100%
2.5~5.5s
until response
2.5s
0.5s
2.5s
until response
(Time)

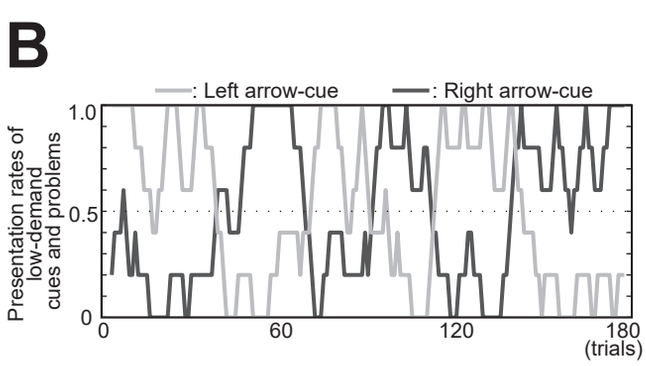

B
: Left arrow-cue
: Right arrow-cue
Presentation rates of low-demand cues and problems
1.0
0.5
0
0
60
120
180
(trials)

Figure 1.

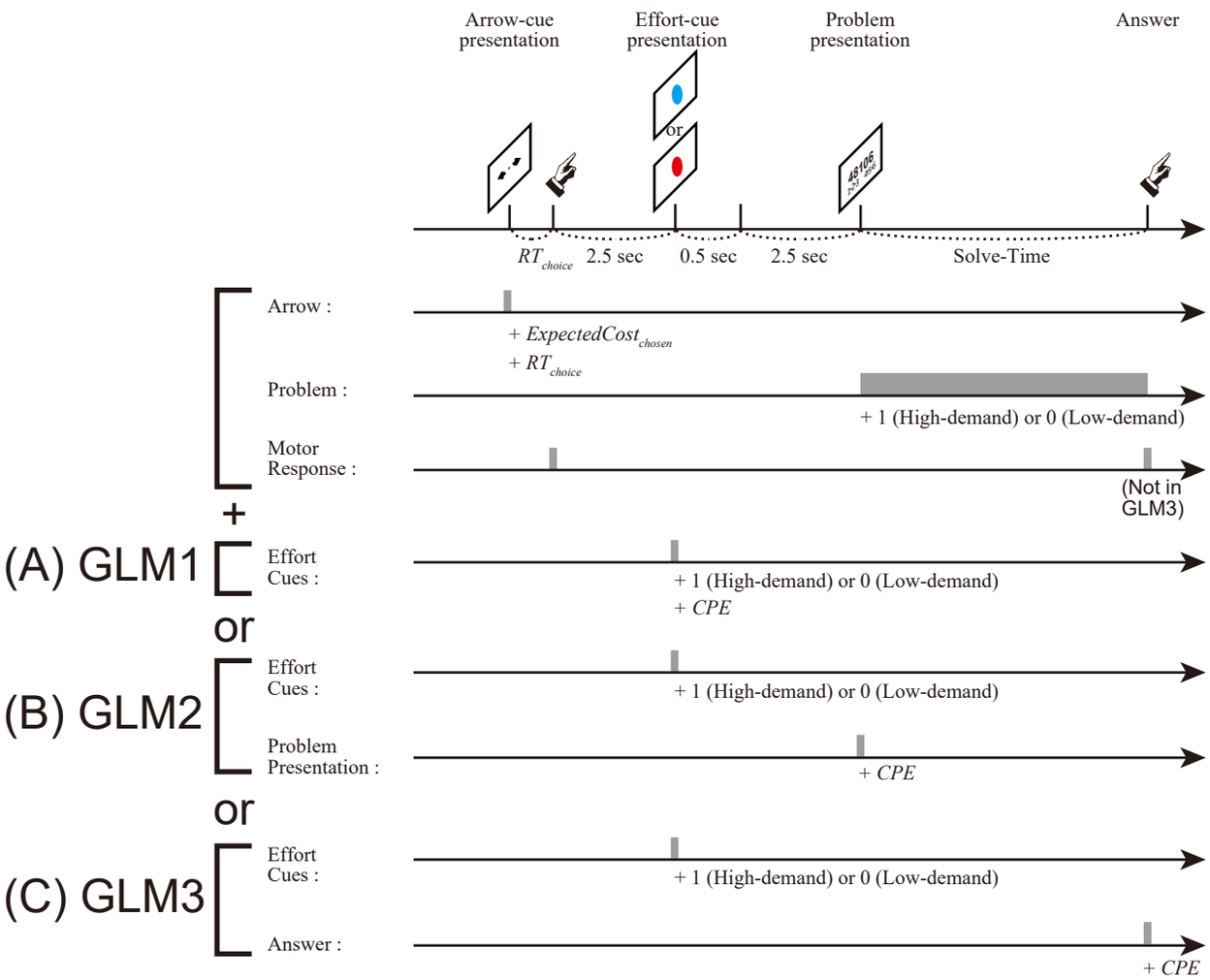
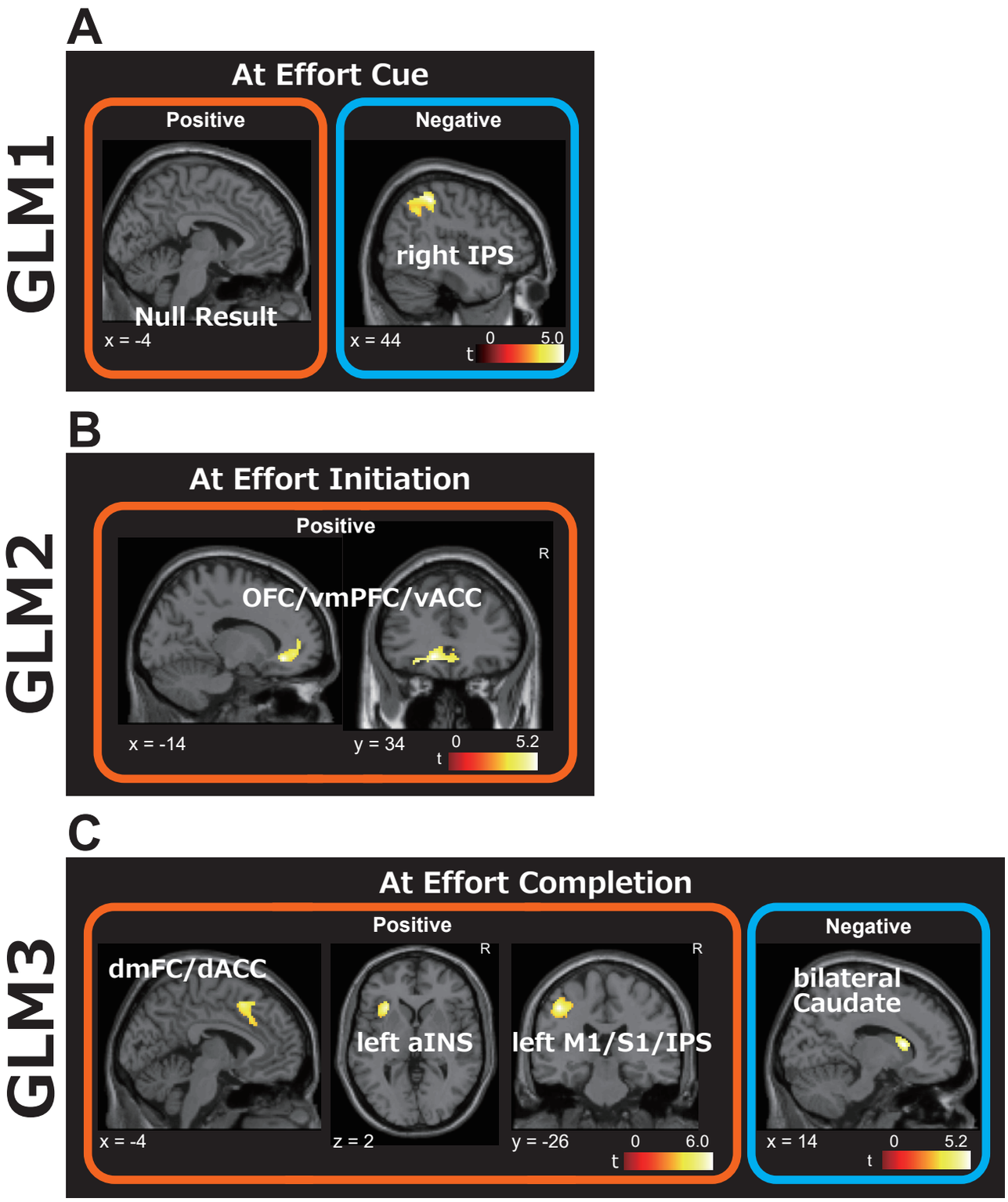

Arrow-cue presentation
Effort-cue presentation
Problem presentation
Answer
or
RT_{choice}
2.5 sec
0.5 sec
2.5 sec
Solve-Time
Arrow :
+ $ExpectedCost_{chosen}$
+ RT_{choice}
Problem :
+ 1 (High-demand) or 0 (Low-demand)
Motor Response :
(Not in GLM3)
+
(A) GLM1
Effort Cues :
+ 1 (High-demand) or 0 (Low-demand)
+ CPE
or
(B) GLM2
Effort Cues :
+ 1 (High-demand) or 0 (Low-demand)
Problem Presentation :
+ CPE
or
(C) GLM3
Effort Cues :
+ 1 (High-demand) or 0 (Low-demand)
Answer :
+ CPE

Figure 2.

Neural Correlates of CPE

Conjunction Analysis among Exp. 1 and Exp. 2

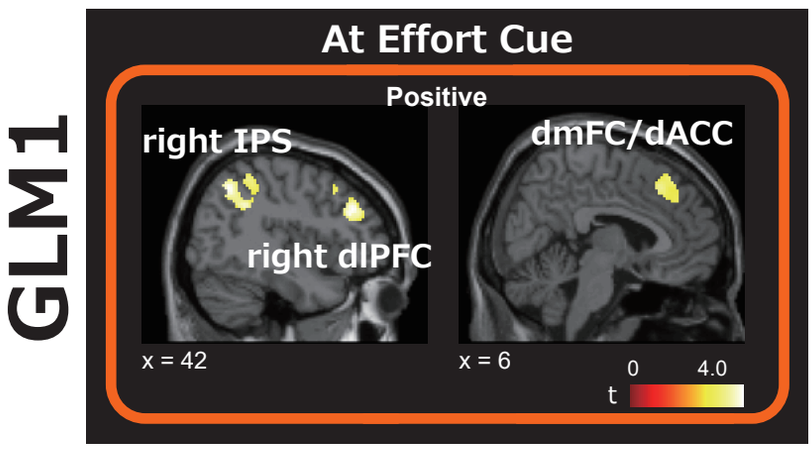


cluster-level FWE corrected p<0.05, voxel-level uncorrected p<0.001

Figure 3.

Neural Correlates of Task Preparation

Conjunction Analysis
among Exp. 1 and Exp. 2



cluster-level FWE corrected p<0.05, voxel-level uncorrected p<0.001

Figure 4.

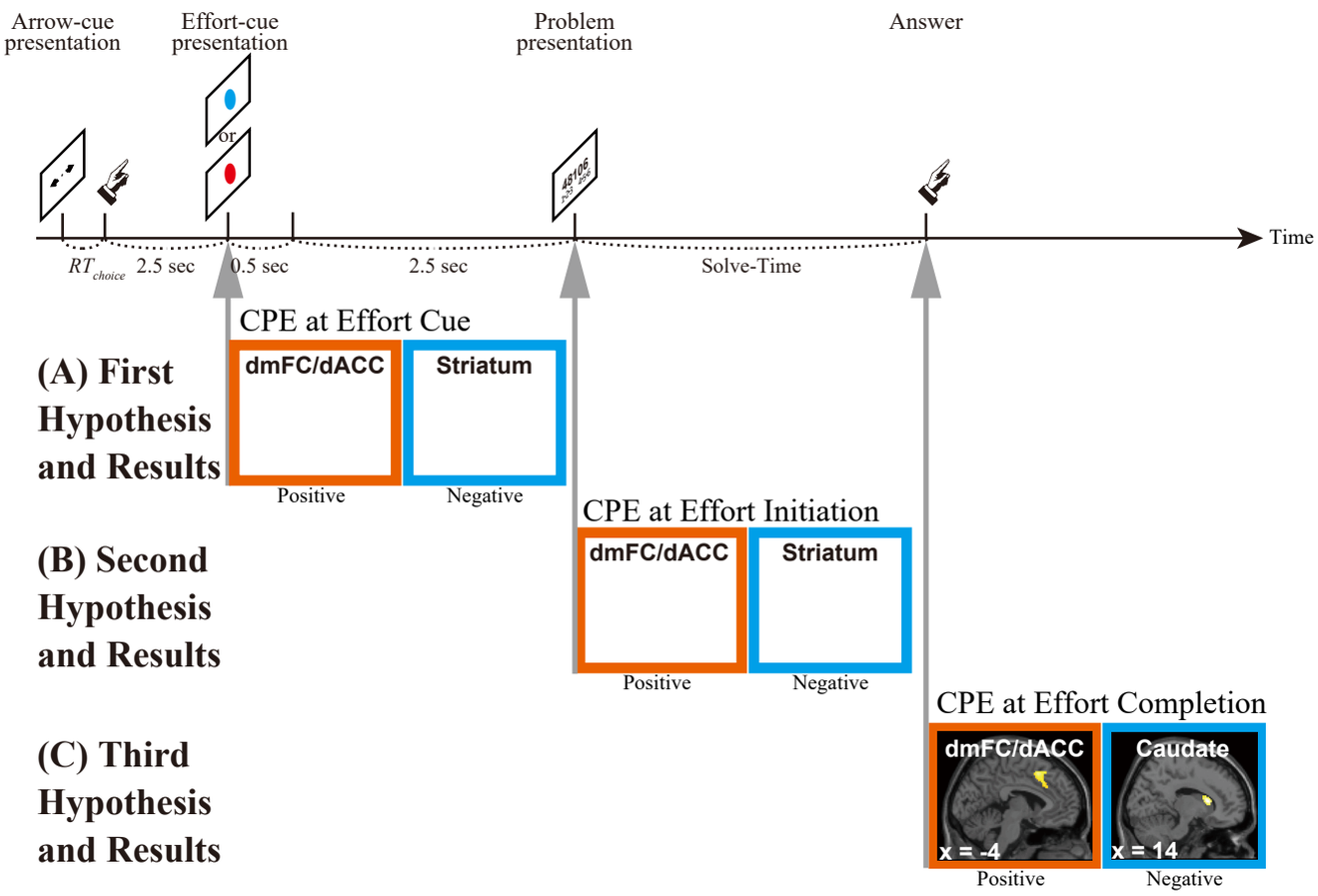

Arrow-cue presentation
Effort-cue presentation
Problem presentation
Answer
or
Time
RT choice
2.5 sec
0.5 sec
2.5 sec
Solve-Time
CPE at Effort Cue
(A) First Hypothesis and Results
dmFC/dACC
Striatum
Positive
Negative
CPE at Effort Initiation
(B) Second Hypothesis and Results
dmFC/dACC
Striatum
Positive
Negative
CPE at Effort Completion
(C) Third Hypothesis and Results
dmFC/dACC
Caudate
x = -4
x = 14
Positive
Negative

Figure 5.