Title:

Mental Effort Cost Learning is Retrospective

Abbreviated Title:

Mental Effort Learning is Retrospective

Authors:

Asako Mitsuto (光戸麻子)[*1,2,3,4,5,6], Rei Akaishi (赤石れい)[7], Keiichi Onoda (小野田慶一)[5,8], Kenji Morita (森田賢治)[4,6,9], Toshikazu Kawagoe (川越敏和)[5,10], Tetsuya Yamamoto (山本哲也)[11], Shuhei Yamaguchi (山口修平)[#5,12], Ritsuko Hanajima (花島律子)[#2], and Andrew Westbrook[#1]

[1] Department of Psychiatry, Center for Advanced Human Brain Imaging Research, Rutgers University, Piscataway, NJ 08854, USA

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

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

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

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

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

[7] Center for Brain Science, RIKEN, Wako, Saitama 351-0198, Japan

[8] Faculty of Psychology, Otemon Gakuin University, Ibaraki, Osaka 567-8502, Japan

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

[10] School of Humanities and Science, Kumamoto Campus, Tokai University, Toroku 9-1-1, Kumamoto-city, Kumamoto 862-8652, Japan

[11] Brain Function Imaging and Support Center, Laboratory of Biofunctional Information Analysis, National Institute for Physiological Sciences

[12] Department of Neurology, Shimane Prefectural Central Hospital, Izumo 693-8555, Japan

*Correspondence: asako.3210@rutgers.edu

[#]These senior authors contributed equally.

Keywords:

frontal cortex, striatum, mental effort, reinforcement learning, fMRI, prediction error, uncertainty, temporal-difference learning

Abstract

To understand why people avoid mental effort, it is crucial to reveal the mechanisms by which we learn and decide about mental effort costs. This study investigated whether mental effort cost learning aligns with temporal-difference (TD) learning or alternative mechanisms. Model-based fMRI analyses showed no correlation between cost prediction errors (CPEs) and activity in the dorsomedial frontal cortex/dorsal anterior cingulate cortex (dmFC/dACC) or striatum at the time of a fully informative effort cue about upcoming effort demands, contradicting the TD hypothesis. Instead, CPEs correlate with dmFC/dACC (positively) and caudate (negatively) activity at effort completion. Furthermore, only activity patterns at effort completion predict subsequent choices. These results show that decision policies are updated retrospectively at effort completion, updating expected costs with prediction error between experienced effort and prior expectations, demonstrating mental effort cost learning is retrospective, and imply that adaptive learning of mental effort cost does not follow canonical TD learning.

Significance Statement

Understanding how people learn about mental effort costs is essential for advancing theories of motivation and cognitive control. However, the algorithms supporting such learning remain unclear. This study addressed this gap and found that temporal-difference learning, commonly used to explain reward learning, could not account for how people learn about effort. Instead, decision policies were updated retrospectively at effort completion, based on a prediction error between experienced effort and prior expectations. These findings reveal that mental effort cost learning is fundamentally retrospective and imply that it relies on mechanisms distinct from canonical temporal-difference learning.

Introduction

People exert effort adaptively to achieve their goals. Understanding goal-directed behavior requires revealing the mechanisms by which people learn and decide about effort. To make effort-based decisions, humans and animals integrate effort costs and benefits. Reward learning mechanisms are well understood and effort cost-benefit evaluation has been studied extensively[1-8]. However, the neural mechanisms by which mental effort costs are learned remains unclear[9-15].

A central question is what algorithms govern mental effort cost learning. Like rewards, effort may be learned through reinforcement learning (RL), with the brain generating cost prediction errors (CPEs)[11-13]. However, when and how CPEs are generated remains unknown. Negatively-valenced stimuli such as pain, air puff, or reward omission are learned via temporal-difference (TD) learning[16-18]. TD algorithms involve an update of the estimated value of future outcomes, based on the difference between value estimates immediately before and after information about reinforcers becomes available. Thus, in TD learning, informative cues should evoke a CPE signal. However, evidence suggests that physical or mental effort demands appear to be encoded after effort is exerted, or even retrospectively, at the time of subsequent reward[19-21]. Because prior studies have not dissociated effort production from effort information, it remains unclear whether information about mental effort demands alone is sufficient to trigger CPE

signaling, as predicted by TD learning. Moreover, because effort encoding has mostly been studied in terms of reward discounting[20-24], it is unclear how effort cost learning proceeds in the absence of reward processing. In this study, we use fMRI to examine neural dynamics while humans learn about mental effort costs, in the absence of rewards, in a task where information about upcoming effort demands is separated in time from effort exertion itself.

Initially, we hypothesized that mental effort cost learning, like reward learning, is governed by TD learning. Thus, we predicted that informative cues about upcoming effort demands would trigger CPE signaling. Moreover, we predicted that CPEs would, at the time of 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 the firing rates of distinct populations of mPFC neurons encode the PE of rewards and punishments positively[25,26]. Second, our previous fMRI study for humans showed that the activity in the dmFC/dACC and other regions correlates positively with mental effort CPE signals[11] – though in that study, we did not dissociate effort exertion from effort information processing. Third, human fMRI data reveal a positive correlation of PE of physical effort with the activity in the dmFC[24].

There is also reason to suspect that CPE signals correlate negatively with activity

62 in the striatum. In terms of anatomy, parallel cortico-basal ganglia loops are a likely
63 substrate for effort cost learning given their role in action selection via dopamine-
64 mediated plasticity[4,27,28]. If reward PEs reflect integrated utility, CPEs should correlate
65 negatively with striatal activity, contrasting with a positive correlation for reward PEs[29-
66 31]. We note, however, that PEs for positive punishment, such as pain and loss of financial
67 reward, correlate positively with striatal activity[17,29]. As such, the sign of CPE encoding
68 in the dmFC/dACC and the striatum at the time of the effort cues is uncertain.

69 To test our hypotheses, we designed two fMRI experiments of demand-selection
70 tasks with informative cues about upcoming effort which would, according to TD learning,
71 permit participants to generate CPE signals based on the differences between expected
72 effort [11]costs before and after effort cues. In a trial, participants chose between two options
73 associated probabilistically with either high or low demand mental arithmetic and spatial
74 reasoning tasks (Figure 1A). High/low demand probabilities for each option changed over
75 time. We next presented effort cues which were fully informative of upcoming effort
76 levels. The first hypothesis predicted that, at the time of effort cues, brains represent CPE
77 signals (Figure 4Aa). Contrary to the first hypothesis, CPE signals did not correlate with
78 the activity in the dmFC/dACC or striatum, at the time of effort cues. We thus speculated
79 that mental effort cost learning does not follow canonical TD learning and considered
80 alternative hypotheses. Specifically, we hypothesized that expected costs may instead be

updated at the point of effort initiation, when participants have committed to an action (Figure 4Ab). Although we found that CPE signals correlated positively with the activity of the ventromedial prefrontal cortex (vmPFC)/ventral anterior cingulate cortex (vACC)/orbitofrontal cortex (OFC), this result did not confirm our second hypothesis about the anatomical location of the CPE which we predicted would involve the anterior cingulate cortex and the striatum. We next hypothesized that an update of expected effort requires effort completion, and thus we predicted that the brain would encode CPEs at the moment that effort is concluded (Figure 4Ac). Confirming our third hypothesis, CPE signals correlated positively with the activity of the dmFC/dACC and other regions, and negatively in the bilateral caudate nucleus (Cd), at the time of effort completion. Moreover, a test of the timing of choice policy updating revealed that residual blood oxygenation level dependent (BOLD) signals predicted the next trials' choice (stay/switch) only at the time of effort completion. These results suggest that the adaptive learning of mental effort cost is retrospective.

Results

Summary of experimental designs

We conducted two fMRI demand-selection experiments using mental arithmetic and spatial reasoning tasks ((Experiment (Exp.) 1: n=30 (males: females = 17:13), Exp. 2: n=28 (males: females = 12:16)) (Figure 1A). Both experiments had identical task structures but differed in cognitive demands. 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. Choices transitioned probabilistically to either high or low effort cues (Figure 1A). These transition probabilities changed across 180 trials with five reversals (Figure 1B). Participants were instructed "the probability that each type of problem is more likely to appear is set independently for the left and right arrows. These probabilities change slowly over time. The same situation will continue for at least 10 trials". After the effort cue, a problem was presented. Participants were instructed to "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. We instructed that choices did not affect session duration.

For both experiments, we prepared two sets of problems that required different

levels of cognitive demand: 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 colored 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

In Exp. 1, 97% of participants, and in Exp. 2, 100% answered “yes” when asked whether the instruction was consistent with their scanner experience. These results suggest that most participants 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 participants 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 (0 = easiest, 6 = most difficult). 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 participants avoided high-demand problems but stated preferences and choice tendencies to avoid high-demand problems did not strongly relate.

Response time of arrow choices and task performance

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 sec 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 correctly answered high- and low-demand problems 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 4.11 ± 0.18 sec in Exp. 2. In both experiments, participants took longer for the high-demand problems (paired $t(29) = 11.02$, $p = 7.0 \times 10^{-12}$, in Exp. 1, $t(27) = 11.94$, $p = 2.8 \times 10^{-12}$, in Exp. 2). We set inclusion criteria (> 80% accuracy for all main experiment trials of both high- and low-demand problems) to ensure

engagement. By this criterion, we excluded one and two participants in Exp. 1 and Exp. 2 from further analyses, respectively.

Avoidance Behavior

We assumed that participants who tried to avoid high-demand problems would stay on the same side (left or right) after solving a low-demand problem in the previous trial and would switch to the opposite side after solving 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[11]. We found that 27 (93.1 %, out of 29) and 24 (92.3%, out of 26) participants in Exp. 1 and Exp. 2, respectively, significantly tended to avoid high-demand problems ($p < 0.01$), 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. Most participants avoided high-demand problems (Chi-square test, $p < 0.01$)[11].

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

We tested the effects of experiences two trials back for demand avoiders who significantly avoided high-demand problems (chi-square test). We categorized choices into four bins: problem type (high- or low-demand) in the (k–2)-th trial × problem type in the (k–1)-th

trial for trials in which the (k–1)-th choice was the same or opposite of the choice on the (k–2)-th trial. We found that the proportions of the k-th choice that were the same as (k–2)-th choice depended on both choice and experience history (Figure 2ABC). 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 (Figure 2Aa: $p = 1.7 \times 10^{-8}$, $d = -1.79$, $t(26) = -8.01$; Figure 1Ac: $p = 2.6 \times 10^{-2}$, $d = -0.74$, $t(22) = -2.40$; Figure 2Ad: $p = 4.7 \times 10^{-2}$, $d = -0.56$, $t(22) = -2.10$) and in Exp. 2 (Figure 2Ba: $p = 2.6 \times 10^{-8}$, $d = -1.49$, $t(23) = -8.24$; Figure 2Bc: $p = 4.1 \times 10^{-4}$, $d = -1.08$, $t(22) = -4.15$). Most participants significantly avoided high-demand problems, adapting based on experiences from two preceding trials in a changing environment. These results all replicate our previous findings[11] that humans avoid mental effort based on a demand history.

Demand avoiders reflected multiple past experiences to their current choices

Using Generalized Linear Mixed Model (GLMM) analyses, we tested whether one, two, and more than three past experiences affect the current choices of demand avoiders. The model included fixed effects of the choice and outcome from (k–1)-th to (k–4)-th trials, as well as the (k–1)-th choice. We assumed random effects to properly allocate variance to between- and within-subjects predictors. The random effects were the individual variations in intercepts, in the choice and outcome of both (k–1)-th and (k–2)-th trials,

and in (k–1)-th choice. Choices were significantly influenced by the choice and outcome of (k–1)-th, (k–2)-th, and (k–3)-th trials, and the choice at (k–1)-th trial significantly affected the current choices in both Exp. 1 and 2 (Figure 2D: logit coefficients 1.55, 0.72, 0.20, and 1.50, $p < 0.05$, in Exp. 1; Figure 2E: logit coefficients 1.73, 0.85, 0.30, and 1.58, $p < 0.05$, in Exp. 2). The choice and outcome at the (k–4)-th trial was also significant in Exp. 1 (Figure 2D: the logit coefficient was 0.23, $p < 0.05$), but not in Exp. 2 ($p = 0.156$). These results demonstrated that choices reflect multiple past experiences. More recent experiences had greater influence, suggesting a RL-like process.

Model-based analysis of choice data

We performed model fitting and comparison to determine which models best explain choice behavior in effort avoidance. Eleven models were tested, incorporating cost functions based on solution times, inaccuracy rates, rest time, and binarized demand levels (see Methods and Mitsuto et al., 2018 for full details)[11]. Nearly the same models were used as in our prior study, though we cut off the five worst-performing models. Bayesian Information Criterion (BIC) was used for model comparison. RL models assumed participants learned each factor in a simple RL update algorithm based on the prediction errors. We assumed Solution-Times (ST; continuous), Incorrect-Correct (IC; binary), and High-Low demand levels (HL; binary) as *ActualCost* in RL models. We also tested ST-IC and HL-IC combination models. Another model assumed rial-by-trial rest-

time (Rest) as a reinforcer and tested the combination between RL-HL and Rest. We also tested one and two-parameter probabilistic models (Win-Stay-Lose-Shift: pWSLS and Probabilistic-Selection: PS), assuming participants chose strategically based on HL or IC (HL and IC: binary). In the pWSLS-HL, a participant stays with their choice after a low-demand problem and switches after solving high-demand problems with the probability p. In the PS-HL model, the probabilities were a and b (the repeat rate after high and low demand, respectively).

Among RL models, the RL-HL model (binarized demand) had the best BIC scores for most participants in Exp. 1, and all the participants in Exp. 2 (25/27 (92.6%) in Exp. 1; 24/24 (100.0%) in Exp. 2; Figure 3Aa and Figure 3Ba, Table S1). The result that the RL-HL model outperformed the other RL models indicates that the cost of mental effort is learned as a binary variable, rather than variables reflecting the solution time or errors. The RL-HL model was also better than the two rest-time RL models (27/27 (100.0%) in Exp. 1; 21/24 (87.5%) in Exp. 2; Figure 3Ab and Figure 3Bb). Among probabilistic models, the PS model with binarized effort demands was superior (25/27 (92.6%) in Exp. 1; 23/24 (95.8%) in Exp. 2; Figure 3Ac and Figure 3Bc). Direct BIC comparisons showed that the PS model outperformed RL-HL (17/27 (63.0%) in Exp. 1; 14/24 (58.3%) in Exp. 2). Thus, both PS and RL models captured effort avoidance, but PS explained choices better for more participants.

228 Decisions in the PS model rely on one preceding trial, while RL models
229 incorporate multiple past trials (see Methods). According to BIC values, the PS model
230 explained choices better than the RL model in both our previous and current studies. On
231 the other hand, the RL model captured the sequential effect of past experienced demands
232 that the PS model was incapable of capturing. Previous simulations showed that RL, but
233 not PS, captured the effect of two-preceding trials (Figure 8 of Mitsuto et al.[11]). Thus, we
234 conducted model-based fMRI analysis using the RL-HL model. Figure 3C presents a
235 sample RL-HL model fitting. Figure 3D shows estimated CPE values of the same sample
236 data. Model-based fMRI analyses were conducted using CPE values from the RL-HL
237 model.

238 ***Lack of CPE encoding at effort cues***

239 We initially hypothesized that effort costs are learned by TD learning and expected costs
240 are updated at the time of effort cues. We thus predicted that CPE signals correlate
241 positively with dmFC/dACC, and negatively with striatal BOLD signal, irrespective of
242 demand type, at the time of effort cues (Figure 4Aa). To test this hypothesis, we examined
243 the activity of the whole brain to search for regions where BOLD signal at effort cue
244 correlated positively or negatively with the CPE signals in a conjunction analysis between
245 Exp. 1 and Exp. 2. The conjunction analysis involved a strict mask comprising common
246 voxels from Exp. 1 and 2, with the threshold of cluster-level FWE corrected $p < 0.05$, and

247 voxel-level uncorrected p < 0.001 from GLM 1 (Figure 4Ba). At the time of effort cues,
248 activity in neither the dmFC/dACC nor the striatum were correlated with CPE signals
249 significantly with the threshold of cluster-level FWE corrected $p < 0.05$, and voxel-level,
250 uncorrected $p < 0.001$ (Figure 5A, Figure S1ABGH, Table S2; the results of conjunction
251 analyses were reported with this threshold, unless otherwise stated). Given the
252 conservative nature of this test, we used a relaxed mask. To confirm the null results, we
253 analyzed the conjunction using a relaxed mask that consisted of common voxels between
254 the results of Exp. 1 and 2 with the threshold of voxel-level uncorrected p < 0.01 using
255 GLM 1. Under relaxed criteria, activity in the vmPFC/vACC was positively correlated
256 with the CPE signals at the time of effort cues (voxel-level uncorrected p < 0.001, Table
257 S2D), but there was no super-threshold correlation with activity in the dmFC/dACC.
258 These results thus fail to support the first hypothesis (Figure 4Aa).

259 In contrast to these null results, activity in the right inferior parietal sulcus (IPS)
260 correlates negatively with CPE signals at the time of effort cues, using the strict mask
261 (Figure 5A, Figure S1GH, Table S2). This finding is discussed later.

262 One possibility for the null results is that participants ignored effort cues. However,
263 conjunction analysis showed significant task preparation responses in dmFC/dACC, right
264 dlPFC, and right Angular Gyrus (AG) activity, which positively correlated with problem
265 demand (Figure 6). This result indicates that most demand avoiders encoded information

266 about upcoming effort demands at the time of effort cues, predicting the upcoming effort,
267 and prepared for the task accordingly. Therefore, these null results imply that the brain
268 does not encode CPE signals to update expected costs at the moment when full
269 information becomes available and, as such, mental effort cost learning does not follow
270 traditional TD learning.

271 ***CPE encoding at effort initiation***

272 We next hypothesized that expected cost updates require effort initiation (Figure 4Ab). A
273 conjunction analysis with the strict mask by using GLM 2 (Figure 4Bb) revealed that, at
274 the time of problem presentation when effort is initiated, the activity in the ventromedial
275 prefrontal cortex (vmPFC)/ ventral anterior cingulate cortex (vACC)/orbitofrontal cortex
276 (OFC) correlates positively with CPE signals (Figure 5B, Figure S1CDIJ, Table S3).
277 Although we detected a positive correlation between the CPE and the activity of the large
278 cluster including the dmFC/dACC in Exp. 1, the CPE was neither positively correlated
279 with the activity in the dmFC/dACC in Exp. 2 nor negatively correlated with the activity
280 in the striatum in either experiment. These results fail to support the second hypothesis
281 (Figure 4Ab). However, they suggest that the vmPFC/vACC/OFC encode CPEs, at the
282 time of effort initiation, with the information about upcoming effort levels.

283 ***CPE encoding at effort completion***

284 We next tested the hypothesis that expected costs are updated at effort completion (Figure

285 4Ac). A conjunction analysis with the strict mask (GLM 3; Figure 4Bc), revealed that
286 activity in the dmFC/dACC, the left anterior insula (aINS), and the left primary motor
287 cortex (M1)/ primary somatosensory cortex (S1)/IPS correlates positively, and activity in
288 the bilateral caudate nucleus (Cd) correlates negatively with CPEs at the time of effort
289 completion (Figure 5C, Figure S1EFKL, Table S4). Therefore, these results support the
290 third hypothesis and support the hypothesis that expected costs are updated at the time of
291 effort completion (Figure 5C).

292 ***Residual BOLD signals at effort completion predicted sequential choices***

293 If CPE encoding reflects an update to choice policy, then BOLD signal at the time of CPE
294 encoding should predict choice on the next trial. We tested the prediction that averaged
295 residual BOLD signal extracted from each cluster ROI that encoded CPEs anticipates
296 subsequent stay-switch choices, using mixed effects logistic regressions. For our analysis,
297 we first fit a new GLM, identical to the model we fit previously, but without a CPE
298 regressor (Figure 4B – without +*CPE*), and extracted the residual BOLD signal from all
299 ROIs defined by our prior conjunction analysis to correlate with CPEs at various time
300 points. Finally, we regressed binary stay-switch choices onto extracted residuals for each
301 ROI. In this analysis, a positive (negative) regression weight indicates that a participant
302 is more (less) likely to switch their choice on the next trial, reflecting an update in choice
303 policy. We find that residual BOLD signals in the Cd at effort completion predicted

304 sequential choices significantly in both Exp. 1 and 2 (Figure 7A). The residual BOLD in
305 the dmFC/dACC, left aINS, and left M1/S1/IPS also predicted the upcoming choices in
306 Exp. 2 (Figure 7AB). Neither BOLD signal residuals in the right IPS at effort cues nor
307 the vmPFC/vACC/OFC at effort initiation predicted upcoming choices (Figure 7A,
308 Figure S2AB), though all clusters' residual BOLD signals were correlated with CPEs
309 (Supplemental Result 1). These results imply that the neural signals at effort completion
310 reflect an update to choice policy while CPE-correlating activity either at effort cues or
311 at effort initiation does not.

312 **Discussion**

313 We first hypothesized that mental effort-cost learning follows TD learning and that cued
314 information about upcoming effort is sufficient to trigger an update of expected effort
315 associated with a given choice. We therefore predicted that, at the time of effort cues,
316 CPEs correlate positively with activity in the dmFC/dACC and negatively with activity
317 in the striatum. However, activity in neither the dmFC/dACC nor striatum correlate with
318 CPEs at the time of effort cues, which was revealed by a conjunction analysis across Exp.
319 1 and 2. This result failed to support the first hypothesis. Therefore, we hypothesized that
320 expected costs are updated either at effort initiation or completion. At effort initiation, the
321 activity of the vmPFC/vACC/OFC correlates positively with CPEs. At effort completion
322 the CPE correlates positively with the activity in the dmFC/dACC and other regions, and

correlates negatively with activity in the bilateral Cd. Additionally, we tested whether such correlates of CPE reflect an update of effort expectations, and thus choice policy. Thus, we tested the prediction that BOLD signal in regions that encode CPEs also predict choice on the next trial. We found that residual BOLD signals at effort completion predicted the sequential choices in both Exp. 1 and 2. These results demonstrate that mental effort cost learning requires effort exertion and imply that adaptive learning of mental effort costs does not follow traditional TD learning.

Consideration of results and limitations

Multiple Past Demanding Experiences Affect Current Choices. Behavioral data showed that humans tend to avoid mental effort, reflecting the experiences of multiple past events in the task design of the current study with the addition of effort cues, replicating our previous study[11]. The choice data of demand avoiders were approximated with a RL model based on demand levels, implying that humans track mental effort cost information. Questionnaire results suggested that most participants prefer low-demand problems but self-reported demand preferences have a weak relationship with behavioral choice tendencies.

Lack of Significant Results at Effort Cues. 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 when effort information is first available contradicts

342 the prediction from the TD hypothesis (Figure 4A and 5A). Furthermore, residual BOLD
343 in the right IPS at effort cue did not predict the sequential choices, and the residual BOLD
344 signals in right IPS did not show a consistent activity pattern with respect to CPEs
345 between two experiments (Figure S2A). We speculate that the negative correlation
346 between right IPS activity and CPE reflects a network reset signal in the frontoparietal
347 network at effort cue timing[32]. Consistent with the network reset hypothesis, increased
348 IPS activity in response to a negative CPE (indicating an easier-than-expected task) may
349 reflect disengagement from a high-attentional state. Collectively, our results are
350 inconsistent with the hypothesis of traditional TD learning whereby mental effort cost
351 expectations are updated when effort information first becomes available. We
352 acknowledge the possibility that our methods are insufficiently sensitive to detect CPE
353 encoding in the dmFC/dACC and the striatum at informative effort cues. Indeed, the
354 absence of evidence is not the evidence of absence. However, our methods were
355 sufficiently sensitive to detect CPEs at effort completion, suggesting at the very least that
356 CPE encoding is more robust at effort completion.
357 *vmPFC/vACC/OFC Activity at Effort Initiation.* Activity patterns around the rostromedial
358 prefrontal cortex were consistent across two experiments in our previous study that did
359 not use effort cues[11] and four experiments in total. However, the positive correlation
360 between CPEs and activity in vmPFC/vACC/OFC at effort initiation contradicts the

361 second hypothesis that expected costs are updated at the time of effort initiation (Figure
362 4Ab and 5B). These results stand in contrast with the general perspective in the field of
363 value-based decision-making that the vmPFC positively encodes predicted reward value,
364 or reward PEs, discounted by effort costs[31,33-35]. A positive correlation with costs thus
365 argues against the prediction that the vmPFC activity at effort initiation reflects value
366 encoding or value updating. Furthermore, residual BOLD signals of the
367 vmPFC/vACC/OFC at effort initiation do not predict the subsequent choices in either
368 experiment, implying that although BOLD signal correlates with CPEs, this correlation
369 does not reflect an update of expected costs that subsequently guides value-based
370 decision-making.

371 Activity in the vmPFC/vACC/OFC also correlated with CPEs at effort initiation,
372 which we speculate reflects a shift to a task engaged state. More generally, the OFC is
373 hypothesized to track current state for elaborating progress through a model of the
374 external world, cognitive map[36,37,38,39]. The OFC may track current states and update a
375 cognitive map such as state-space map and task set during effort initiation, supporting
376 transitions into task-engaged states and credit assignment for actions directed toward task
377 goals. Perhaps the mental action, triggered by effort initiation, recruits the OFC to update
378 a world model, reflecting a transition into task-engaged states.

379 *Sequential Choice Predicted by Residual BOLD Signal at Effort Completion.* When

380 participants complete a task, the associated CPE correlates positively with the activity in
381 the dmFC/dACC, left aINS, and left M1/S1/IPS and negatively with activity in the
382 bilateral Cd, irrespective of demand type (Figure 4Ac and 5C). Furthermore, residual
383 BOLD signals predicted the upcoming choices in the Cd, in both experiments, as well as
384 the residual BOLD signals at the dmFC/dACC, left aINS, and left M1/S1/IPS clusters in
385 Experiment 2 (Figure 7AB, Figure S2C). These results demonstrate that the brains encode
386 the CPE signals at effort completion, not at other timepoints, and moreover that this CPE
387 encoding at effort completion reflects an update of expected values and choice policy.

388 Potential confounds, such as retrospective confidence PE or satisfaction PE,
389 were considered. However, confidence-related activations are typically observed in
390 regions like the frontopolar cortex or ventrolateral PFC, not in the Cd[40-42]. Similarly,
391 satisfaction as a confound is unlikely since it serves as a positive reinforcer. Because
392 participants avoided higher demands, they did not behave as though they found greater
393 satisfaction when completing higher demands.

394 *Mental Effort Learning Algorithm.* We speculate that expected cost updates are delayed
395 until task completion because people can adjust effort on-the-fly to handle unexpected
396 difficulties or unexpected efficacies. With practice and insight, some tasks may become
397 easier, so the brain may wait to evaluate mental effort retrospectively, after completion,
398 to access full information about the effort that was just exerted. Waiting until effort

399 completion also avoids multi-tasking between the task itself and the evaluation of effort
400 costs.

401 Our data imply that mental effort costs are learned in cortico-basal ganglia loops
402 encompassing the dmFC/dACC and Cd, being consistent with known circuit anatomy[27].
403 This hypothesis is supported by evidence of positive encoding of effort costs and effort
404 preparation in the dmFC/dACC and the negative correlation of the CPE with activity in
405 the Cd, detected at effort completion.

406 Traditional TD learning theory failed to explain our results. Recently, the
407 retrospective causal learning algorithm was proposed to account for midbrain dopamine
408 signaling, during reward learning, better than classic TD learning[43]. While our results
409 indicate that mental effort learning entails retrospective inference, the retrospective causal
410 learning algorithm does not explain why CPEs are encoded at effort completion. As with
411 TD learning, retrospective causal learning also predicts a CPE when information is first
412 available. Therefore, we propose that the algorithm governing mental effort cost learning
413 involves retrospectively Bayesian or non-TD RL[1,44], updating expected effort costs as a
414 function of the difference between prior expectations, and experienced effort costs,
415 reported as a CPE, when exertion is completed.

416 ***Future Directions and Perspectives***

417 Refining the effort cost learning algorithm is essential. Future studies should simulate and

test multiple candidates learning algorithms, based on the observation that updates require direct experience, to develop new experimental paradigms and determine which algorithm best explains learning.

Conclusion

Our data indicate that mental effort cost learning is retrospective and 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. Furthermore, we find that the residual BOLD signals at effort completion predicted subsequent choices, indicating that the CPE encoding reflects an update of choice policies at the time of effort completion. Thus, effort exertion is required for humans to learn the costs of effortful tasks. These findings refine our understanding of the mechanisms underlying decision-making and learning based on effort with implications for adaptive learning of mental effort in health and disease.

Supplemental Result 1. Residual BOLD signal of each cluster region of interest (ROIs) explained CPEs. To confirm the trial-wise relationship between each averaged residual BOLD signal of each cluster ROIs at each timing and CPE signals, a Linear Mixed Model (LMM) analysis was conducted. The cluster ROIs at each time point were as follows: the right IPS at effort cue, the vmPFC/vACC/OFC at effort initiation, and the dmFC/dACC, left aINS, left M1/S1/IPS, and bilateral Cd at effort completion. Residual BOLD signals within each ROI significantly explained CPE signals across trials.

Author Contributions

1. Conceptualization

AM

2. Data curation

AM

3. Formal analysis

AM

4. Funding acquisition

SY, AW, and AM

5. Investigation

AM, TK, and KO

6. Methodology

AM, AW

7. Project administration

AM, SY, and RH

8. Resources

AM

9. Software

AM

10. Supervision

AM, SY, RH, and AW

11. Validation

AM, AW, KM, TY, and KO

12. Visualization

AM

13. Writing – original draft

AM

14. Writing – review & editing

AM, RA, KM, TK, RH, and AW

Declaration of interests

The authors declare no competing interests.

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Asako Mitsuto (asako.3210@rutgers.edu).

Disclosure

During the preparation of this work the authors used ChatGPT (OpenAI. ChatGPT, Version 3.5 and 4. Available at https://openai.com/) to check grammar, spelling, and logical flow of drafts and create a part of GLMM analysis scripts in R Studio. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Acknowledgements

This work was supported by Grant-in-Aid of JSPS for Research Activity Start-up No. 18H06090 [to AM], for JSPS Fellows No. 19J00964 [to AM], for Early-Career Scientists No. 20K16475 and No. 23K14680 [to AM], Challenging Research (Exploratory) No. 19K21809 [to AM], and for Scientific Research on Innovative Areas No. 15H05876 [to AM], the Impulsing Paradigm Change through disruptive Technologies (ImPACT) program in Japan [to SY], and Grant No. R00 MH125021 of NIMH [to AW]. We are grateful to K Sakai, W Zajkowski (NIH), DL Aguilar (CBS), RP Badman (Harvard Univ),

K Miyamoto (CBS), M Cai (Univ Miami), A Funamizu (Univ Tokyo), T Toyoizumi (CBS), Y Morishima (Bern Univ), K Tamura (Univ Cambridge), H Tomoki (ATR), A Nagai (Shimane Univ), T Shimizu (Tottori Univ) and the members of the Neurology Division of Tottori University, Akaishi Lab of RIKEN CBS, Doya Lab of OIST, and Rushworth Lab of the University of Oxford for the advice, support, discussion, or feedback to this study.

492 **Methods**

493 ***Participants***

494 We recruited 30 participants (13 females; mean age, 23.0 ± 2.2) in Experiment (Exp.) 1
495 and 28 participants (16 females; mean age, 22.6 ± 1.6) in Exp. 2. Among these participants,
496 11 participants took part in both Exp. 1 and Exp. 2. We paid all participants 5,000 yen by
497 bank transfer after their participation. No participants were taking any medicine or had a
498 prior history of neuropsychiatric disorders. All participants were students of Shimane
499 University, age 20-30 years, right-handed, native Japanese speakers, and received their
500 compulsory education in Japan. Informed written consent was obtained from all
501 participants before the experiment. The present study was approved by the medical ethics
502 committee of Shimane University.

503 ***Behavioral tasks***

504 We conducted two experiments, Exp. 1 and Exp. 2 (Figure 1). Except for changes I~VII
505 as we described below, these experiments were almost similar behavioral protocols with
506 our previous work [11]. Consequently, we summarized the methods and listed changes I~VII
507 from our preceding procedure (please see Mitsuto *et al.*, 2018 for full details). The most
508 important change was to add the effort cues that indicated the upcoming effort levels (i.e.,
509 high- or low-demand problems) between choices and problem presentations (please see
510 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 [1]. 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. Second, the addition of effort cues let us 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.

We provided a summary of the experimental paradigms. 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-

530 folding) problems to judge whether a 3D cube with three visible colored faces and a
531 concurrently presented unfolded cube matched or not in Exp. 2. The second point is that
532 we used different features of effort cues that indicated upcoming effort levels; circle cues
533 with red and green colors in Exp. 1 and upward and downward triangle cues in Exp. 2.

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

541 The flow of one trial in both experiments was described below (Figure 1A). There
542 were 180 trials. At the start of each trial, participants choose either the left or right arrow.
543 After the choice, an effort cue was presented. The effort cue was 100% associated with
544 either low- or high-demand problems. After the effort cue, a problem was presented (high-
545 demand or low-demand problem). The arrow determined the presentation probability of
546 either low- or high-effort cues. The probabilities of arrows associated with upcoming
547 effort cues, or effort levels, were alternated dynamically over time (Figure 1B). The set
548 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 were instructed that their choices did not affect how fast they could finish the experiment.

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 outlined 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 color blind, we planned to use blue instead of green but there were no color blindness participants. The color combinations between upcoming difficulties and colors were counterbalanced across participants. We used upward and downward triangle cues with grey color 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 mapping of arrows to demand levels were counterbalanced across, and stable in, both Exp. 1 and 2. After practice block(s), we asked participants how red and green problems in Exp. 1 and how upward and downward triangle problems differed to prevent misunderstanding the difference between the two kinds of problems (low- and high-demand problems). Notably, we did not explain the difference. If participants did not reply that the problems were low- versus high-demand, we responded that something else differentiated the problem types.

IV: Participants completed180 total trials in the scanner (the same number 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 [45].

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.

Participants then responded yes or no by circling their answer.

If the subject answered No, the participants 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 or 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.

Behavioral Analysis

Basic Analysis. Behavioral analyses were the same protocols as our previous work (please see Mitsuto et al., 2018), except for the analyses of Generalized Linear Mixed Model (GLMM). 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.

Generalized Linear Mixed Model. We conducted GLMM analyses on the binominal choice data to test whether or not multiple past events affect current choices for demand avoiders who showed the avoidance behavior significantly in the aforementioned χ^2 test. We use lmer4 package of R Studio for the GLMM analyses with maximum likelihood estimation methods. The response variable was coded as 1 for left choices and 0 for right choices. The fixed effects were an intercept, influences of choices and outcomes from both (k–1)-th to (k–4)-th trials' choices on the k-th choices, and influences of choice repetitions. Specifically, the choice and outcomes were valued as 0 for both low-demand problems after left choices and high-demand problems after right choices and 1 for both high-demand problems after right choices and low-demand problems after left choices. The choice repetitions were valued as 0 for the opposite choices with the previous trials

and 1 for the same choices as the previous trials. We used the z-score normalized values of them as the fixed effects. Additionally, the random effects were individual variations in the intercept, in the choices and outcomes of both (k–1)-th and (k–2)-th choices, and in choice repetitions.

The GLMM formulation was as follows:

$$Choice(k) \sim \beta 0 + \beta 1 \times ChoiceOutcome(k-1) + \beta 2 \times ChoiceOutcome(k-2)$$
$$+ \beta 3 \times ChoiceOutcome(k-3) + \beta 4 \times ChoiceOutcome(k-4)$$
$$+ \beta 5 \times Choice(k-1) + (\beta 0i + \beta 1i + \beta 2i + \beta 5i \mid SubjectID)$$

where β_0 is the constant term, β_1 to β_4 are the effects of choice and outcomes from (k–1)-th to (k–4)-th trials on the choice at k-th trial, β_5 is the effect of the choice at (k–1)-th trial, and β_{0i}, β_{1i}, β_{2i}, and β_{5i} are the individual variations in these effects since individual differences in β1, β2, and β5 explained the most individual differences variance. Here, choices were binary ([Left, Right] = [0, 1]). The choice and outcome from (k–1)-th to (k–4)-th trials ([Left_Low, Left_High, Right_Low, Right_High] = [0, 1, 1, 0]) and the choice at (k–1)-th trial ([Left, Right] = [0, 1]) were normalized in z-scores, respectively.

Modelling, Model Fitting, and Comparisons. The models that we used were the almost same set as our previous work (please see Mitsuto et al., 2018), except for the following change. We cut off the five worst-performing models, for which the expected cost for the unchosen option on the first trial was an additional free parameter. Concerning the

682 following five reinforcement learning (RL) models, *ExpectedCost* for the unchosen
683 option at the first trial [$ExpectedCost_{unchosen}(1)$] was assumed to be equal to
684 $ExpectedCost_{Chosen}(1)$.
685 (1) solution time in individual trials (RL-Solution-Time-(ST) model)
686 (2) demand level of the problem (RL-High-Low (HL) model)
687 (3) incorrect solving (RL-Incorrect-Correct (IC) model)
688 (4) sum of (1) and (3), with a weighting parameter for (3) ($w_{incorrect}$) (RL-ST-IC model)
689 (5) sum of (2) and (3), with a weighting parameter for (3) ($w_{incorrect}$) (RL-HL-IC model)
690 *ExpectedCost* for the unchosen and chosen options were assumed to be equal because
691 means of the Bayesian information criterion (BIC) of the models where
692 $ExpectedCost_{unchosen}(1)$ was set to be a free parameter were worse than the BIC means of
693 the above-mentioned models that had the same initial values in our previous results and
694 the presentation probabilities of which effort levels came after choices were 0.5 in practice
695 sessions (please see the change III). Finally, we tested the following eleven models that
696 were used in our previous work: a RL model based on solution time (RL-ST model), a
697 low/high-demand model (RL-HL), a model based on incorrect responses (RL-IC),a model
698 that combines solution time and incorrect responses ST + IC (RL-ST-IC), a model that
699 combines low/high demands and incorrect choices (RL-HL-IC), a rest-time RL model
700 (RL-Rest) and a rest-time model that also has information about low/high demands (RL-

Rest-HL). Rest time model used a rest-time after the answer as a reinforcer in the reinforcement learning model. 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 behavior was explained better by trial history. We used several probability models (WSLS and PS) because WSLS mode was widely-used models of choice behavior and PS model captured qualitative patterns in our 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 [11]. 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.

The PS-HL model was a PS 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.

In sum, we analyzed: 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 behavioral 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 fit general linear models (GLM) to our 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 (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 expected costs are updated at effort cues, by including a CPE regressor at the time of effort cues. Next, we designed GLM2 and GLM3 with the CPE regressor at the time of problem presentation or the when participants answered 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

773 of arrow-cue presentation with parametric modulations by $ExpectedCost_{chosen}$ and RT_{choice},
774 the regressors at the time of effort cue presentation with parametric modulations by
775 problem-demand (0 and 1 for low- and high-demand problems, respectively) and CPE
776 (this parametric modulation of CPE is only for GLM1), the regressor at the time of
777 problem presentation with parametric modulation by CPE (this regressor with parametric
778 modulation of CPE is only for GLM2), the regressors with the duration from problem
779 presentation to the answer of problems with parametric modulations by problem-demand,
780 the regressors at time of the answer of problems with parametric modulation by CPE (this
781 regressor with parametric modulation of CPE is only for GLM3), the regressors for motor
782 response at the time of the choice of an arrow and the answer of problems (GLM1 and
783 GLM2 included them both timing and GLM3 included them only at the time of the choice
784 of an arrow) and head movements. Each regressor in each GLM was convolved with the
785 canonical hemodynamic response function, and one-sample t-tests were performed on
786 individual maps of the regressors of interest for 25 participants in Exp. 1 and 23
787 participants in Exp. 2. The variance inflation factor (VIF) was calculated to determine
788 whether the collinearity of the regressors of interest was within an acceptable range where
789 the cut-off value for the VIF used in collinearity issues was 5 or 10.

790 For GLM 1-3, the variance inflation factors (VIF) of CPE parametric regressors
791 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 was the same as our previous work (please see Mitsuto *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 $p < 0.001$.

Residual Blood Oxygenation Level Dependent (BOLD) analyses. To test whether residual BOLD signals predict choice on each subsequent trial, we fit a version of our GLM1 above, but without a CPE regressor, and extracted residuals from each of the CPE-correlated cluster's for each timing: at effort cue, effort initiation, and effort completion. CPE-correlated cluster regions of interest (ROIs) were the significant cluster of strict masked conjunction analyses with a threshold of FWE corrected $p < 0.05$ and voxel-level uncorrected $p < 0.001$: the right IPS from at effort cue, and the vmPFC/vACC/OFC from at effort initiation and the dmFC/dACC, left aINS, left M1/S1/IPS, and bilateral Cd from at effort completion.

We calculated the average residual BOLD signal across all voxels within each cluster at each time point, starting from time zero as the initial time point, for effort cue, effort initiation, and effort completion, up to 17.5 seconds. Subsequently, we averaged the residual BOLD signals between 5.0 and 7.5 seconds for each cluster, targeting the peak of canonical hemodynamic response function.

Next, we fit a generalized linear mixed effects model (with logit link function), regressing choice onto residual BOLD signal as follows:

$$Choice(k+1) \sim \beta 0 + \beta 1 \times ResidualBOLD(k) + (\beta 0i \mid SubjectID)$$

where β_0 is the constant term, β_1 is the effects of residual BOLD signals at k-th trial, and β_{0i} is the individual variation in the constant term effects. Here, choices at (k + 1)-th trials were binary ([Stay, Switch] = [0, 1]). The residual BOLD signals at the k-th trial were normalized in two steps: first, within each block using z-scores, and second, across all blocks for each subject using z-scores.

As a sanity check, we also sought to confirm the relationship between trial-wise CPEs and trial-wise residual BOLD signals for each cluster at the time point, by performing Linear Mixed Model (LMM) analyses with the LMM was formulated as follows:

$$CPE(k) \sim \beta 0 + \beta 1 \times ResidualBOLD(k) + (\beta 0i + ResidualBOLD(k) \mid SubjectID)$$

where β_0 was the constant term, β_1 was the effects of residual BOLD signals at k-th trial,

and β_{0i} and β_{1i} were the individual variations in these effects. The residual BOLD signals at the k-th trial were normalized using the same two-step process described above, consistent with the normalization used in the GLMM.

Our mixed effects model analyses were conducted using MATLAB (version 2019b) with the Statistics and Machine Learning Toolbox, specifically using the *fitglme* and *fitlme* functions for GLMM and LMM, respectively.

Regarding missing data, trials with missing BOLD signals due to scan durations exceeding ten minutes per block were excluded from these fMRI analyses.

Resource Availability

Upon request, the codes, the behavior data, and the preprocessed MRI data of this study are available from the corresponding author.

References

1 Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction second edition Introduction*. (2018).

2 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). https://doi.org:10.1016/j.neuron.2013.07.007

3 Glimcher, P. W. & Fehr, E. in *Neuroeconomics (Second Edition)* (eds Paul W. Glimcher & Ernst Fehr) xvii-xxviii (Academic Press, 2014).

4 Westbrook, A., Frank, M. J. & Cools, R. A mosaic of cost–benefit control over cortico-striatal circuitry. *Trends in Cognitive Sciences* (2021). https://doi.org:10.1016/j.tics.2021.04.007

5 Walton, M. E. & Bouret, S. What Is the Relationship between Dopamine and Effort? *Trends in Neurosciences* **42**, 79-91 (2019). https://doi.org:10.1016/j.tins.2018.10.001

6 Westbrook, A. *et al.* Dopamine promotes cognitive effort by biasing the benefits versus costs of cognitive work. *Science* **367**, 1362-+ (2020). https://doi.org:10.1126/science.aaz5891

7 Doya, K. Modulators of decision making. *Nature Neuroscience* **11**, 410-416 (2008). https://doi.org:10.1038/nn2077

8 Wallis, J. D. & Rushworth, M. F. S. in *Neuroeconomics (Second Edition)* (eds Paul W. Glimcher & Ernst Fehr) 411-433 (Academic Press, 2014).

9 Kool, W., McGuire, J. T., Rosen, Z. B. & Botvinick, M. M. Decision Making and the Avoidance of Cognitive Demand. *Journal of Experimental Psychology-General* **139**, 665-682 (2010). https://doi.org:10.1037/a0020198

10 Westbrook, A., Kester, D. & Braver, T. S. What Is the Subjective Cost of Cognitive Effort? Load, Trait, and Aging Effects Revealed by Economic Preference. *Plos One* **8** (2013). https://doi.org:10.1371/journal.pone.0068210

11 Nagase, Mitsuto., A. *et al.* Neural Mechanisms for Adaptive Learned Avoidance of Mental Effort. *Journal of Neuroscience* **38**, 2631-2651 (2018). https://doi.org:10.1523/jneurosci.1995-17.2018

12 Sayali, C. & Badre, D. Neural systems of cognitive demand avoidance. *Neuropsychologia* **123**, 41-54 (2019). https://doi.org:10.1016/j.neuropsychologia.2018.06.016

13 Sayalı, C. & Badre, D. Neural systems underlying the learning of cognitive effort costs. *Cogn Affect Behav Neurosci* (2021). https://doi.org:10.3758/s13415-021-00893-x

14 Cavanagh, J. F., Masters, S. E., Bath, K. & Frank, M. J. Conflict acts as an implicit cost in reinforcement learning. *Nature Communications* **5** (2014). https://doi.org:10.1038/ncomms6394

15 Kool, W. & Botvinick, M. Mental labour. *Nature Human Behaviour* **2**, 899-908 (2018). https://doi.org:10.1038/s41562-018-0401-9

16 Matsumoto, M. & Hikosaka, O. Representation of negative motivational value in the primate lateral habenula. *Nature Neuroscience* **12**, 77-84 (2009). https://doi.org:10.1038/nn.2233

17 Seymour, B. *et al.* Temporal difference models describe higher-order learning in humans. *Nature* **429**, 664-667 (2004). https://doi.org:10.1038/nature02581

18 Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275**, 1593-1599 (1997). https://doi.org:10.1126/science.275.5306.1593

19 Patrick, L. M., Anderson, K. M. & Holmes, A. J. Local and distributed cortical markers of effort expenditure during sustained goal pursuit. *Neuroimage* **244** (2021). https://doi.org:10.1016/j.neuroimage.2021.118602

20 Palidis, D. J. & Gribble, P. L. EEG correlates of physical effort and reward processing during reinforcement learning. *Journal of Neurophysiology* **124**, 610-622 (2020). https://doi.org:10.1152/jn.00370.2020

21 Suzuki, S., Lawlor, V. M., Cooper, J. A., Arulpragasam, A. R. & Treadway, M. T. Distinct regions of the striatum underlying effort, movement initiation and effort discounting. *Nat Hum Behav* **5**, 378-388 (2021). https://doi.org:10.1038/s41562-020-00972-y

22 Hernandez Lallement, J. *et al.* Effort increases sensitivity to reward and loss magnitude in the human brain. *Social Cognitive and Affective Neuroscience* **9**, 342-349 (2013). https://doi.org:10.1093/scan/nss147

23 Botvinick, M. M., Huffstetler, S. & McGuire, J. T. Effort discounting in human nucleus accumbens. *Cognitive Affective & Behavioral Neuroscience* **9**, 16-27 (2009). https://doi.org:10.3758/cabn.9.1.16

24 Hauser, T. U., Eldar, E. & Dolan, R. J. Separate mesocortical and mesolimbic pathways encode effort and reward learning signals. *Proceedings of the National Academy of Sciences of the United States of America* **114**, E7395-E7404 (2017). https://doi.org:10.1073/pnas.1705643114

25 Matsumoto, M., Matsumoto, K., Abe, H. & Tanaka, K. Medial prefrontal cell activity signaling prediction errors of action values. *Nature Neuroscience* **10**, 647-656 (2007). https://doi.org:10.1038/nn1890

26 Rushworth, M. F. S. & Behrens, T. E. J. Choice, uncertainty and value in prefrontal and cingulate cortex. *Nature Neuroscience* **11**, 389-397 (2008). https://doi.org:10.1038/nn2066

27 Alexander, G. E., Delong, M. R. & Strick, P. L. PARALLEL ORGANIZATION OF FUNCTIONALLY SEGREGATED CIRCUITS LINKING BASAL GANGLIA AND CORTEX. *Annual Review of Neuroscience* **9**, 357-381 (1986). https://doi.org:10.1146/annurev.ne.09.030186.002041

28 Reynolds, J. N. J. & Wickens, J. R. Dopamine-dependent plasticity of corticostriatal synapses. *Neural Networks* **15**, 507-521 (2002). https://doi.org:10.1016/s0893-6080(02)00045-x

29 Seymour, B., Daw, N., Dayan, P., Singer, T. & Dolan, R. Differential encoding of losses and gains in the human striatum. *Journal of Neuroscience* **27**, 4826-4831 (2007). https://doi.org:10.1523/jneurosci.0400-07.2007

30 McClure, S. M., Berns, G. S. & Montague, P. R. Temporal prediction errors in a passive learning task activate human striatum. *Neuron* **38**, 339-346 (2003). https://doi.org:10.1016/s0896-6273(03)00154-5

31 Lopez-Gamundi, P. *et al.* The neural basis of effort valuation: A meta-analysis of functional magnetic resonance imaging studies. *Neuroscience & Biobehavioral Reviews* (2021). https://doi.org:10.1016/j.neubiorev.2021.10.024

32 Bouret, S. & Sara, S. J. Network reset: a simplified overarching theory of locus coeruleus noradrenaline function. *Trends in Neurosciences* **28**, 574-582 (2005). https://doi.org:10.1016/j.tins.2005.09.002

33 Hauser, T. U. *et al.* Temporally Dissociable Contributions of Human Medial Prefrontal Subregions to Reward-Guided Learning. *Journal of Neuroscience* **35**, 11209-11220 (2015). https://doi.org:10.1523/jneurosci.0560-15.2015

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

35 Klein-Flügge, M. C., Bongioanni, A. & Rushworth, M. F. S. Medial and orbital frontal cortex in decision-making and flexible behavior. *Neuron* (2022). https://doi.org:10.1016/j.neuron.2022.05.022

36 Behrens, T. E. J. *et al.* What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. *Neuron* **100**, 490-509 (2018). https://doi.org:10.1016/j.neuron.2018.10.002

37 Tolman, E. C. Cognitive maps in rats and men. *Psychol Rev* **55**, 189-208 (1948). https://doi.org:10.1037/h0061626

38 Sharpe, M. J. *et al.* An Integrated Model of Action Selection: Distinct Modes of Cortical Control of Striatal Decision Making. *Annu Rev Psychol* **70**, 53-76 (2019). https://doi.org:10.1146/annurev-psych-010418-102824

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

40 Miyamoto, K. *et al.* Identification and disruption of a neural mechanism for accumulating prospective metacognitive information prior to decision-making. *Neuron* (2021). https://doi.org:10.1016/j.neuron.2021.02.024

41 Miyamoto, K. *et al.* Causal neural network of metamemory for retrospection in primates. *Science* **355**, 188-193 (2017). https://doi.org:10.1126/science.aal0162

42 Bang, D. & Fleming, S. M. Distinct encoding of decision confidence in human medial prefrontal cortex. *Proceedings of the National Academy of Sciences* **115**, 6082-6087 (2018). https://doi.org:10.1073/pnas.1800795115

43 Jeong, H. *et al.* Mesolimbic dopamine release conveys causal associations. *Science* **378**, eabq6740 (2022). https://doi.org:10.1126/science.abq6740

44 Ma, W. J. & Jazayeri, M. in *Annual Review of Neuroscience, Vol 37* Vol. 37 *Annual Review of Neuroscience* (ed S. E. Hyman) 205-220 (2014).

45 Okubo, M., Suzuki, H. & Nicholls, M. E. R. A Japanese version of the FLANDERS handedness questionnaire. *The Japanese journal of psychology* **85**, 474-481 (2014). https://doi.org:10.4992/jjpsy.85.13235

Figure legends

Figure 1. The behavioral 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 (low- or high-demand) problem was presented. 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 indicated upcoming effort levels exactly, 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. Demand avoiders made a choice, learning based on multiple past experiences.

Behavioral analyses for the choices of demand avoiders in Exp. 1 (n = 27) and Exp. 2 (n = 24). (AB) The influences of the two preceding trials on current choices. Each pair of

bars indicates the across-participants mean rates that the current (k-th) choice was the same as the choice at 2 trials before ((k–2)-th) when the experienced demand at the (k–2)-th trial was high (left bar) or low (right bar), for each case sorted by the choice at the (k–1)-th trial (same as (a, b) or opposite from (c, d) the (k–2)-th choice) and the experienced demand at the (k–1)-th trial (high (a, c) or low (b, d)). Each bar has a label like HsH, following the rule explained in (C). The average was taken across the participants who were judged as demand avoiders as a result of Chi-square test (see Result section). The error bars indicate the mean ± standard error of the mean (SEM). The paired data of individual participants were shown as dots and were connected by lines. Unpaired data were shown as crosses without lines. The paired data were compared by paired t-test, and asterisks marked significant differences. (C) Bar label explanation. The HsH is an example of bar labels on Figure 2A and 2B. (D,E) The influences of multiple past experiences on the current choices, analyzed using Generalized Linear Mixed Model (GLMM). Bars indicate the fixed effects and the SEMs. The fixed effects were an intercept, influences of the choice and outcome from both the (k–1)-th to the (k–4)-th trials and the choice at (k–1)-th trial on the choice at k-th trial. Random effects were the individual variations in intercepts, in the choice and outcome of both (k–1)-th and (k–2)-th trials, and in the choice at (k–1)-th trial. Asterisks mark the significant t-test at $p < .05$.

Figure 3. Model comparisons, examples of model fitting and the estimated cost prediction error values.

(A,B) Model Comparison. The bars indicate the mean ± SEM of the Bayesian Information Criterion (BIC) scores. Results for 5 models of reinforcement learning (RL) based on ActualCost (Solution-Time (ST), High-Low (HL), Incorrect-Correct (IC), ST-IC, and HL-IC). Individual demand avoiders are represented by the black dots linked with lines. (Ab, Bb) Results for the rest-time RL models which regarded the rest time in the scanner as reward for participants. (Ac, Bc) Results for probabilistic Win-Stay-Lose-Shift (pWSLS) models and Probabilistic-Selection (PS) models based on HL and IC. (B) Illustration of choices and choice probability predicted by the reinforcement learning model based on High-Low demands (RL-HL model). Short vertical bars at the top and bottom indicate participant's left and right choices, respectively, with dark- or light-gray indicating that high- or low-demand problems were experienced, respectively. The black dashed line in the middle indicates the Left – Right difference in the presentation rates (moving average of latest 5 trials) of low-demand problems plotted against the left scale. The blue and red solid lines indicate the participant’s actual left-choice rate (moving average of latest 5 trials) and the left-choice probability predicted by the RL-HL model plotted against the blue and red scales on the right, respectively. (C) Illustration of estimated values of cost prediction error of the same participant with Figure 3B.

Figure 4. Hypotheses and General Linear Models for fMRI Analyses.

(Aa) Hypothesis 1 is that effort cost learning follows temporal-difference learning such that expected costs are updated at the time of the effort cues. If participants predict the upcoming effort demand is high-demand at 80% and they receive the information of low demand at the effort cue presentation, the prediction error (δ) of -80% generates at the effort cue timing. Expected cost and the cost prediction error (CPE) as δ during a trial are shown schematically. Lines are colored for high-demand trials with the gradation from red to pale pink and for low-demand with the gradation from pale blue to blue, representing expected effort levels at the arrow-cue from high- to low-demand effort levels. To simplify, we used a monotonic linear decrease function as mental effort cost function during problem solution. (Ab) Hypothesis 2 is that expected costs are updated at effort initiation. (Ac) Hypothesis 3 is that expected costs are updated at effort completion. (B) 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 CPE estimated in the reinforcement learning model based on effort levels by using three GLMs (GLM1-3), according to three different hypotheses regarding the time of CPE generation. All GLMs included regressors at the presentation of arrow cues with non-orthogonalized parametric modulations by ExpectedCostchosen and RTchoice, 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 solution-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 5. Neural correlates of the cost prediction error at the time of effort cue, at effort initiation, and at effort completion.

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 (A: GLM1), problem presentation (B: GLM2) or answer (C: GLM3) 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 inferior parietal sulcus (IPS) was detected as a negative correlation. (B) The vmPFC/vACC/orbitofrontal cortex (OFC) 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 6. Neural correlates of the task preparation at presentation of an 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 (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 angular gyrus (AG), dmFC/dACC were detected as a positive correlation.

Figure 7. Cluster ROI-based residual BOLD signal analyses: Prediction of sequential choices.

(A) The influence of residual BOLD signals from each cluster region of interest (ROI) for each time point (effort cue: green bars, effort initiation: blue bars, and effort completion: red bars) on subsequent choices, analyzed using a Generalized Linear Mixed Model (GLMM). Fixed effects included the residual BOLD signal at the k-th trial to predict subsequent choice behavior (switch or stay). Random effects accounted for individual variation in the intercepts. Asterisks mark significant t-tests at $p \leq 0.05$. Each p-value was shown. Lighter bars indicate results from Experiment 1, while darker bars indicate results from Experiment 2. Positive estimates suggest a higher residual BOLD

signal is associated with an increased likelihood of switching choices in subsequent trials, while negative estimates indicate an increased likelihood of staying. (B) Time-course analyses of residual BOLD signals at effort completion (Experiment 2) for selected ROIs: the dmFC/dACC, left aINS, left M1/S1/IPS, and bilateral Cd. Curves represent residual BOLD signals averaged by CPE tertiles: CPE Low ($-1 \leq$ CPE $<$ 33rd percentile, blue line), CPE Middle (33rd $\leq$ CPE $\leq$ 67th percentile, yellow line), and CPE High (67th percentile $<$ CPE ≤ 1, red line). Before taking group averages, the residual BOLD signals for each participant were normalized in two steps: first, within each block using z-scores, and second, across all blocks for each subject using z-scores. Shaded regions indicate time points (5s and 7.5s) where the averaged values were used for analysis. Asterisks mark significant t-tests at $p \leq 0.05$ in the GLMM.

Figure S1. Neural correlates of the cost prediction error (CPE) at the time of effort cue, at effort initiation, and at effort completion.

Panels A–F show positive correlations with CPE at the time of effort cue (GLM1), effort initiation (GLM2), and effort completion (GLM3) in Experiments 1 and 2. Panels G–L show corresponding negative correlations at the same stages. Results are thresholded at cluster-level FWE-corrected $p < 0.05$ or $p \approx 0.05$ and voxel-wise uncorrected $p < 0.001$.

Figure S2. Residual BOLD signals predicted the sequential choices.

Time-course analysis of residual BOLD signals for selected ROIs: the right IPS at effort cues, the vmPFC/vACC/OFC at effort initiation, and the dmFC/dACC, left aINS, left M1/S1/IPS, and bilateral Cd at effort completion. Curves represent residual BOLD signals averaged by CPE tertiles: CPE Low ($-1 \leq$ CPE $<$ 33rd percentile, blue line), CPE Middle (33rd $\leq$ CPE $\leq$ 67th percentile, yellow line), and CPE High (67th percentile $<$ CPE ≤ 1, red line). Before taking group averages, the residual BOLD signals for each participant were normalized in two steps: first, within each block using z-scores, and second, across all blocks for each subject using z-scores. Shaded regions indicate time points (5s and 7.5s) where the averaged values were used for analysis. Asterisks mark significant t-tests at $p \leq 0.05$ in the GLMM. The bottom panels of Figure S2C are identical to those in Figure 7B.

Table S1. BIS scores and best-fit parameters for demand avoiders.

This table shows the median, 25th and 75th percentiles of BIC scores and best-fit parameter estimates for 5 variants of reinforcement learning (RL) models with ActualCost (Solution-Time (SV), High-Low (HL), Incorrect-Correct (IC), ST-IC, and HL-IC), 2 variants of rest time considered RL models (Rest and HL-Rest models), and probabilistic Win-Stay-Lose-Shift (WSLS) and Probabilistic-Selection (PS) models. ECunchosen: the expected cost for the unchosen option, ECchosen: the expected cost for the chosen option, α: learning rate, β: inverse temperature, wincorrect: weighting

parameter, αcost: the learning rate for cost, αreward: the learning rate for reward, wreward: weighting parameter, p: probability to select the different and same option after Win and Lose, respectively, a: probability to select the same option after Win, b: probability to select the same option after Lose.

Table S2. Neural correlates of cost-prediction-error at the time of effort cues using GLM1.

Table S3. Neural correlates of cost-prediction-error at the time of effort initiation using GLM2.

Table S4. Neural correlates of cost-prediction-error at the time of effort completion using GLM3.

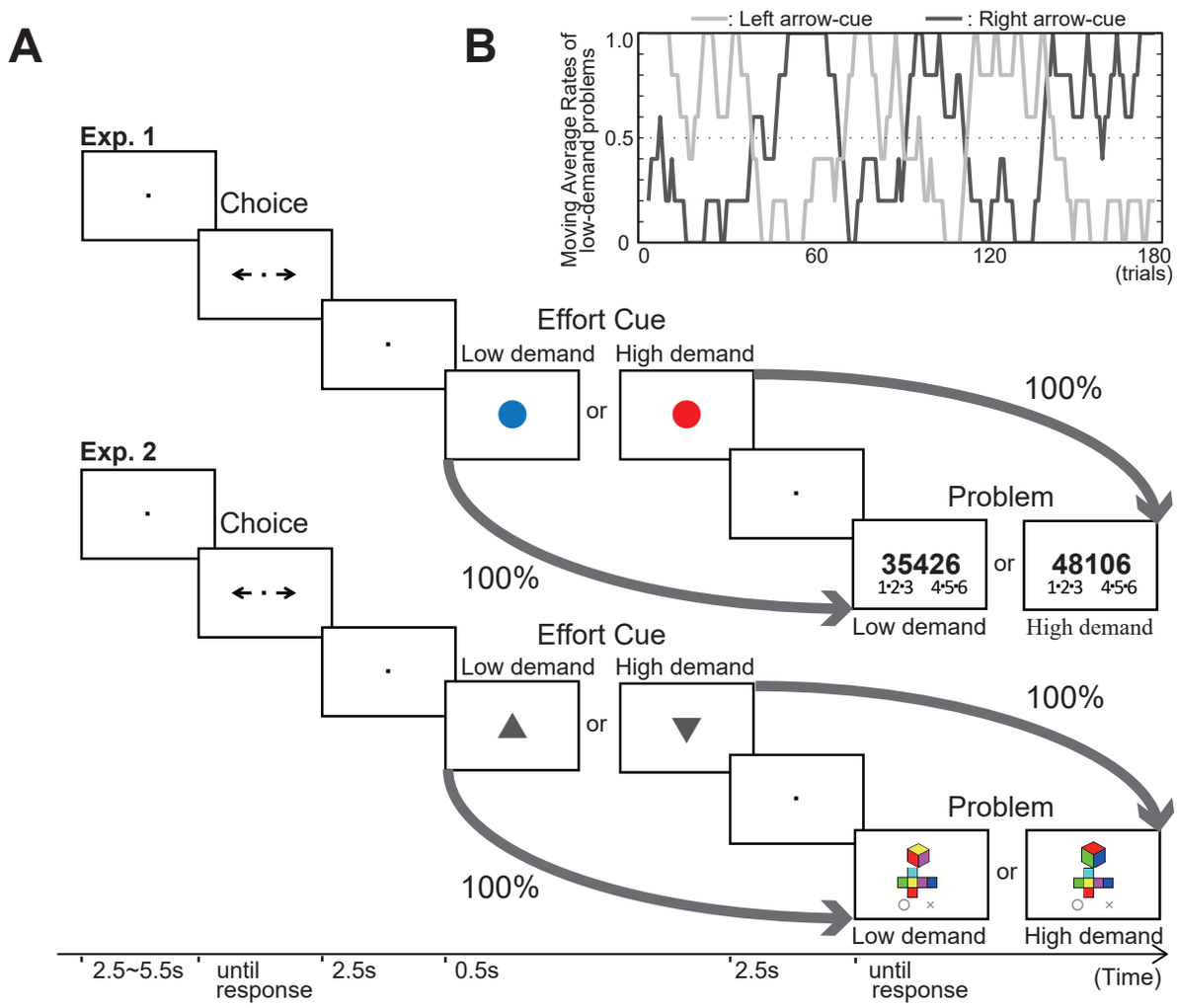

A
B
Exp. 1
Choice
Effort Cue
Low demand
High demand
or
100%
Problem
35426
1•2•3 4•5•6
48106
1•2•3 4•5•6
Low demand
High demand
Exp. 2
Choice
Effort Cue
Low demand
High demand
or
100%
Problem
Low demand
High demand
2.5~5.5s
until response
2.5s
0.5s
2.5s
until response
(Time)
: Left arrow-cue
: Right arrow-cue
Moving Average Rates of low-demand problems
1.0
0.5
0
0
60
120
180
(trials)

Figure 1

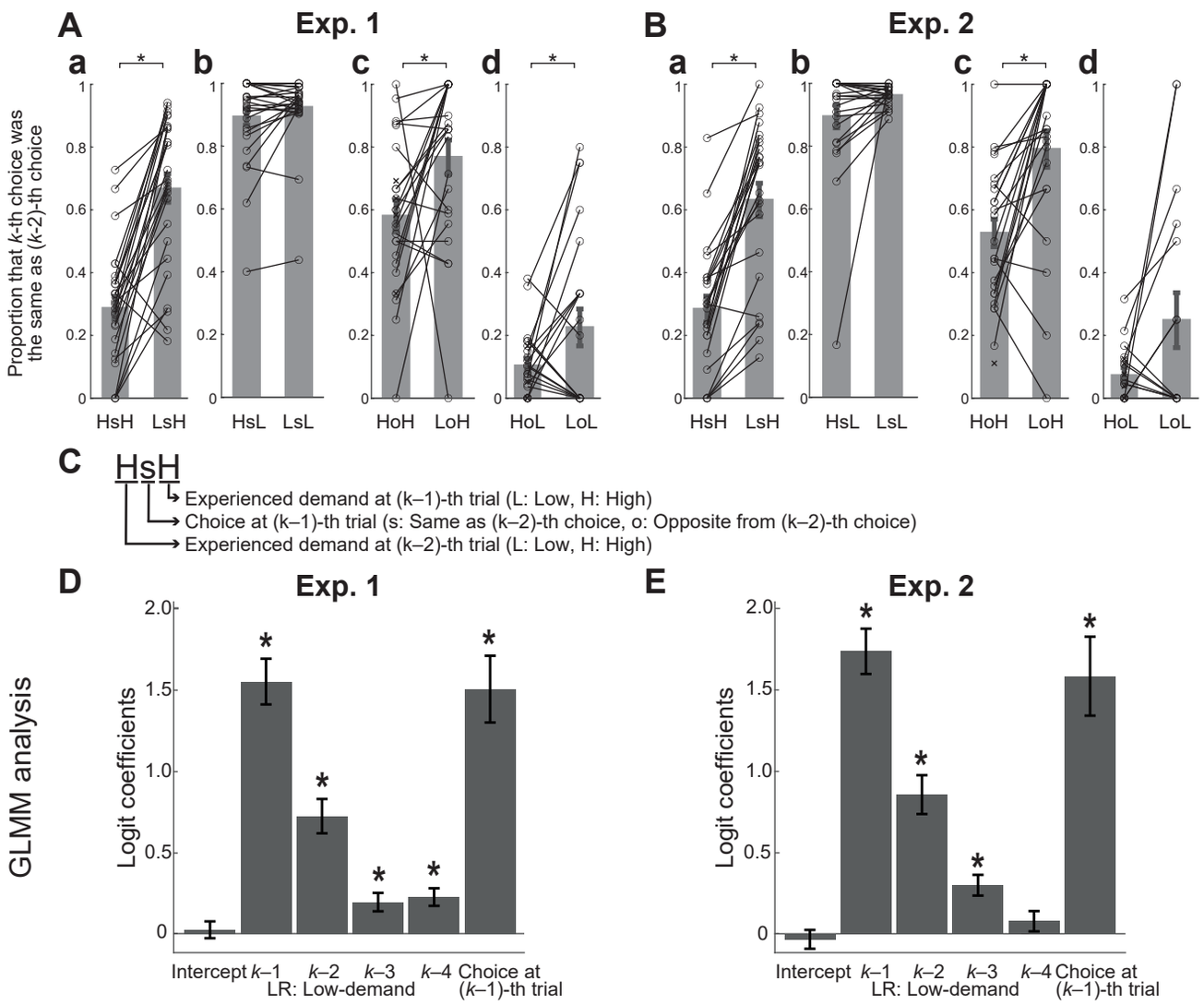
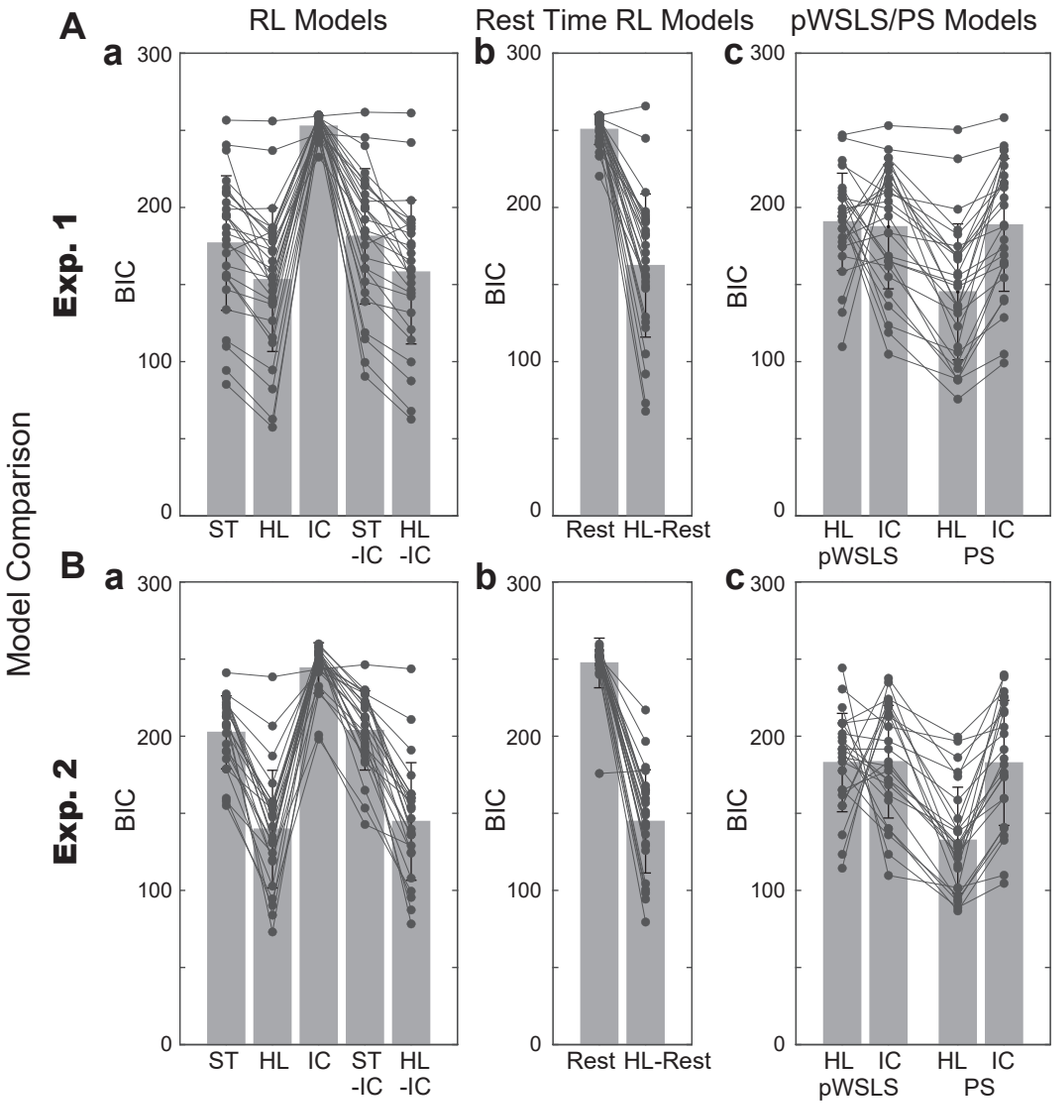
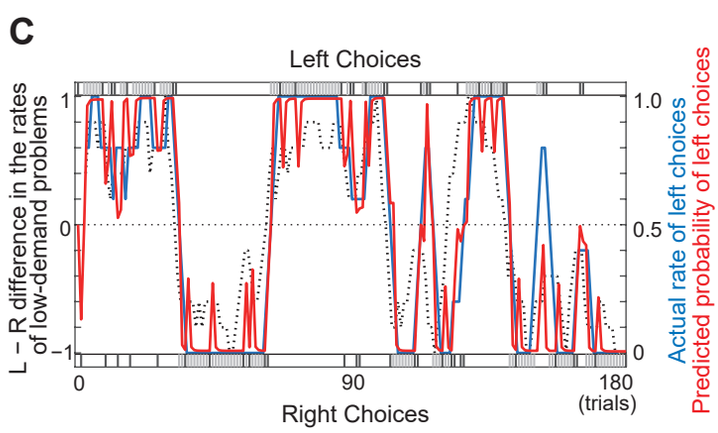

A
Exp. 1
B
Exp. 2
a
b
c
d
Proportion that k-th choice was the same as (k-2)-th choice
HsH
LsH
HsL
LsL
HoH
LoH
HoL
LoL
C
HsH
Experienced demand at (k–1)-th trial (L: Low, H: High)
Choice at (k–1)-th trial (s: Same as (k–2)-th choice, o: Opposite from (k–2)-th choice)
Experienced demand at (k–2)-th trial (L: Low, H: High)
GLMM analysis
D
Exp. 1
E
Exp. 2
Logit coefficients
Intercept
k–1
k–2
k–3
k–4
Choice at (k–1)-th trial
LR: Low-demand

Figure 2


A
RL Models
Rest Time RL Models
pWSLS/PS Models
a
b
c
Exp. 1
Exp. 2
Model Comparison
BIC
300
200
100
0
ST
HL
IC
ST -IC
HL -IC
Rest
HL-Rest
HL pWSLS
IC pWSLS
HL PS
IC PS
B
Abbreviations are as follows; pWSLS: probabilistic Win-Stay-Lose-Shift,
PS: Probabilistic-Selection, ST: Solution-Time, HL: High-Low, IC: Incorrect-Correct


C
Left Choices
L − R difference in the rates of low-demand problems
1
0
−1
0
90
180
(trials)
Right Choices
1.0
0.5
0
Actual rate of left choices
Predicted probability of left choices

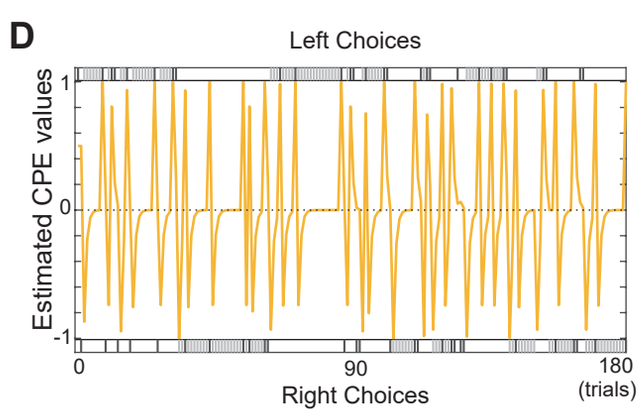

D
Left Choices
Estimated CPE values
1
0
-1
0
90
180
(trials)
Right Choices

Figure 3

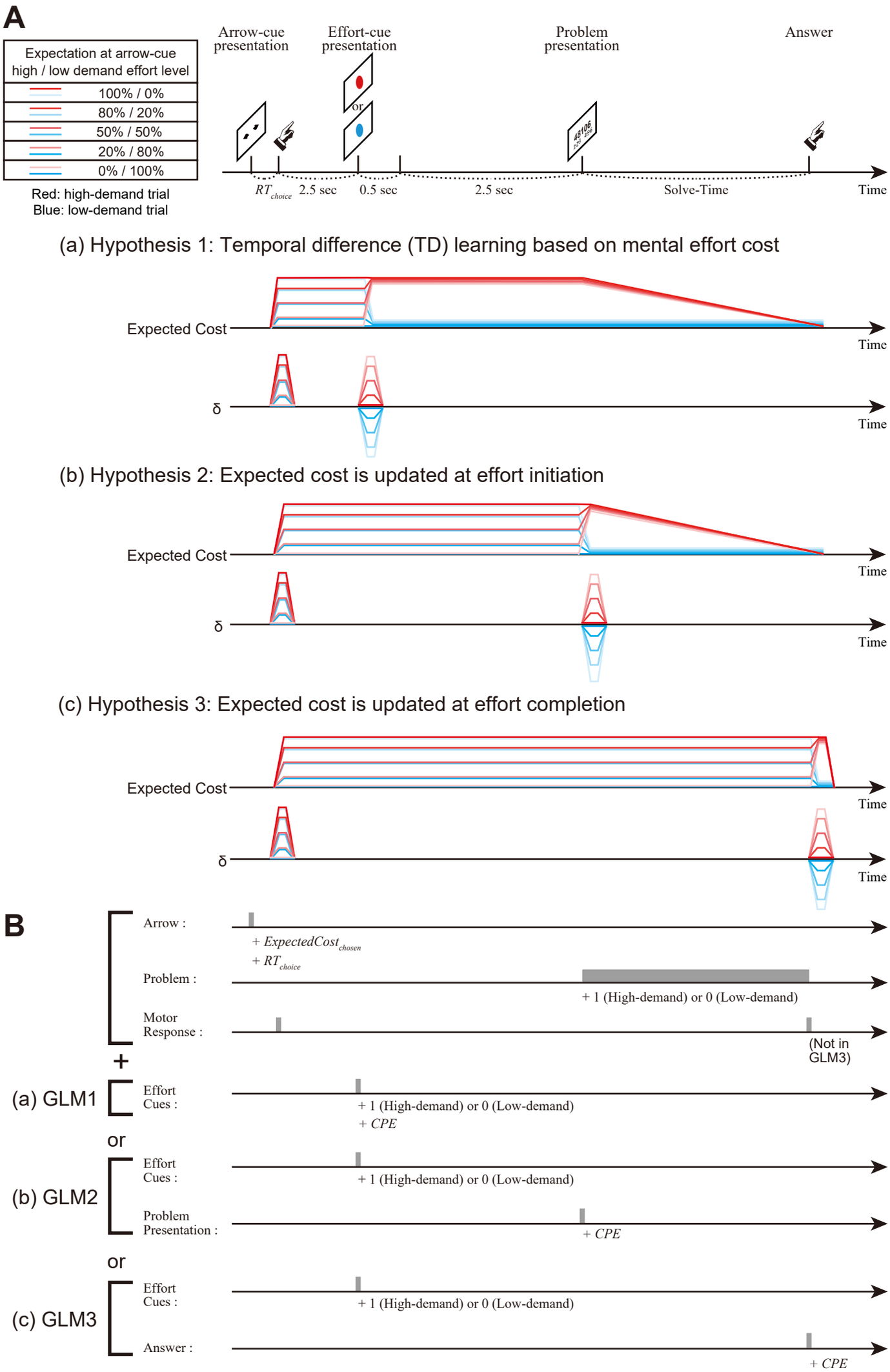

A
Expectation at arrow-cue
high / low demand effort level
100% / 0%
80% / 20%
50% / 50%
20% / 80%
0% / 100%
Red: high-demand trial
Blue: low-demand trial
Arrow-cue presentation
Effort-cue presentation
or
Problem presentation
Answer
RT choice
2.5 sec
0.5 sec
2.5 sec
Solve-Time
Time
(a) Hypothesis 1: Temporal difference (TD) learning based on mental effort cost
Expected Cost
Time
δ
Time
(b) Hypothesis 2: Expected cost is updated at effort initiation
Expected Cost
Time
δ
Time
(c) Hypothesis 3: Expected cost is updated at effort completion
Expected Cost
Time
δ
Time
B
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 4

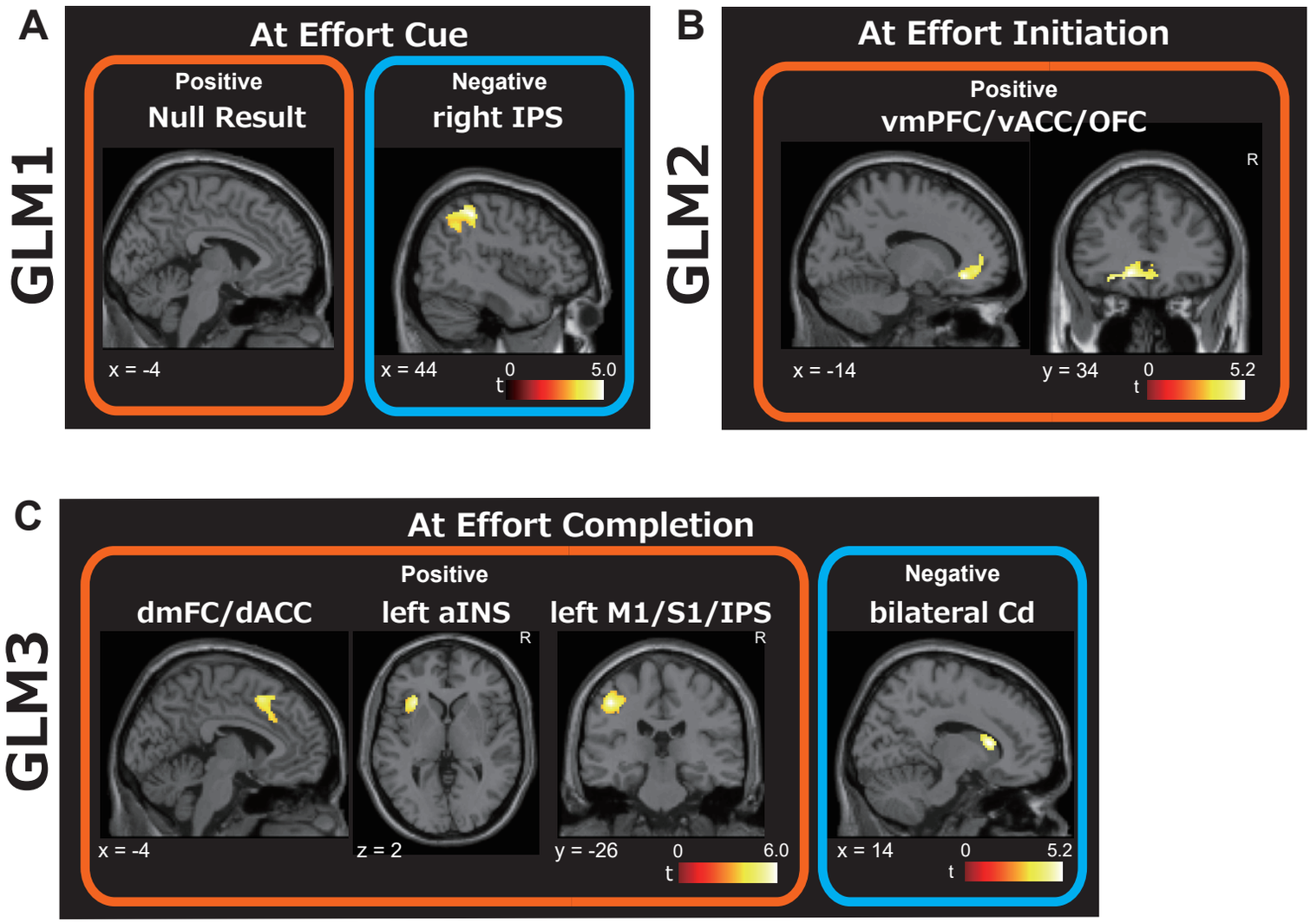

A
At Effort Cue
GLM1
Positive
Null Result
x = -4
Negative
right IPS
x = 44
0
5.0
t
B
At Effort Initiation
GLM2
Positive
vmPFC/vACC/OFC
R
x = -14
y = 34
0
5.2
t
C
At Effort Completion
GLM3
Positive
dmFC/dACC
left aINS
left M1/S1/IPS
R
R
x = -4
z = 2
y = -26
0
6.0
t
Negative
bilateral Cd
x = 14
0
5.2
t

Figure 5

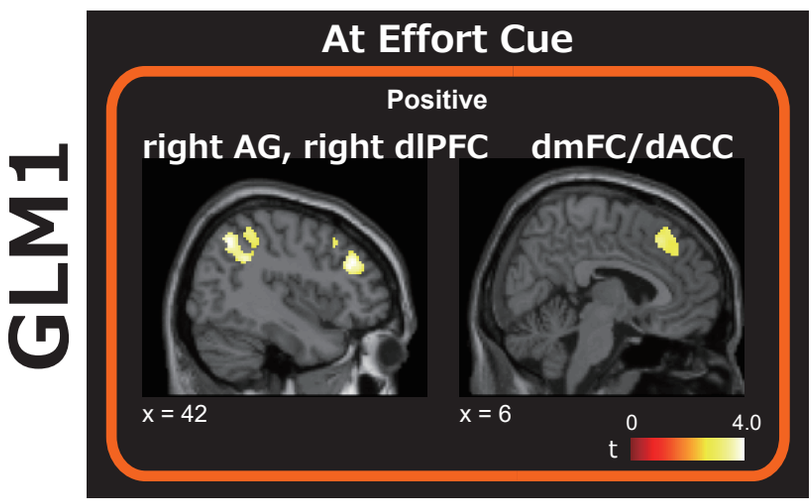

GLM1
At Effort Cue
Positive
right AG, right dlPFC
dmFC/dACC
x = 42
x = 6
0
4.0
t

Figure 6

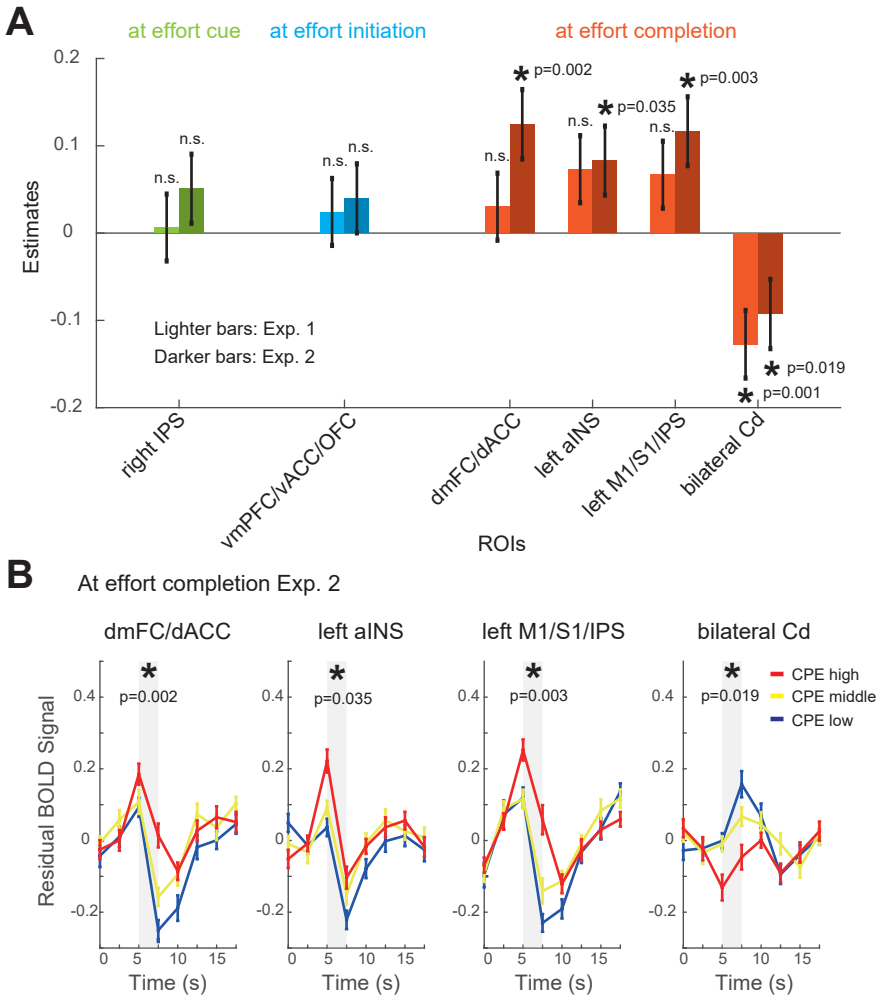

A
at effort cue
at effort initiation
at effort completion
Estimates
0.2
0.1
0
-0.1
-0.2
n.s.
p=0.002
p=0.035
p=0.003
p=0.019
p=0.001
Lighter bars: Exp. 1
Darker bars: Exp. 2
right IPS
vmPFC/vACC/OFC
dmFC/dACC
left aINS
left M1/S1/IPS
bilateral Cd
ROIs
B
At effort completion Exp. 2
dmFC/dACC
left aINS
left M1/S1/IPS
bilateral Cd
Residual BOLD Signal
CPE high
CPE middle
CPE low
Time (s)

Figure 7

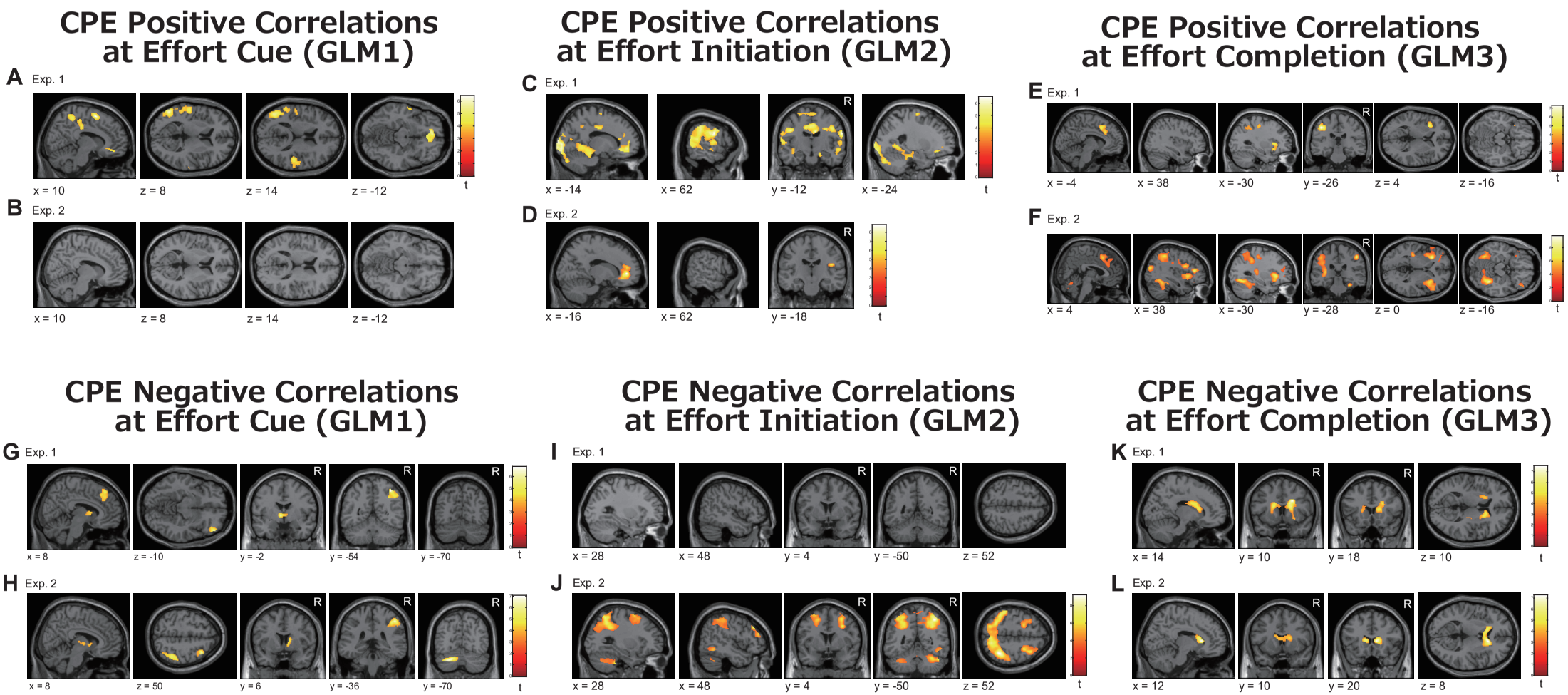

CPE Positive Correlations at Effort Cue (GLM1)
A Exp. 1
x = 10
z = 8
z = 14
z = -12
t
B Exp. 2
x = 10
z = 8
z = 14
z = -12
CPE Positive Correlations at Effort Initiation (GLM2)
C Exp. 1
x = -14
x = 62
y = -12
x = -24
t
D Exp. 2
x = -16
x = 62
y = -18
t
CPE Positive Correlations at Effort Completion (GLM3)
E Exp. 1
x = -4
x = 38
x = -30
y = -26
z = 4
z = -16
t
F Exp. 2
x = 4
x = 38
x = -30
y = -28
z = 0
z = -16
t
CPE Negative Correlations at Effort Cue (GLM1)
G Exp. 1
x = 8
z = -10
y = -2
y = -54
y = -70
t
H Exp. 2
x = 8
z = 50
y = 6
y = -36
y = -70
t
CPE Negative Correlations at Effort Initiation (GLM2)
I Exp. 1
x = 28
x = 48
y = 4
y = -50
z = 52
J Exp. 2
x = 28
x = 48
y = 4
y = -50
z = 52
t
CPE Negative Correlations at Effort Completion (GLM3)
K Exp. 1
x = 14
y = 10
y = 18
z = 10
t
L Exp. 2
x = 12
y = 10
y = 20
z = 8
t

Figure S1

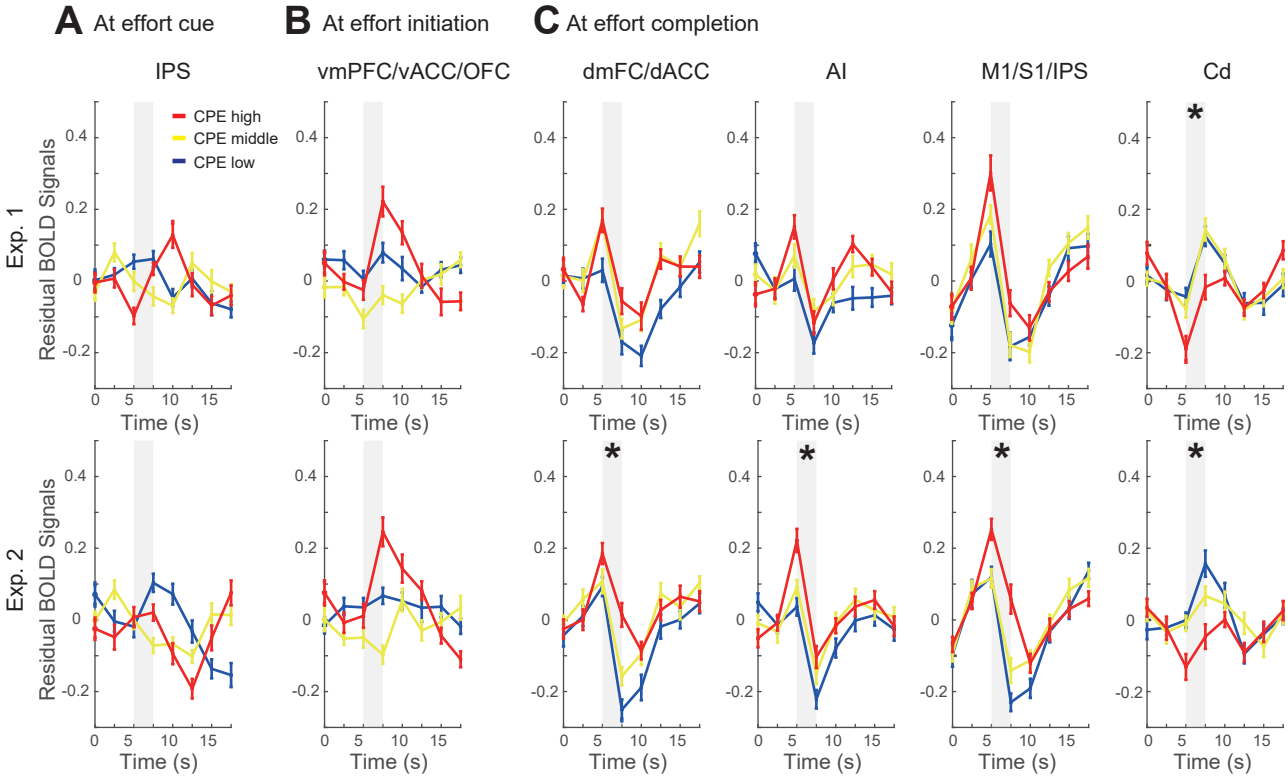

A At effort cue
B At effort initiation
C At effort completion
IPS
vmPFC/vACC/OFC
dmFC/dACC
AI
M1/S1/IPS
Cd
CPE high
CPE middle
CPE low
Exp. 1
Exp. 2
Residual BOLD Signals
Time (s)
*

Figure S2

Table S1

		Exp.1			Exp.2		
models	BIC/parameters	median	25 %ile	75 %ile	median	25 %ile	75 %ile
RL Models							
ST $EC_{\text{Unchosen}}(1) = EC_{\text{Chosen}}(1)$	BIC	**183.1**	**154.8**	**207.7**	**207.6**	**187.8**	**221.1**
	α	0.6	0.5	0.8	0.5	0.4	0.7
	β	0.2	0.1	0.4	0.2	0.1	0.4
HL $EC_{\text{Unchosen}}(1) = EC_{\text{Chosen}}(1)$	BIC	**155.2**	**129.2**	**182.2**	**138.1**	**119.5**	**156.1**
	α	0.8	0.6	1.0	0.7	0.6	0.9
	β	2.8	2.3	4.9	3.5	2.8	4.8
IC $EC_{\text{Unchosen}}(1) = EC_{\text{Chosen}}(1)$	BIC	**253.6**	**248.2**	**259.1**	**249.7**	**242.1**	**254.5**
	α	0.4	0.2	1.0	0.3	0.1	0.4
	β	0.8	0.0	4.0	3.0	-2.8	5.9
ST-IC $EC_{\text{Unchosen}}(1) = EC_{\text{Chosen}}(1)$	BIC	**185.1**	**154.0**	**212.6**	**209.5**	**189.0**	**224.8**
	α	0.6	0.5	0.7	0.5	0.4	0.7
	β	0.2	0.1	0.4	0.2	0.1	0.3
	$w_{\text{Incorrect}}$	0.8	0.0	16.6	5.0	0.0	16.1
HL-IC $EC_{\text{Unchosen}}(1) = EC_{\text{Chosen}}(1)$	BIC	**160.4**	**134.4**	**187.4**	**143.3**	**124.5**	**161.2**
	α	0.7	0.5	1.0	0.7	0.6	0.9
	β	2.8	2.3	4.8	3.4	2.8	4.7
	$w_{\text{Incorrect}}$	0.0	0.0	0.4	0.0	0.0	0.2
Rest-Time RL Models							
Rest	BIC	**254.2**	**244.3**	**257.7**	**251.5**	**246.3**	**254.4**
	α	0.3	0.2	0.9	0.2	0.1	0.3
	β	0.0	-0.1	0.0	0.1	-0.1	0.2
HL-Rest	BIC	**165.6**	**133.6**	**191.1**	**146.2**	**127.5**	**165.0**
	α_{Cost}	0.7	0.5	0.9	0.7	0.6	1.0
	α_{Reward}	0.0	0.0	0.3	0.0	0.0	0.2
	β	2.9	2.3	4.8	3.5	2.9	4.8
	w_{Reward}	0.2	0.0	188.3	0.1	0.0	4.2
pWSLS and PS Models							
pWSLS-HL	BIC	**194.2**	**175.4**	**207.9**	**187.6**	**163.5**	**205.0**
	p	0.2	0.2	0.3	0.2	0.2	0.2
pWSLS-IC	BIC	**194.2**	**162.7**	**222.7**	**182.1**	**160.1**	**216.7**
	p	0.2	0.2	0.3	0.2	0.2	0.3
PS-HL	BIC	**148.7**	**107.0**	**169.3**	**128.8**	**101.5**	**152.6**
	a	0.9	0.9	1.0	1.0	0.9	1.0
	b	0.5	0.4	0.6	0.5	0.4	0.6
PS-IC	BIC	**188.7**	**163.0**	**225.5**	**183.5**	**150.2**	**219.1**
	a	0.8	0.7	0.9	0.8	0.7	0.9
	b	0.5	0.3	0.9	0.7	0.4	0.9

Table S2

Area	LR	cluster size	cluster p (FWE)	peak T	MNI {mm} x	y	z
A Exp. 1							
Positive correlation							
SMC/PCC	B	3720	<0.001	6.40	10	6	60
MTG/AG	L	2113	<0.001	5.92	-50	-64	8
vmPFC	B	497	<0.001	5.64	-6	46	-12
PO	R	426	<0.001	4.83	50	-24	14
Negative correlation							
IPS	R	1095	<0.001	6.80	38	-54	54
Thalamus	L	289	0.038	6.10	-6	-2	2
OFC/dlPFC	R	974	<0.001	5.84	50	44	-10
dmFC	B	591	0.002	5.17	8	38	50
B Exp. 2							
Positive correlation							
null							
Negative correlation							
Cerebellum	L	568	<0.001	7.02	-22	-70	-30
IPS/AG	R	1439	<0.001	6.35	48	-36	50
FEF/dlPFC	R	720	0.001	6.31	36	18	54
Caudate/TP	R	338	0.034	4.89	12	6	10
C Conjunction analysis (strict mask [1])							
Positive correlation							
null							
Negative correlation							
IPS	R	710	0.002	5.88	44	-42	52
D Conjunction analysis (relaxed mask [2])							
Positive correlation [3]							
vmPFC	B	114	0.365	4.10	-4	44	-12
CO	R	4	0.974	3.46	42	0	12
Negative correlation [3]							
IPS	R	821	<0.001	5.88	44	-42	52
OFC	R	189	0.163	5.01	50	44	-10
FEF/dlPFC	R	236	0.101	4.71	36	20	54

Area	LR	cluster size	cluster p (FWE)	peak T	MNI {mm} x	y	z
Cerebellum	L	137	0.285	4.56	-30	-68	-30
dlPFC	R	65	0.610	4.35	48	40	20
IC	R	95	0.448	4.26	12	8	4
dmFC	B	267	0.074	4.14	6	30	46
dlPFC	L	57	0.659	3.90	-42	52	-2
IPS	L	24	0.871	3.90	-44	-48	52
Cd	L	15	0.923	3.90	-8	8	4
aI	R	49	0.711	3.72	32	20	-6
IFG	R	11	0.943	3.46	54	20	8
Cerebellum	L	4	0.974	3.41	-8	-80	-30
Thalamus	R	3	0.978	3.38	6	-14	6

(1): a binary mask consisting of common voxels at cluster-level FWE corrected $p < 0.05$ or $p \approx 0.05$ and voxel-level uncorrected $p < 0.001$

(2): a binary mask consisting of common voxels at voxel-level uncorrected $p < 0.01$

(3): voxel-level uncorrected $p < 0.001$

Table S3

Area	LR	cluster size	cluster p (FWE)	peak T	MNI {mm} x	y	z
A Exp. 1							
Positive correlation							
S1/V2/EC/FC	B	17659	<0.001	6.52	62	-6	28
S1/PC	L	2544	<0.001	5.54	-44	-12	32
SMC/PMC	B	262	0.045	4.56	-24	-10	68
Negative correlation							
null							
B Exp. 2							
Positive correlation							
vmPFC/vACC	B	2986	<0.001	8.77	16	44	2
Operculum	R	489	0.008	5.95	42	-18	20
Negative correlation							
null							
C Conjunction analysis (strict mask (1))							
Positive correlation							
vmPFC/vACC/OFC	B	961	<0.001	5.37	-14	34	-8
Negative correlation							
null							
D Conjunction analysis (relaxed mask (2))							
Positive correlation (3)							
vmPFC/vACC/OFC	B	1522	<0.001	5.37	-14	34	-8
OFC	R	9	0.955	4.31	28	34	-12
dACC	L	30	0.836	3.99	-2	30	16
PO	R	47	0.724	3.91	46	-22	20
CO	R	17	0.914	3.43	48	-10	16
STC	R	3	0.979	3.41	66	-40	20
dmPFC	R	7	0.964	3.40	2	52	18
MTC	R	4	0.976	3.33	46	-38	4
MTC	L	1	0.987	3.32	-60	-2	-18
MTC	L	1	0.987	3.30	-58	0	-20
Negative correlation (3)							
null							

(1): a binary mask consisting of common voxels at cluster-level FWE corrected $p < 0.05$ or $p \approx 0.05$ and voxel-level uncorrected $p < 0.001$

(2): a binary mask consisting of common voxels at voxel-level uncorrected $p < 0.01$

(3): voxel-level uncorrected $p < 0.001$

Table S4

Area	LR	cluster size	cluster p (FWE)	peak T	MNI {mm} x	y	z
A Exp. 1							
Positive correlation							
S1/M1	L	1397	<0.001	7.25	-44	-26	46
aI	L	368	0.012	5.75	-32	18	4
dmFC/dACC	B	635	<0.001	5.49	-6	14	46
Negative correlation							
MFG/Cd	B	2212	0.000	7.51	24	10	30
B Exp. 2							
Positive correlation							
FG/Celebellum	R	1899	<0.001	9.74	34	-48	-16
aI	L	1005	<0.001	8.18	-28	24	0
IPS/MOG	R	3319	<0.001	8.00	48	-28	48
dlPFC/AI	R	2323	<0.001	7.45	38	36	14
FG/Celebellum	L	953	<0.001	6.88	-30	-70	-12
PI/IPS/M1	L	3103	<0.001	6.29	-36	-24	-2
dmFC/dACC	B	1346	<0.001	5.87	4	18	46
IPS	L	384	0.011	5.10	-22	-70	32
MFG/SFG	L	289	0.033	5.00	-22	46	10
IFG/PCG	L	340	0.018	4.99	-56	12	24
Negative correlation							
Cd	B	1062	<0.001	7.24	12	20	8
C Conjunction analysis (strict mask [1])							
Positive correlation							
M1/S1/IPS	L	940	<0.001	6.01	-44	-26	46
aI	L	322	0.032	5.62	-30	22	2
dmFC/dACC	B	423	0.012	5.07	-4	14	50
Negative correlation							
Cd	B	371	0.019	5.42	14	18	10
D Conjunction analysis (relaxed mask [2])							
Positive correlation [3]							
S1/M1/IPS	L	1092	<0.001	6.01	-44	-26	46

Area	LR	cluster size	cluster p (FWE)	peak T	MNI {mm} x	y	z
aI	L	434	0.011	5.62	-30	22	2
dmFC/dACC	B	644	0.002	5.07	-4	14	50
FG	L	78	0.518	4.72	-30	-72	-10
AL/FO	R	283	0.048	4.52	36	22	4
FG	R	220	0.097	4.47	34	-48	-20
dlPFC	R	60	0.633	4.26	50	34	16
PCG/MFG	L	29	0.850	4.12	-46	6	32
PCG/IFG	R	105	0.377	4.01	46	8	26
SMG	R	39	0.781	3.93	46	-40	42
Precuneus	R	13	0.945	3.91	20	-62	26
OFC	R	21	0.901	3.68	32	18	-20
MFG	L	10	0.959	3.56	-28	4	44
AG	R	9	0.963	3.49	56	-48	28
FG	R	3	0.985	3.49	32	-34	-22
FG	R	1	0.991	3.37	32	-30	-24
MFG	L	2	0.988	3.34	-22	48	12
MTG	L	3	0.985	3.32	-60	-28	-14
SMG	R	1	0.991	3.31	50	-26	48
Negative correlation (3)							
Cd	B	553	0.004	5.42	14	18	10
Thalamus	B	28	0.857	4.18	0	-10	6
Cd	L	5	0.979	3.42	-16	28	6
Pallidum	L	1	0.991	3.40	-20	-6	-10

(1): a binary mask consisting of common voxels at cluster-level FWE corrected $p < 0.05$ or $p \approx 0.05$ and voxel-level uncorrected $p < 0.001$

(2): a binary mask consisting of common voxels at voxel-level uncorrected $p < 0.01$

(3): voxel-level uncorrected $p < 0.001$