AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

2023, Vol. 152, No. 12, 3440–3458
https://doi.org/10.1037/xge0001449

Policy Abstraction as a Predictor of Cognitive Effort Avoidance

Ceyda Sayalı[1], Jordan Rubin-McGregor[2], and David Badre[3, 4]
[1] Center for Psychedelic and Consciousness Research, Johns Hopkins University School of Medicine
[2] Department of Psychology, Miami University
[3] Department of Cognitive, Linguistic, and Psychological Sciences, Brown University
[4] Carney Institute for Brain Science, Brown University

Consistent evidence has established that people avoid cognitively effortful tasks. However, the features that make a task cognitively effortful are still not well understood. Multiple hypotheses have been proposed regarding which task demands underlie cognitive effort costs, such as time-on-task, error likelihood, and the general engagement of cognitive control. In this study, we test the novel hypothesis that tasks requiring behavior according to higher degrees of policy abstraction are experienced as more effortful. Accordingly, policy abstraction, operationalized as the levels of contextual contingency required by task rules, drives task avoidance over and above the effects of task performance, such as time-on-task or error likelihood. To test this hypothesis, we combined two previously established cognitive control tasks that parametrically manipulated policy abstraction with the demand selection task procedure. The design of these tasks allowed us to test whether people avoided tasks with higher order policy abstraction while controlling for the contribution of factors such as time-on-task and expected error rate (ER). Consistent with our hypothesis, we observed that policy abstraction was the strongest predictor of cognitive effort choices, followed by ER. This was evident across both studies and in a within-subject cross-study analysis. These results establish at least one task feature independent of performance, which is predictive of task avoidance behavior. We interpret these results within an opportunity cost framework for understanding aversive experiences of cognitive effort while performing a task.

Public Significance Statement
In this study, we provide evidence for a novel hypothesis that task demands related to policy abstraction, operationalized as the levels of contextual contingency required by task rules, drive task avoidance beyond the effects of task performance, such as time-on-task or error likelihood. Using a demand selection task procedure, we show that policy abstraction is the strongest predictor of cognitive effort choices, followed by error rates. Our results provide insight into the features that make a task cognitively effortful and suggest that cognitive effort costs can be influenced by factors beyond task performance. These findings may have implications for designing interventions aimed at reducing cognitive effort avoidance.

Keywords: cognitive control, executive function, cognitive effort, motivation, decision making

Supplemental materials: https://doi.org/10.1037/xge0001449.supp

There is now consistent evidence that performing certain tasks is associated with a greater subjective experience of cognitive effort than others are. People often find these effortful tasks aversive and avoid them (Cavanagh et al., 2014; Gold et al., 2015; Kool et al., 2010; McGuire & Botvinick, 2010; Sayalı & Badre, 2019, 2021; Shenhav et al., 2013; Westbrook et al., 2013). Likewise, people discount rewarding outcomes from a task based on the mental effort required by the task (Cavanagh et al., 2014; Westbrook et al., 2013). Cognitive effort is then interpreted as a cost during decision making, as in a reduction in the value associated with a task. Although subjective and objective indices of cognitive effort costs do not always perfectly map onto one another, it is generally inferred

This article was published Online First August 24, 2023.
Ceyda Sayalı https://orcid.org/0000-0002-3420-5499
This work was supported by a Multidisciplinary University Research Initiative (MURI) award from the Office of Naval Research (N00014-16-1-2832) and the National Institute of Mental Health (R01 MH125497). Data are available on request from the authors. Analysis codes can be found at https://github.com/zceydas/HierCost. This article has been previously published in a preprint form on psyarxiv.com. This work has not been preregistered.
Ceyda Sayalı served as lead for conceptualization, formal analysis, methodology, writing–original draft, and writing–review and editing. Jordan Rubin-McGregor served as lead for data curation, project administration, and visualization, contributed equally to formal analysis, and served in a supporting role for writing–original draft and writing–review and editing. David Badre served as lead for funding acquisition, resources, and supervision and contributed equally to conceptualization, methodology, writing–original draft, and writing–review and editing.
Correspondence concerning this article should be addressed to Ceyda Sayalı, Center for Psychedelics and Consciousness Research, Johns Hopkins University School of Medicine, 5510 Nathan Shock Drive, Baltimore, MD 21224, United States. Email: zsayali1@jh.edu

that subjective effort costs drive people to avoid effortful tasks if the rewarding outcomes from the task are insufficient.

Importantly, although the cost of cognitive effort may be inferred from task avoidance, it remains unknown what characteristics of a task make it feel effortful. Accordingly, scientists have no established way of predicting whether a particular task will be avoided because people find it effortful, apart from observing that they avoid it. Therefore, to avoid circular reasoning about cognitive effort, an important goal is to identify factors, separate from the tendency to avoid a task itself, which marks that task as effortful and predicts that people will avoid it. In other words, what task features lead to cognitive effort avoidance?

One hypothesized source of cognitive effort costs is simply the degree to which one expects to perform poorly on a task. Harder tasks are defined by impoverished performance relative to easier tasks, as would be evident in higher error rates (ER). If it is harder to complete a task, the reward one stands to gain is less likely and should be devalued accordingly. In other words, harder tasks require more cognitive processing and resources to achieve arbitrary levels of performance. As such, the brain might penalize effort on such tasks with an intrinsic cost. Hence, tasks that are associated with greater ER (or longer response times [RT] to avoid errors) will be associated with an experience of effort and will be avoided in the future.

Consistent with this view, both ER (T. L. Dunn, Inzlicht, & Risko, 2019) and RT (Sayalı & Badre, 2021) have been shown to predict effort avoidance learning across individuals. Furthermore, there is some evidence that the brain monitors error likelihood within the cognitive control system. In particular, medial prefrontal cortex (PFC) neurons may increase their activity in proportion to the error likelihood of a context during the cue period (Alexander & Brown, 2015; Brown & Braver, 2005). Thus, it is plausible that the brain monitors error likelihood, and this performance-derived signal is interpreted as a cost during decision making regarding what task to perform.

A related performance-based account derives from the fact that longer RT not only signal a potentially higher likelihood of error, but also delay the receipt of rewards or other positive outcomes that follow the completion of a task. It is well established that decision makers prefer smaller but more immediate rewards over larger but delayed rewards, termed delay discounting (e.g., Myerson & Green, 1995). Thus, tasks with longer performance times may also be subject to this bias.

Longer task completion times may also be experienced as effortful because of the impact of performing these tasks on our ability to perform other tasks. More specifically, some tasks may tie up more cognitive systems or resources that are capacity-limited and may do so for longer periods. As such, performing one task may compromise the performance of other rewarding tasks that can be performed simultaneously. For example, while we perform geometry proofs, we can simultaneously write a paper or do programming. The longer the proof is, the longer we must go without performing other tasks.

It follows that performing a task that taxes capacity-limited systems incurs an opportunity cost, as in the loss of potential value from other unchosen tasks that cannot be performed while the chosen task occupies resources. Opportunity cost accounts of cognitive effort (e.g., Boureau et al., 2015; Kurzban et al., 2013; Musslick & Cohen, 2021; Niv et al., 2007) propose that the experience of effort and its cost in decision making derive from the counterfactual value of the foregone opportunity while being on task.

As people are generally bad at multitasking, the longer they perform any given task, the greater the opportunity cost that will be imposed to the extent that performing other tasks is valuable. Consequently, the expected time required to complete a task determines whether it is effortful (Kurzban et al., 2013). Thus, we might generally penalize tasks according to the time they take, regardless of the particulars of the tasks or the cognitive systems required. This general time-on-task account is consistent with the evidence that the average RT on a task correlates with avoidance behavior (Otto & Daw, 2019; Sayalı & Badre, 2021).

Both error likelihood and time-on-task as sources of cognitive effect costs are empirically supported and are the baseline hypotheses against which other more process-oriented accounts of effort costs should be compared. Nonetheless, there may be alternatives derived from theories about the specific capacity limits of cognitive systems and the ways in which a particular task might demand those systems.

In one such case, task demands that increase the likelihood of multitasking costs should result in greater opportunity costs; therefore, tasks with such demands should be experienced as more mentally effortful, even beyond what is explained by time-on-task or error likelihood more generally. Although multitasking costs have been widely observed, they are not universally observed. We can perform certain tasks simultaneously. Chewing gum, for example, would likely not limit our ability to perform geometric proofs or other demanding tasks. One loses no value from one of these tasks while performing the other, regardless of the amount of time one spends on them.

Thus, it is important to ask whether there are particular capacity-limited cognitive systems or resources that, when demanded by a task, contribute differentially to cognitive effort costs, because they signal that multitasking will be particularly ineffective. One such possibility, which we will test in the present experiments, is the demand that a task places on *hierarchical control*, which is expressed in terms of the degree of policy abstraction the task requires.

In general, cognitive control allows mapping from an input state to an appropriate response (i.e., an action policy) to change as a function of contextual contingency (Badre et al., 2021; Miller & Cohen, 2001). Hierarchical cognitive control refers to situations in which the influence of a context on the choice of response is itself determined by one or more higher order contexts (Badre & D'Esposito, 2009; Badre & Nee, 2018; Botvinick, 2007; Dayan, 2007).

As an example of hierarchical control, a green crossing light may indicate that pedestrians may cross, and a red-crossing light may indicate that pedestrians should stop. This comprises a simple rule or *policy* in which two stimuli are mapped to two responses. However, in the case where there are no cars, pedestrians sometimes cross, regardless of the crossing signal. In this case, the presence of cars is a superordinate context that determines how the crossing signal is interpreted. Indeed, contingency levels can be continuously added. For example, one might know that in one town, enforcement laws are strictly enforced, even when no cars are present, but not in another. Thus, the town now forms a superordinate context that determines how the presence or absence of cars should be interpreted with respect to different crossing signals.

This nesting of rules is a feature of hierarchical control, termed *policy abstraction*, because, as additional levels of contextual contingency are added, each higher level of contingency contextualizes a more abstract class of lower order policies (Badre & Nee, 2018; Newell, 1994). Thus, the number of contingencies required by a rule determines its degree of policy abstraction. In experimental settings, policy abstraction can be operationalized in terms of the depth of the decision tree traversed en route to a response (Figure 1).

Within this framework, the specific hypothesis we test is that greater degrees of policy abstraction (i.e., deeper rule trees) will be associated with higher mental effort costs. Consequently, tasks with higher degrees of policy abstraction will be avoided at higher rates than those with lower degrees of policy abstraction.

This hypothesis is motivated by the assumption that behaving according to higher order policy will compromise multitasking more than behaving according to lower order policy, and accordingly, will incur a greater opportunity cost. This assumption is justified by several reasons. First, multitasking is a form of hierarchical control. In other words, if one cannot perform two or more tasks at once, such as because there is overlap among components of the tasks that causes mutually incompatible interference, then context is needed to resolve that conflict (Badre, 2008; Badre & Nee, 2018; A. G. Collins & Frank, 2013; Koechlin & Hyafil, 2007). This means that an additional level of contingency is added to the policy-governing behavior during this multitasking scenario. To the degree that humans do not have a limitless capacity to manage multiple levels of contingency, tasks that involve deeper levels of policy abstraction will prevent effective multitasking more than those that do not. There is some evidence for such capacity limits in human decision making (Koechlin & Hyafil, 2007).

Second, it was observed that indices of multitasking costs, such as the cost of switching between more than one task, are larger at higher levels of policy abstraction. In other words, the cost of performance is greater when switching superordinate versus subordinate levels of context (A. G. E. Collins & Frank, 2016), and costs from higher level decisions are inherited by lower order decisions (Kleinsorge & Heuer, 1999). Furthermore, although there is evidence that humans can make decisions at different levels of hierarchical contingency in partial parallel (Rac-Lubashevsky & Frank, 2021; Ranti et al., 2015), they do not do so in perfect parallel. Rather, there are interactions among the levels of contingency, and conflict at any level increases decision thresholds (Rac-Lubashevsky & Frank, 2021).

These observations indicate that multitasking costs and vulnerability to conflict increase with higher policy abstraction. As such, it is reasonable to hypothesize that such tasks incur a greater opportunity cost, will be accompanied by an experience of greater mental effort, and will be avoided at higher rates, even when controlling for overall performance differences.

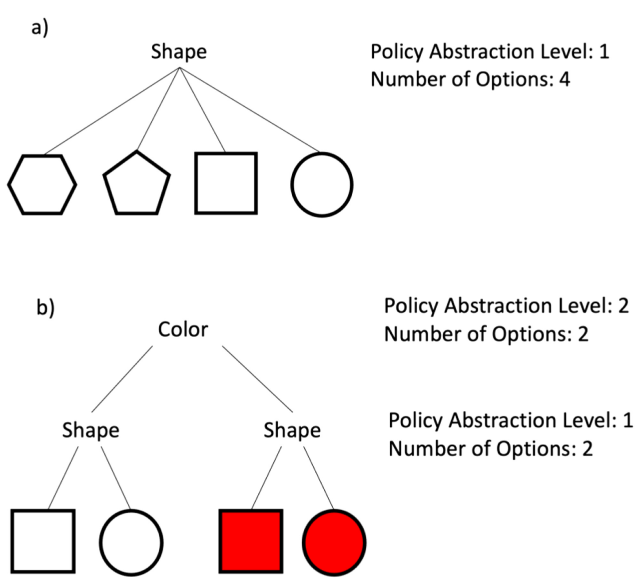
To test the hypothesis that policy abstraction drives cognitive effort avoidance, we combined two cognitive control paradigms with a demand selection task (DST). In a DST, participants repeatedly choose between two tasks that differ in their cognitive effort (Gold et al., 2015; Kool et al., 2010; McGuire & Botvinick, 2010; Schouppe et al., 2014). In this adapted version of the DST that we employed previously (Sayalı & Badre, 2019, 2021), an initial learning phase allowed participants to gain experience with the task. In the subsequent selection phase, participants chose between alternative tasks and then performed the task they selected. The cognitive control paradigms we adopted (Badre & D'Esposito, 2007; Nee & D'Esposito, 2016) operationalize policy abstraction as the depth of the conditional rule tree that must be traversed to make a response.

Both tasks manipulate hierarchical cognitive control and lend themselves to the operational definitions of task conditions in terms of policy abstraction. Both tasks have also been observed to differentially engage frontoparietal networks in the human brain as a function of policy abstraction changes. Specifically, Badre and D'Esposito (2007) observed that increasing policy abstraction across four increasing levels was associated with the recruitment of distinguishable frontoparietal networks that included progressively more rostral regions in the PFC (see Choi et al., 2018). Nee and D'Esposito (2016) used a different set of tasks to show that progressively abstract levels of control (e.g., feature, contextual, and temporal control) were associated with similar distinct networks (Badre & Nee, 2018).

Relevant to the present experiments, policy abstraction in both paradigms can be defined as the depth of the conditional rule tree that must be traversed to make a response. Furthermore, several conditions included in each paradigm varied the competition within each abstraction level, operationalized as the number of potential responses at a given level of contingency. For example, if one level of the decision tree is a choice based on the context of "shape," the competition within the level could be varied by having a different number of possible shapes (Figure 1). Hence, competition within an abstraction level and the abstraction level of the task are not collinear because there are pairs of task conditions that vary the level of response competition, but keep the level of policy abstraction constant.

These are the task conditions for which previous studies observed changes in task performance measures, despite no change in policy abstraction. Because of these manipulations, behavioral performance in both studies was not collinear with changes in policy abstraction. For the present purposes, this feature allows us to test

Figure 1
Visual Example of Levels of Policy Abstraction and Number of Options



Note. (a) In the first example, the policy abstraction level is 1. Thus, there are four options for each abstraction level. (b) In the second example, the policy abstraction level for the task is 2. There are two options at the second abstraction level (red/dark grey vs. white) and two options for each branch at the first level (circle vs. square). This means that in the second example, shape information alone is insufficient to make a response. See the online article for the color version of this figure.

the effects of time-on-task and error likelihood on avoidance rates separately from the effect of policy abstraction.

Furthermore, the experiments differ both within and between studies on several specific features, such as cues, responses, stimuli, and working memory (WM) demands. This allowed us to test the general effect of policy abstraction on avoidance decisions across these differences. Across these separate experimental tasks, we predict that demand selection rates will correlate separately and negatively with the abstraction level of the task and task performance.

Experiment 1

Our primary question of interest was whether policy abstraction (decision tree depth), number of options (decision tree breadth), error likelihood, and/or time on task underlie effort avoidance. In Experiment 1, we used an adapted and shortened version of the behavioral task used in Badre and D'Esposito (2007). This task parametrically varied the number of options at each of the multiple levels of policy abstraction, while also varying the number of abstraction levels within the same task type. Therefore, this paradigm allowed us to test the effect of increasing competition within and across levels of policy abstraction, and explore their relative weights in combination with performance measures in predicting selection rates.

Method

Subjects

Based on the pilot study results ($N = 5$), we calculated the sample size needed ($N = 16$) to achieve 80% power using the effect size ($d = 0.98$) observed in the demand avoidance rates between conditions that differ in terms of policy abstraction and not task type (see Method for the definition of the F4/F1 condition and Results 1 in the online supplemental materials for the sample size plot). Participants were asked to verbally self-report their age and sex. For Experiment 1, 19 right-handed participants (eight male, 11 female, ages 19–35) with normal or corrected-to-normal vision and without neurological or psychiatric illness were recruited to participate in a two-part behavioral study. Two of the 19 participants (both male) were unable to participate in Experiment 1: one due to illness and one due to technical failure. One participant's data were excluded to show extreme performance differences from the group mean (>2.5 standard deviations [*SDs*] worse than the mean for both RT and accuracy) prior to the analysis of task selection rates. This participant was also excluded from Experiment 2. The remaining 16 participants were included in subsequent analyses. All behavioral tests were conducted according to the procedures approved by the Human Research Protection Office of Brown University. The participants provided informed consent in accordance with the Research Protection Office of Brown University. All participants were compensated by course credit or monetary payments at a rate of $10/hr.

Procedure

Experiment 1 was based on tasks outlined in Badre and D'Esposito (2007). This study used tasks at four levels of policy abstraction. However, we utilized only the first two levels (response and feature) to test our hypotheses.

Experimental Tasks and Logic

Response Task. In the response task, competition along the response dimension was parametrically manipulated by changing the number of response options available to the participants from one to two to four (Figure 2a). For all mappings, four differently colored squares cued what response was to be made. In the R1 condition, all four colors cued the same response; in R2, two responses were cued by either of the two colors (e.g., red or blue boxes cued one response, while yellow or green cued another); and in R4, each color cued a different response. Based on these mappings, the participants responded to a colored target box presented at the center of the screen by pressing the respective key on the computer keyboard using their right hand.

Feature Task. In the feature task, competition along the feature dimension was varied parametrically by manipulating the number of sets of stimulus-to-response mappings. Specifically, a colored square cued in which the direction of a presented arrow took a target response and which one took a nontarget response. This effectively manipulated the mapping between the arrow direction and response, with the mapping changing as a function of the color cue. Across the feature task conditions, we manipulated whether one arrow direction, two arrow directions, or four arrow directions could be cued as a target by four colored squares (Figure 2b). Participants responded to the arrow within the square by pressing one of the two buttons ("yes" [target) or "no" [not target]) using their right hand.

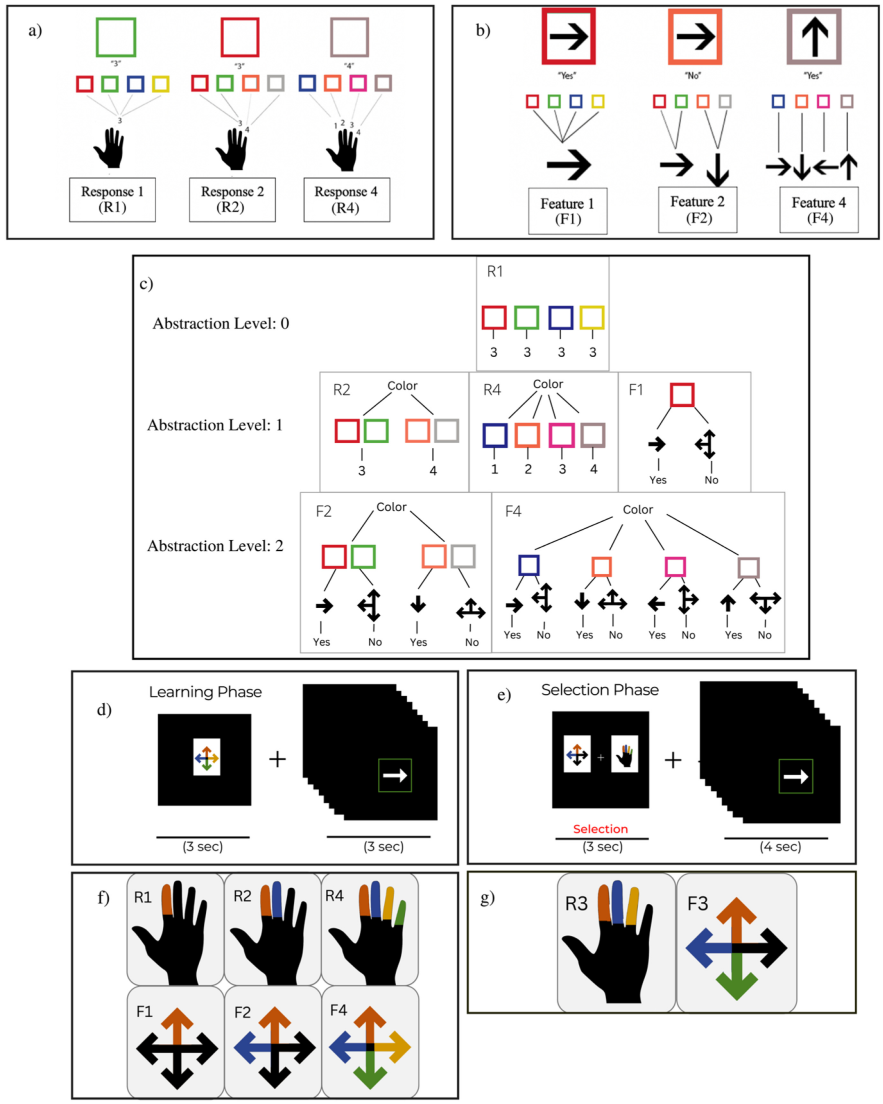
Policy Abstraction in Experiment 1. Tasks were coded for policy abstraction level by the depth of the hierarchical decision tree required to complete the task (i.e., the number of contextual contingencies), as shown in Figure 2f. Three levels of abstraction were used in the experiment.

In one condition, R1 was assigned an abstraction level of 0, as there was no contextual contingency about the response to be made; participants pressed the same key regardless of what was presented on the screen. Three conditions were assigned to an abstraction level of one, as there was only one contextual contingency. Crucially, two of these conditions were encountered during the response task (R2 and R4) and one during the feature task (F1). This allowed us to distinguish superficial features related to the task type (response or feature) from the abstraction level. The final conditions, F2 and F4, were assigned an abstraction level of two, as each had two relevant contextual contingencies. Specifically, the color cued the target arrow direction, and the direction of the arrow determined the response.

In addition to abstraction, we define progressive changes in choice options within each abstraction level as the number of options. Specifically, there were three response options among the conditions associated with each type of task. R1 and F1 each had one response option, R2 and F2 each had two, and F4 and R4 each had four. Note that some of these changes in the number of options, such as from R2 to R4 and F2 to F4, do not increase the level of policy abstraction. Furthermore, as noted by Badre and D'Esposito (2007), the F1 and R2 blocks are theoretically analogous in terms of the number of options and abstraction level, as each one requires a choice between two responses based on one contextual contingency.

The six experimental conditions were presented in separate blocks; each block consisted of trials only in that condition. Each condition was identified by a particular symbol that indicated which condition would govern performance in a particular block of trials. Participants were shown the symbol associated with the forthcoming condition centered on a black background prior to

Figure 2
Experiment 1 Method



Note. (a) The response task consisted of three conditions in which the number of responses was manipulated. In every trial, a colored box appeared on the screen, and the button was cued to be pressed. (b) The feature task consisted of three conditions that manipulated the number of sets of stimulus-response mappings. In each trial, a colored box containing an arrow appeared. The color of the box cued the target direction of the arrow so that the response mapping set was relevant. (c) Policy abstraction in Experiment 1. Condition R1 was at the 0th level of policy abstraction, as no competition was required. The R2 and R4 conditions from the response task and the F1 condition from the feature task are at the first level of policy abstraction, as they each had one level of branching on the presented tree. The F2 and F4 conditions were at the second level of policy abstraction, as there are two levels of branching in each condition. (d) Learning phase schematic. The task condition symbol was presented first, followed by the participants engaging in 12 practice trials. (e) Selection phase schematic. First, the participants were presented with two task symbols, and they chose the task they would rather execute. The participants then executed the task associated with the symbol. (f) Symbols used in Experiment 1. Top row: these symbols were associated with response task conditions, with far left representing the first number of options, middle, the second number of options, and far right, the third number of options. Second row: these symbols were associated with the feature task conditions. Far left represents the first number of options, middle, the second, and far right, the third. (g) Novel symbols. Left: novel symbol for the response level, representing an unlearned rule set. Right: novel symbol for the feature level, representing an unlearned rule set. Hand icons used in this figure are a mirrored version of "Silhouette hand" by Simon Waldherr, which is licensed under CC BY-SA 3.0. To view a copy of this license, visit https://creativecommons.org/licenses/by-sa/3.0/deed.en. R = response; F = feature. See the online article for the color version of this figure.

performing a block of that condition. The participants received no feedback regarding their accuracy throughout the experiment.

DST. The six conditions of the response and feature tasks were incorporated into a DST procedure to measure participants' tendency to avoid tasks that differed in the variables of error likelihood, time-on-task, number of options, and policy abstraction. There were two phases to the DST: learning phase and selection phase.

Learning Phase. In the learning phase, participants were presented with feature and response tasks in a randomized order. Within each task, participants were given the rules for three distinct conditions in random order by listening to the experimenter read from a prewritten script describing the rules. For each task type, the participants learned the rules for the three distinct conditions in a random order. They then practiced each condition twice for a total of six practice blocks per task. In the first practice set, the rules associated with the condition were presented on the screen. In the second set, the rules were not presented on the screen, and participants had to remember the rule. Each block consisted of 12 trials. The trials were separated by a 2 s fixation cross, and there was a response deadline of 4 s (Figure 2d). Each task was associated with a unique symbol. At the beginning of each block, the participants viewed the symbol associated with the task condition to be performed on the upcoming block (Figure 2f). In this way, participants learned which unique symbol cued each task. After learning and practicing the rules for the three conditions in either the feature or response task, the participants completed the same process for the other task.

Selection Phase. After completing the learning phase, participants immediately performed the selection phase. Participants were presented with two symbols, each representing one of the six different task conditions, and were asked to choose the condition block they would like to perform next. The symbols for each condition were presented on the left and right sides of the screen an equal number of times. Participants had a maximum of 4 s to make a decision. The unique combination of the six conditions (F1, F2, F4, R1, R2, R4) resulted in 15 possible pairs, each of which was presented 10 times, resulting in 150 random pairwise selections divided into 15 blocks with optional breaks between them (Figure 2e). The participants performed the selected task according to their decisions.

Finally, at the end of the 15 blocks of the selection phase, participants chose between two novel stimuli, intended to represent a third (and previously unexperienced) level of the response and feature tasks (R3 and F3; see Figure 2g). This generalization test allowed us to test the choice behavior based on the inferred experience of these task types. We reasoned that if people made decisions about tasks in which they have no prior experience, based on the inferred policy abstraction level of those tasks, participants would prefer a novel-response task over a novel-feature task when those options were paired. This decision pair was presented in 10 paired choices without performing the task after each choice, although participants were instructed that there may be execution of whichever rule they chose based on one of those choices.

Postexperimental Questionnaire. After completing both sessions of the experiment, the participants were asked several inventory questions, including whether they preferred certain task conditions and how those preferences emerged. Participants were asked to rank the relative difficulty of each task condition as represented by the symbol associated with each condition. Participants also completed the Need for Cognition Scale (Cacioppo & Petty, 1982) to identify individuals who enjoy engaging with cognitively challenging materials in their everyday lives.

Data Analysis

Basic task performance in both the learning and selection phases was analyzed in terms of RT and error rate. Task performance was analyzed using within-subject factors of abstraction level, number of options, and phase, referring to the learning or selection phase. Analyses were conducted using repeated measures of analysis of variance (ANOVAs; aov() command in R), where the subject number was entered as a random effect. Differences in task conditions were assessed using overall selection rates, defined as the probability of selecting a condition out of all the times that the condition was presented as an option within the selection phase. Analysis of selection rates was conducted using ANOVAs with within-subject variables of abstraction and number of options. Significant interactions were followed by simple effects analysis, and results were presented with false discovery rate (FDR) correction. All analyses used .05 alpha levels, and all plotted error bars represented within-subject standard error of the mean (*SEM*). Effect sizes were reported as partial eta squares, unless one-way ANOVAs were performed, in which case partial eta squared is equivalent to eta squared, and eta squared was reported.

Pairwise analyses used the number of times a condition was chosen when paired with a condition of a higher abstraction level. If the two conditions were at the same abstraction level (e.g., both R2 and R4 had an abstraction level of 1), then the comparison referred to the condition with a lower number of options. For example, the pairwise selection of R2/R4 would indicate how often the R2 condition was selected out of the 10 presentations of this combination. Tasks R2 and F1, which had equivalent difficulty levels, were defined as the proportion of times when task R2 was chosen. Pairwise analyses conducted using *t*-tests were performed with Bonferroni multiple correction. The effect sizes were reported as Hedge's *g* to account for the small sample size. Given the small sample size, we ran a Bayesian factor analysis using the BayesFactor package in R. Bayes factors (BFs; De Santis, 2004; Morey & Rouder, 2022; Weiss, 1997) that are between 1–3 and 1/3–1 are considered to be anecdotal/inconclusive evidence for either direction of the test. BFs that are between 3–10 and 1/10–1/3 are considered moderate, between 10–3 and 1/30–1/10 are considered strong, 30–100 and 1/100–1/30 are considered very strong, and >100 and $<1/100$ are considered extreme evidence.

To identify the combination of explanatory variables that best accounted for the variance in avoidance behavior, we used a full information criterion-based approach to select a best-fitting general linear model (GLM). This approach avoids the limitations of common stepwise approaches to variable selection by identifying the best-fitting GLM, providing assessments of uncertainty, and permitting multimodel inference (Buckland et al., 1997; Calcagno & de Mazancourt, 2010; Johnson & Omland, 2004). Owing to the small sample size, we used the corrected Akaike's information criterion (AICc) in the glmulti-model search. AICc is a variant of the AIC, which corrects model fits for small sizes and the ratio of the sample size to the number of predictors (Anderson & Burnham, 2002). We used the glmulti function in R to determine the relative importance of each variable across all possible models using a genetic algorithm (Calcagno & de Mazancourt, 2010). The glmulti function was used to compare all possible GLM models from the input candidate explanatory variables and pairwise interactions and calculate the AIC for each model. To unbias the model search, we performed list-wise deletions (i.e., if one of the variables, such as ER, yielded fewer observations than other variables owing to task selection rates, then the row numbers of other variables were matched to the observation number of that variable). To observe the impact of candidate variables across many models, glmulti was used to determine the model-weighted importance, confidence intervals (CIs), and estimates for each explanatory variable. The variables included in the model as potential predictors of the selection rate were abstraction, number of options, selection phase RT, selection phase error rates, and all possible pairwise interactions. The subject number was used as a random effect in all models. Response times were scaled across the tasks. Model comparison was performed using a Type I sequential ANOVA (the anova() command in R); list-wise deletion was performed when necessary to compare across models with different variables. We also repeated the same analysis with learning phase performance measures in Results 2 in the online supplemental materials and reported no qualitative differences between outcomes. Data are available on request from the authors.

Results

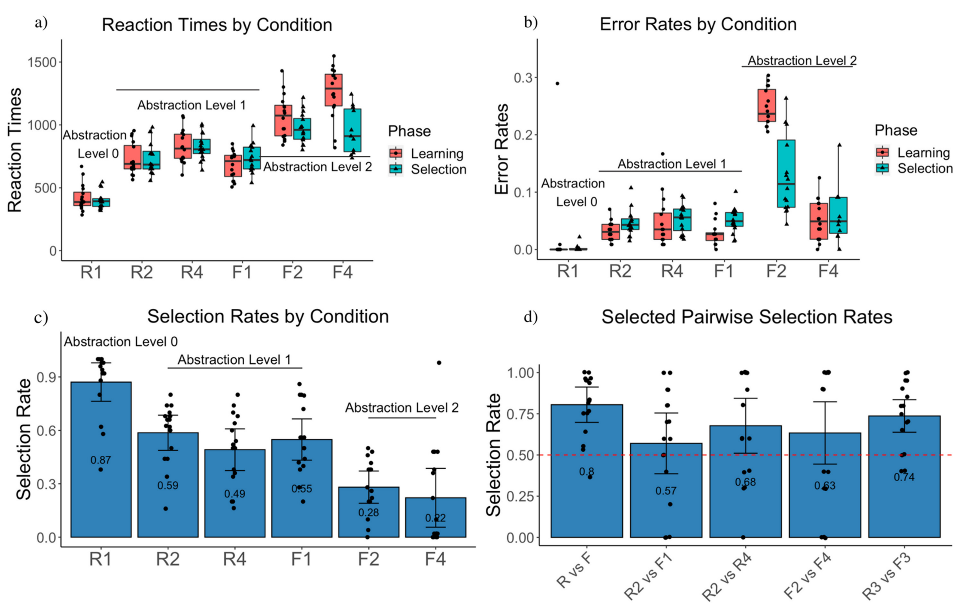
The Effect of Task Levels on Task Performance

As seen in Figure 3a, RT increased in both phases as a function of abstraction and the number of options within each abstraction level, replicating observations from previous research (Badre & D'Esposito, 2007). We tested these observations using a 2 × 3 × 3 within-group (Phase × Number of Options × Abstraction) repeated-measures ANOVA. Abstraction had a significant impact on RT, $F(2, 175) = 258.0$, $p < .001$, $\eta_p^2 = .75$. The number of options, $F(2, 175) = 11.8$, $p < .001$, $\eta_p^2 = .12$ and phase, $F(1, 175) = 8.0$, $p = .005$, $\eta_p^2 = .04$, also significantly affected RT. There was a significant interaction between phase and abstraction, $F(2, 175) = 11.3$, $p < .001$, $\eta_p^2 = .11$, such that RT were longer in the learning phase than in the selection phase for higher abstraction conditions, and a three-way interaction between abstraction, phase, and the number of options, $F(1, 175) = 4.2$, $p = .04$, $\eta_p^2 = .02$, such that conditions with a higher number of options took longer than conditions with a lower number of options in learning than selection phases, but only at higher abstraction levels.

As in Badre and D'Esposito (2007), ER (Figure 3b) followed a less clear trend across levels of abstraction and number of options. Overall, errors were more likely to occur at higher levels of abstraction, $F(2, 175) = 106.9$, $p < .001$, $\eta_p^2 = .55$. Error rates also differed with the number of options, although a monotonic increase or decrease across levels was not evident, $F(2, 175) = 30.2$, $p < .001$, $\eta_p^2 = .26$, and this was driven by high ER in the F2 condition. While there was no difference in ER for the learning versus selection phases, $F(1, 175) = 1.5$, $p = .2$, $\eta_p^2 = .01$, there was a significant interaction between abstraction and phase, such that participants were less accurate during the

Figure 3

Results of Experiment 1



Note. (a) Response times for the learning and selection phases by condition. Conditions are broken into abstraction levels. (b) Error rates for the learning and selection phases, by condition. Conditions are broken into abstraction levels. (c) Overall selection rates for each condition, with a significant main effect of condition. Conditions are separated by the abstraction level. (d) Selected pairwise avoidance rates. Far left: selection rates for the overall tasks; middle left: control condition (R2 vs. F1); and right: the two abstraction levels. Individuals generally chose response tasks over feature tasks but were not significantly different from 50% when comparing R2 to F1. The likelihood of choosing an easier option within the abstraction level was not significantly different between the two abstraction levels. The value of 50% is represented by a dashed line. In all the graphs, the dots represent the mean values for each participant. R = response; F = feature. See the online article for the color version of this figure.

learning phase at higher abstraction levels, $F(2, 175) = 9.2$, $p < .001$, $\eta_p^2 = .10$. There was also a significant interaction between both phase and number of options, $F(2, 175) = 9.2$, $p < .001$, $\eta_p^2 = .10$, such that ER were higher for a lower number of options to a greater degree in the learning than selection phases, and abstraction and number of options, $F(1, 175) = 98.7$, $p < .001$, $\eta_p^2 = .36$, also seemingly driven by the F2 condition. Finally, there was an observed three-way interaction between abstraction, phase, and the number of options, $F(1, 175) = 16.2$, $p < .001$, $\eta_p^2 = .09$, where participants were less accurate in conditions with a lower number of options in learning than in selection phases, but only at higher abstraction levels.

Hierarchical Switch Costs. As previously mentioned, previous research has found that switch costs are greater at higher levels of policy abstraction, which provides some evidence that policy abstraction affects multitasking. To verify whether this observation holds in the present manipulation of policy abstraction, we tested for the presence of hierarchical switch costs in Experiment 1.

Specifically, the F2 and F4 conditions of the feature task can have two hierarchical levels of switching. The lower level response switches, as in a change across trials from one key press to another versus repeating the same key press response. Second, there is a higher level switch when the direction of the target arrow changes from one trial to the next. Thus, we predicted larger switch costs for higher order target switches versus lower order response switches under these task conditions. Consistent with our expectations, we observed that the target switch cost RTs were significantly larger than response switch cost RTs in both feature task conditions with two abstraction levels, F2: $t(15) = 5.41$, $p < .001$; F4: $t(15) = 6.32$, $p < .001$. Thus, consistent with other studies of hierarchical task structure, higher order policy abstraction had a negative effect on multitasking, as measured by the switch cost.

Task Selection Rates

As predicted, the selection rates decreased with increasing abstraction. A 3×3 within-group (Number of Options $\times$ Abstraction) ANOVA on selection rates verified a significant main effect of abstraction, $F(2, 90) = 51.9$, $p < .001$, $\eta_p^2 = .54$, on the selection rate but no significant main effect of the number of options, $F(2, 90) = 1.18$, $p = .31$, $\eta_p^2 = .03$, or interaction between abstraction and number of levels, $F(2, 90) = .1$, $p = .73$, $\eta_p^2 = .001$. The overall selection rate is shown in Figure 3c.

Pairwise analyses allowed us to examine direct choices between specific pairs of conditions that differed in their level of policy abstraction or number of options while keeping the other variables of interest constant. Pairwise analyses revealed that only pairs that differed in terms of abstraction levels influenced effort selections. For example, pairwise selection rates for R1 (which had an abstraction level of 0) were significantly different from both R2, $t(15) = 5.79$, $p < .001$, Hedge's $g = 1.54$, BF $= 719.40$, R4, $t(15) = 5.56$, $p < .001$, Hedge's $g = 1.86$, BF $= 489.20$, and F1 conditions, $t(15) = 5.56$, $p < .001$, Hedge's $g = 1.59$, BF $= 488.60$, all of which had an abstraction level of 1. Similarly, the F1 condition, which had an abstraction level of 1, was different from the F2, $t(15) = 5.71$, $p < .01$, Hedge's $g = 1.38$, BF $= 619.93$, and F4, $t(15) = 3.10$, $p = .05$, Hedge's $g = 1.28$, BF $= 7.04$, conditions, which had abstraction levels of 2.

However, we did not find evidence that decisions were influenced by other differences between tasks beyond the level of abstraction. The R2/F1 conditions were at the same abstraction level but used different types of tasks with different task features (stimuli, rules, and instructions). These did not differ in choice preference. Their pairwise selection was not significantly different from 50%, $M = .56$, $SD = .34$; $t(15) = .52$, $p = 1.0$, Hedge's $g = .19$, BF $= 0.26$.

Analysis of pairwise selection rates also allowed us to explore the effect of the number of options on task avoidance. We analyzed pairwise selection rates when F2/F4 (both at the second abstraction level) and R2/R4 (both at the first abstraction level) were pitted during decision making. The results showed that participants were equally likely to choose R4 when paired with R2, $t(15) = 2.21$, $p > .05$, and F4 when paired with F2, $t(15) = 1.32$, $p > .05$. Notably, both pairs shared the same abstraction level, but differed in their number of options. We further calculated the BF for the null results. However, this analysis is inconclusive regarding the effect of the number of options. Specifically, there was anecdotal evidence in favor of the presence of an effect for choosing R2 over R4 (BF $= 1.71$) and anecdotal evidence in favor of an absence of effect for choosing F2 over F4 (BF $= 0.53$). Although the results for the number of options were weak, we included this variable in the model selection search described in The Effect of Task Variables on Effort Selections section to further test the influence of this variable.

Next, we tested individuals' overall likelihood of choosing the unexperienced conditions of the response task (R3) over the unexperienced conditions of the feature task (F3) at a higher policy level. This was calculated to determine the participants' response bias based on their inferred experience of these task types. Indeed, participants preferred R3 to F3, though they had not performed these tasks, response, $M = .88$, $SD = .18$; feature, $M = .09$, $SD = .16$; $t(15) = 9.53$, $p < .001$, Hedge's $g = 4.48$, BF $= 151{,}563.8$. This suggests that participants avoided symbol cueing in a task with greater policy abstraction, even in the absence of direct task performance. The five selection rates of interest are shown in Figure 3d.

Finally, it is possible that participants' selections were driven by their preference for one of the task features, such as arrows versus fingers, which differed between the feature and response tasks. To test control for this, we ran a one-sample t test to compare the paired selection rate of choosing R2 over the F1 task against chance (0.5). These conditions have equivalent policy abstraction, but differ in all other respects as the response and feature tasks. Thus, if participants preferred the response task to the feature task (or vice versa) for any reason related to the tasks themselves other than policy abstraction, then this would be evident in a paired selection that differed from zero.

This one-sample t test was not significant and so did not find evidence that participants were more likely to choose R2 over F1 (or vice versa), $t(15) = 0.21$, $p > .05$. The results showed that there is moderate evidence in favor of an absence of effect for choosing R2 over F1 (BF $= 0.26$). On the other hand, participants preferred the response task over feature task when policy abstraction differed between tasks. As such, participants significantly preferred R2 over F2, $t(15) = 8.86$, $p < .001$. BF for this effect was 64,592.14, indicating extreme evidence for a selection difference between tasks.

The Effect of Task Variables on Effort Selections

We next assessed the impact of candidate variables in task selection by using the glmulti function in R, and glmulti function determines the relative importance of each variable across all possible models using a genetic algorithm method (Calcagno & de Mazancourt, 2010). As such, glmulti function compares all possible GLM models from input candidate explanatory variables and pairwise interactions and calculates the AIC for each model. The variables included in the model as potential predictors of the selection rate were abstraction, number of options, selection phase RT, selection phase ER, and all possible pairwise interactions. The model selection results using the glmulti function on the candidate variables are presented in Table 1. Abstraction was the variable with the highest importance value (model-weighted importance: .976) and was included in 512 of the 1,024 models explored in an exhaustive search. Abstraction had a model-weighted β of −.18, indicating that as abstraction increased, individuals' selection rates decreased. Error rate was the variable with the second-highest importance, with an importance value of .713, and was also included in 512 of the 1,024 models. Error rate had a β of −.80, indicating that as errors increased, selection rates decreased.

Model comparison was conducted, in an iterative manner, to determine how variables with importance values above .5 provided explanatory power. Initial analysis suggested that abstraction explained a significant proportion of variance associated with selection rates, $F(1, 91) = 90.00$, $p < .001$. Adding ER to this model explained significantly more variance in selection rates, $\chi^2(2) = 9.04$, $p = .01$, suggesting that ER provide a contribution independent of abstraction.

Finally, in order to isolate the influence of number of options in choice behavior, we devised a new model where we predicted pairwise task selection rates (15 per participant) by the difference in number of options, abstraction, RT, and ER between the paired tasks. In other words, we aimed to predict each pairwise selection rate (e.g., R1 vs. R4) by the corresponding difference between the tasks (e.g., abstraction difference = 1; number of options difference = 3; RT difference and ER difference).

The results for the model selection of mixed-effects linear regression are summarized in Table 2. Abstraction difference was the variable with the highest importance (model-weighted importance: .99) and was the only variable with an importance variable above .5. These results indicate that when two tasks were paired with each other, participants made their decisions based on the policy abstraction difference between tasks.

Subjective Ratings

After completing both sessions of the experiment, participants were asked to rate the subjective difficulty of each task condition. Analyses were conducted using repeated measures of ANOVAs (aov() command in R) where dependent variable was subjective rankings, the predictors were policy abstraction level of the task and the number of response option the task yielded. Subject number was entered as a random effect. Both policy abstraction and number of options were predictors of subjective ratings of difficulty, main effect of abstraction, $F(2, 167.12) = 74.29$, $p < .001$, $\eta_p^2 = .64$, main effect of number of options, $F(2, 7.65) = 3.40$, $p = .04$, $\eta_p^2 = .074$, but no interaction between abstraction and number of levels, $F(1, 0.15) = 0.13$, $p = .72$, $\eta_p^2 = .002$. Note that number of options was not a significant predictor of effort choices, despite being a significant predictor of subjective ratings.

Discussion

This first experiment sought to test whether policy abstraction drove task avoidance over and above the contribution of other factors such as task performance (RT and ER) or the number of options within an abstraction level. We investigated this question by combining two levels of a paradigm previously established to explore policy abstraction (Badre & D'Esposito, 2007) with a DST procedure.

As expected, both higher policy abstraction and number of options were associated with slower RT and larger ER. Despite these performance costs, mixed-effects model selection results showed that policy abstraction was the best predictor of overall demand selections. Across two separate model searches, we assessed the influence of policy abstraction, number of options, and task performance on task selection rates. The first model tested the effect of the candidate variables in overall selection rates. The second model tested the effect of relative value difference in candidate variables in pairwise selection rates. Both model results showed that policy abstraction explained selection rates over and above the effect of number of options or task performance. Subjective difficulty ratings showed a significant effect of both policy abstraction and number of options, indicating that number of options influenced subjective difficulty rankings despite not influencing task selections. Finally, we observed that avoidance of higher policy tasks occurred even in the absence of direct experience with a task. In particular, with

Table 1
Overall Selection Rates: Estimate, Number of Models, Confidence Interval Upper and Lower Bounds, and Importance for Each Variable Input Into glmulti for Consideration

Variable	Estimate	No. of models	CI *LL*	CI *UL*	Importance
Abstraction	−0.1804	512	−0.2905	−0.0704	0.9763
ER	−0.8035	512	−2.1294	0.5525	0.7134
RT	−0.0272	512	−0.1002	0.0457	0.4917
Number of options	−0.0057	512	−0.0306	0.0192	0.2862
Number of options: ER	−0.1395	62	−0.5715	0.2926	0.2447
Abstraction: number of options	−0.0004	448	−0.0092	0.0100	0.1762
Number of options: RT	0.0033	184	−0.0158	0.0091	0.1553
Abstraction: RT	−0.0009	304	−0.0066	0.0048	0.0779
Abstraction: ER	0.0193	116	−0.1185	0.1574	0.0595
ER: RT	0.0029	36	−0.0177	0.0235	0.0184

Note. CI = confidence interval; *LL* = lower limit; *UL* = upper limit; ER = error rates; RT = response time.

Table 2

Pairwise Selection Rates: Estimate, Number of Models, Confidence Interval Upper and Lower Bounds, and Importance for Each Variable Input Into glmulti for Consideration

Variable	Estimate	No. of models	CI *LL*	CI *UL*	Importance
Abstraction difference	−0.1533	512	−0.2223	−0.0843	0.9987
Number of options difference	−0.0097	512	−0.0445	0.0251	0.4610
RT difference	−0.0020	384	−0.0142	0.0103	0.1805
ER difference	0.0479	320	−0.1343	0.2300	0.1761
Abstraction difference: number of options difference	−0.0024	256	−0.0118	0.0069	0.1265
Abstraction difference: RT difference	−0.0003	128	−0.0019	0.0013	0.0257
Number of options difference: RT difference	−0.0002	72	−0.0009	0.0006	0.0185
Abstraction difference: ER difference	0.0003	40	−0.0023	0.0030	0.0053
ER difference: number of options difference	−7.50E−6	22	−0.0005	0.0005	0.0026
ER difference: RT difference	0.0003	12	−0.0011	0.0017	0.0019

Note. CI = confidence interval; *LL* = lower limit; *UL* = upper limit; RT = response time; ER = error rates.

unexperienced R3/F3 choice, the participants avoided the feature task that was associated with higher abstraction levels, despite their lack of direct experience performing those tasks. This suggests that the cost of policy abstraction may be generalized to new tasks based on the predicted task structure, rather than experiencing the task. We return to this observation and its implications in the overall discussion.

Experiment 2

Experiment 1 found evidence that policy abstraction influenced effort avoidance, more so than global effects of performance and the number of options. Furthermore, we observed that this effect was obtainable even when people had not directly experienced the effort associated with a given task. However, these experiments used tasks that directly differ in terms of policy abstraction, and so it is difficult to be certain whether other uncontrolled task features that correlated with abstraction drove this effect.

To address this limitation and replicate the effect observed in Experiment 1, Experiment 2 tested whether the relationship between policy abstraction and effort avoidance generalizes to a different set of tasks that can be analyzed with respect to policy abstraction. To do so, we adopted a paradigm from Nee and D'Esposito (2016, 2017; Nee, 2021). This task includes four distinct cognitive tasks that differ progressively in their degree of policy abstraction. Evidence from functional magnetic resonance imaging (fMRI) and transcranial magnetic stimulation has associated these tasks with hierarchically organized networks that include lateral PFC (Nee & D'Esposito, 2016, 2017; Nee, 2021). These results have prompted a revised view of where the "top" of the prefrontal hierarchical organization might be, associating it with a network including the anterior mid-dorsolateral PFC rather than the most rostral frontal pole. Nevertheless, in terms of the functional distinctions being drawn, the pattern of results from this task paradigm largely agrees with the results of Badre and D'Esposito (2007) (as reviewed in Badre & Nee, 2018). Thus, it is reasonable to hypothesize that though the tasks differ in their details, in terms of the instructions, stimuli, rules, and responses being made, they share a similar underlying structure in terms of changes in policy abstraction across levels. As such, the Nee and D'Esposito task offers an opportunity to test the generalizability of the relationship between policy abstraction and effort avoidance. Furthermore, this task does not require distinguishing as many task conditions as that used in Experiment 1, and so we can cue each condition with an arbitrary symbol. As such, we can also control for any effect that the task cue itself might have had on avoidance rates in Experiment 1.

Method

Subjects

The same 19 participants who completed Experiment 1 took part in Experiment 2. Participants completed both Experiments in a randomized order (see Results 3 in the online supplemental materials for the analysis of order effects), either on the same day or across 2 days at the same time of day. Participants were debriefed following the completion of both studies. One participant's data were removed due to extreme deviation from mean RT and accuracy across all conditions prior to the analysis of task selection rates; this participant was also removed from Experiment 1. As such, 18 participants were included in the analysis for Experiment 2. All behavioral tests were conducted according to the procedures approved by the Human Research Protection Office of Brown University. The participants provided informed consent in accordance with the Research Protection Office of Brown University. All participants were compensated by course credit or monetary payments at a rate of $10/hr.

Procedure

Experiment 2 was based on the procedure outlined in Nee and D'Esposito (2016). The experimental procedure from that fMRI study was adapted to a DST to explore the effects of policy abstraction on effort avoidance.

As in Experiment 1, the study consisted of learning phase and selection phase. In the learning phase, participants learned the association between unique task identifiers and the different task conditions. In the selection phase, participants chose between two symbols and then performed the task associated with their choice.

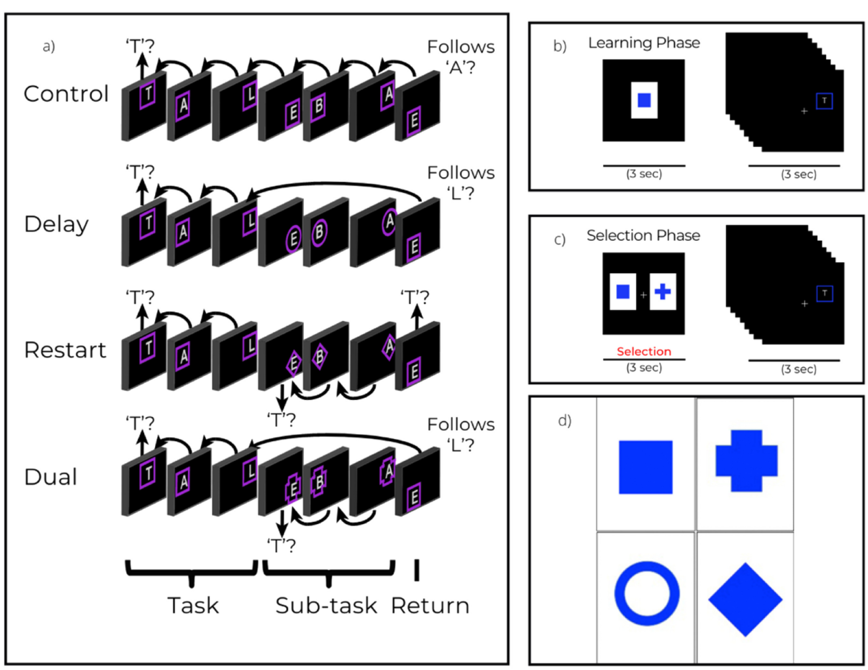
Task Conditions in Experiment 2. There were two task modalities: (a) spatial and (b) verbal. For the baseline spatial sequencing tasks, participants were required to determine if the position of a stimulus (a box with a letter inside of it) followed the previous stimulus as though one were tracing the points of a star in a clockwise direction. The baseline verbal sequencing tasks required participants to determine if the letter inside the box followed the previous letter in the word "tablet." For the first stimulus in a sequence, the participant's response indicated whether it was the letter "t" in tablet or the first position in drawing the star (Figure 4a), depending on whether the block was verbal or spatial. Participants were cued to perform the verbal or the spatial task according to the color of the stimuli within the task. Each block contained trials of only one task modality. The modality was unknown prior to commencing in each task block, so modality was not a relevant feature in the task selection phase (note, however, that Results 4 in the online supplemental materials includes a coding scheme that considered modality as an additional level of abstraction and shows that the results are robust to differences in modality). Thus, as participants needed to consider two contexts (evaluate both the current stimulus and the stimulus immediately prior) to determine if the present stimulus was "correct," the baseline task was assigned to a policy abstraction level of 2.

Within each modality, there were four different task conditions that were organized into blocks: (a) control task, (b) delay task, (c) switch task, and (d) dual task. In each block, participants either performed the baseline task the entire way through (control), or performed the baseline task for two to five trials, one of three subtasks for three to five trials (delay, switch, or dual), and returned to the baseline task for another two to five trials. Participants were cued to begin the subtask and return to the baseline task by the box surrounding the stimulus changing shape. Importantly, subtasks differed from the control task in their level of policy. However, as selection decisions were made about the task condition as a whole (i.e., the whole block rather than only the subtask), we coded the policy abstraction for each task condition as an average of the policy level for the control task and the subtask.

In the delay condition, participants were asked to remember their position in the visual or spatial sequence immediately prior to the start of the subtask. Within the subtask of the delay condition, participants responded to all stimuli with the "no" key, regardless of if the stimulus followed the previous stimulus or not. Upon returning to the basic task, participants continued the sequence from where they had left off prior to the subtask (Figure 4a). Since participants were required to hold three contexts during the baseline task (the shape, the current stimulus, and the prior stimulus) but only two tasks during the subtask (the shape and the current stimulus), the delay condition was considered to have asymmetric mapping and was operationalized as abstraction level 2.5 (control task policy = *3*, subtask policy = *2*).

Figure 4

Experiment 2 Method



Note. (a) The four conditions, with example stimuli for the baseline, subtask, and return to baseline in the verbal modality. Reprinted from "The Hierarchical Organization of the Lateral Prefrontal Cortex," D. E. Nee and M. D'Esposito, 2016, *eLife*, 5, Article e12112 (https://doi.org/10.7554/eLife.12112). © 2016, Nee et al. (b) Schematic of the learning phase. First, the symbolic representation is presented; next, participants engage in 8–11 trials. (c) Schematic of the Selection phase. Participants are presented with two symbolic representations, and they choose which they would rather execute from between the two. Then, participants execute the task associated with that representation. (d) The four possible symbolic representations. Top left: the control condition was always associated with a square. Others: the switch, delay, and dual conditions were each randomly assigned to one of these three symbols. Task figures in panel (a) were provided by Derek Nee (personal communication, June 8, 2023). See the online article for the color version of this figure.

In the switch condition, participants restarted the sequence at the start of the subtask, determining if the initial stimulus in the subtask was the initial stimulus in the sequence (the letter "t," for example), before continuing with the basic task. Upon the return trials to baseline, participants restarted the sequence for a second time (Figure 4a). As participants in both the subtask and baseline task needed to consider three contexts (if the shape changed, the current stimulus, and the prior stimulus), the switch conditions were assigned to an abstraction level of 3 (control task policy = *3*, subtask policy = *3*).

As with the switch condition, the baseline task in the dual condition has a policy abstraction depth of 3, in that participants must decide the task (based on shape) and then use the prior to decide how to interpret the current stimulus with respect to a response. The subtask is different, however. Participants need to maintain the item from the baseline task and then add a decision about which item in WM to use for as a context. Prior work has observed that such decisions within WM require additional contextually guided processing (Rac-Lubashevsky & Frank, 2021). Thus, this condition requires four contextually guided decisions, that is, (a) the shape determines the task, (b) the shape also determines which item in WM to use as the prior stimulus, and then (c) prior stimulus, contextualizes the current stimulus that provides the context (d) for a response. As such, this condition was operationalized with an aggregate abstraction level of 3.5 (control task policy = *3*, subtask policy = *4*).

Learning Phase. In the learning phase, participants learned the association between a unique symbol each task condition. For each of the eight task conditions (four conditions, with either a verbal or spatial task modality), participants completed eight practice blocks three times. The number of trials in each practice block was randomized between eight and 11. The intertrial intervals were jittered around 2 s, and the response deadline was 3 s (Figure 4b). The order of the presentation of the conditions and the mapping between symbol and task condition was counterbalanced across participants. Before each block, the participant was shown the symbol associated with that condition, centered on a black background, for 2 s. Participants received no feedback on their errors or RT during task performance.

Selection Phase. The selection phase immediately followed the learning phase. In the selection phase, all four tasks were paired with each other. There were six unique pairs of combinations, each of which was presented 10 times in a pseudorandomized order.

For each selection, two task symbols were presented, and participants were able to select a condition and then performed the condition they chose. Task symbols were presented on the left and right sides of the screen an equal number of times. Participants knew from the task symbol which task they would be executing, though they did not know the modality of that task. Participants were given 4 s to make their selection. After performing the task, a fixation cross was presented again for a jittered interval before the next task selection (Figure 4c). Participants completed this phase of the experiment in three blocks, and the presentation of pairs was determined randomly.

Data Analysis

Analysis of task performance focused on RT and ER. Responses given beyond the response deadline were excluded from further analysis. 6.2% of trials during learning and <1% of trials during the selection phase were excluded on this basis. Response times were calculated only on correct trials. The overall selection rate for each condition was the probability that it would be selected when presented paired with any of the other three conditions. As in Experiment 1, when multiple pairwise analyses were performed, *t* tests were conducted with the Bonferroni multiple correction applied. Effect sizes were reported using Hedge's *g* to accommodate small sample size.

Further data analysis was performed using three ANOVAs. When the dependent variable was ER or RT, within-factors were abstraction and phase, referring to learning or selection phases. When the dependent variable was the selection rate, the within-factor variable was abstraction level. These repeated-measures ANOVAs were conducted using the aov() command in R. The Greenhouse–Geisser correction was used if sphericity was violated. Significant interactions were followed by simple effects analysis, and results were presented with FDR correction. Analyses used .05 alpha levels, and error bars represented within-subject *SEM*. As in Experiment 1, effect sizes were reported as partial eta squares, unless one-way ANOVAs were performed, in which case partial eta squared is equivalent to eta squared, and eta squared was reported.

The same modeling package, glmulti, was used in this experiment as in Experiment 1. As there was no parametric variance in number of options in this paradigm, the variables that were input to the model were: abstraction, selection phase RT, selection phase ER, and all interactions thereof. Response times were rescaled. The model-weighted importance, CIs, and estimate of each variable were calculated using the glmulti function. The subject number was used as a random effect in all models. Model comparison was conducted using the anova() function in R, which conducts a Type I sequential ANOVA; listwise deletion was performed when necessary to compare across models with different variables. We also repeat the same analysis with learning phase performance measures in Results 2 in the online supplemental materials and report no qualitative differences between outcomes. Data are available on request from the authors.

Results

The Effect of Task Levels on Task Performance

Response times and ER data from both learning and selection phases, collapsed across both the subtask and baseline conditions, can be seen in Figure 5a and b. A 4 × 2 ANOVA (Condition × Phase) was conducted on RT data. Results were consistent with those observed by Nee et al. There was a main effect of abstraction on RT, as participants took longer to complete tasks as the abstraction level increased, $F(3, 127) = 15.3$, $p < .001$, $\eta_p^2 = .27$. Tukey-adjusted post hoc comparisons confirmed that there were significant differences between RT for all pairwise combinations except between the control and delay conditions (control vs. delay $p = .999$; control vs. restart $p = .037$; control vs. dual $p < .001$; delay vs. restart $p = .050$; delay vs. dual $p < .001$; restart vs. dual $p = .021$). There was additionally a main effect of phase, such that participants were slower in the learning than the selection phases, $F(1, 127) = 25.6$, $p < .001$, $\eta_p^2 = .17$. There was no interaction between phase and abstraction, $F(3, 127) = 1.1$, $p = .35$, $\eta_p^2 = .03$.

An additional 4 × 2 ANOVA (Condition × Phase) was conducted on ER. Abstraction had a significant effect on accuracy,

$F(3, 127) = 5.7$, $p < .001$, $\eta_p^2 = .12$, such that participants in more abstract conditions had higher ER. Tukey-adjusted post hoc comparisons indicated that there were only significant differences in accuracy for control versus dual and restart versus dual conditions (control vs. dual $p < .001$; delay vs. dual $p = .031$; all other $ps > .13$). Phase also significantly impacted accuracy, $F(1, 127) = 9.8$, $p = .002$, $\eta_p^2 = .07$, where participants in the learning phase had higher ER. There was no interaction between phase and abstraction, $F(3, 127) = .45$, $p = .72$, $\eta_p^2 = .01$.

Task Selection Rates

Selection rates across conditions are summarized in Figure 5c. The selection rates decreased as abstraction level increased, resulting in a significant main effect of condition on selection rate, $F(3, 68) = 74.6$, $p < .001$, $\eta^2 = .77$. Post hoc comparisons of the effect of condition on selection rate showed that all pairs of conditions were significantly different, with $p < .01$ and all $\mathrm{BF} > 20$. Notably, though post hoc analyses indicated that the control and delay, as well as switch and dual, pairs of conditions, did not differ in RT or ER, both pairs of conditions significantly differed in selection rate.

The Effect of Task Variables on Effort Selections

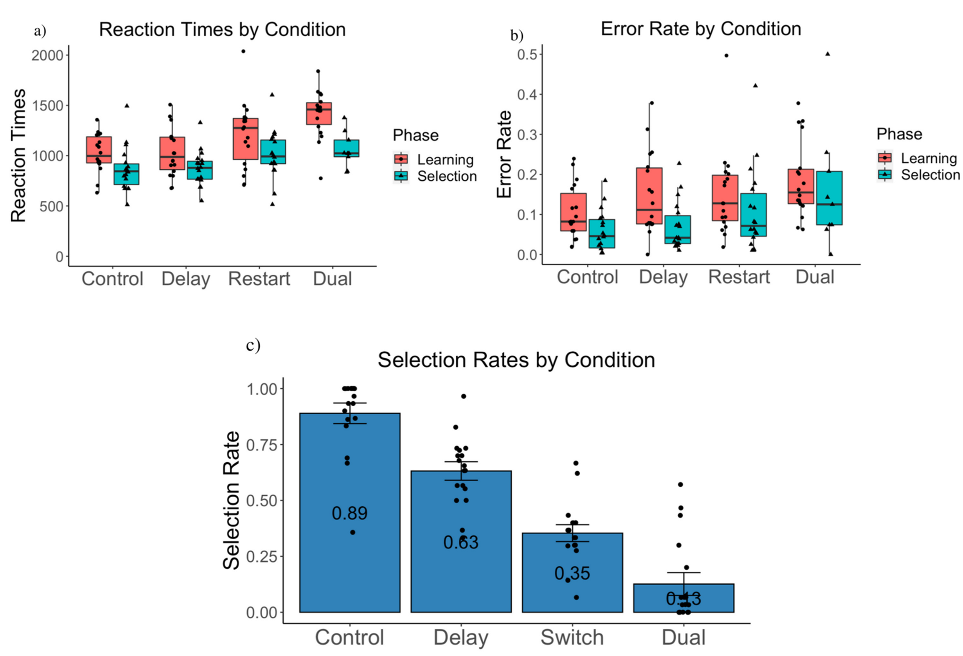
The results for the model selection of mixed-effects linear regression are summarized in Table 3. Abstraction was the variable with the highest importance (model-weighted importance: .92) and was the only variable with an importance variable above .5. Abstraction was included in 32 of the 64 models considered in an exhaustive search, and had a model-weighted β of $-.46$, indicating that as abstraction rates increased, participants' selection rates decreased. Unlike in Experiment 1, ER had a relatively low importance value (importance: .35). As no other variables had importance greater than .5, model comparison was not conducted.

One plausible alternative contributor choice behavior might be the demands placed on WM maintenance, as WM demands of the task might sometimes be correlated with policy abstraction, for example, because holding more contextual representations in WM is needed in order to traverse increasing levels of contingency. Thus, it can be hard to distinguish the number of items held in working memory from the level of policy abstraction.

However, another demand on WM beyond load is the duration information is held over a delay period, particularly in the presence of intervening items. Indeed, sensitivity to delay is comparable to WM load in the consistency of its effect on behavior, and it is a consensus benchmark of WM (Oberauer et al., 2018).

Though delay is a task demand related to WM, it is not correlated with policy abstraction. Thus, we can distinguish general WM demands from policy abstraction by asking whether holding contextual information longer in WM accounts for task avoidance. For example, in Experiment 2, the delay condition increases demands on WM maintenance without simultaneously increasing policy abstraction level of the task, as this task requires the maintenance of an item across 3–5 trials in the presence of intervening stimuli. On the other hand, restart condition increases the policy abstraction level of the task more than it increases WM demands, as this task requires switching between two task rules. In other words, delay task WM demands are always equal to or higher (3–5) than the policy abstraction level of the restart task (3). If participants aimed to reduce WM load and

Figure 5
Results of Experiment 2



Note. (a) RT for learning and selection phases, by condition. (b) ER for learning and selection phases, by condition. (c) Overall selection rates by condition. In all graphs, shapes represent average values for a participant. RT = response times; ER = error rates. See the online article for the color version of this figure.

Table 3

Overall Selection Rates: Estimate, Number of Models, Upper and Lower Confidence Interval Bounds, and Importance for Each Variable Input Into glmulti for Consideration

Variable	Estimate	No. of models	CI *LL*	CI *UL*	Importance
Abstraction	−0.4563	32	−0.6610	−0.2515	0.9211
RT	−0.0513	32	−0.2311	0.1284	0.4875
Abstraction: RT	0.0241	28	−0.0462	0.0944	0.4584
ER	0.1997	32	−0.4896	0.8889	0.3534
Abstraction: ER	−0.0375	20	−0.2263	0.1513	0.2476
Error rates: RT	−0.0094	11	−0.0537	0.0348	0.1055

Note. CI = confidence interval; *LL* = lower limit; *UL* = upper limit; RT = response time; ER = error rates.

not policy abstraction, they should prefer the restart more than the delay task. The effect size for the paired-samples *t* test comparing the overall likelihood of choosing the delay task and the overall likelihood of choosing the restart task was 1.92 (Cohen's *d*), indicating very high effect size. At this effect size, a replication study would require six people to attain a significant effect.

In order to test the potential influence of WM maintenance on effort selections, we scored each task based on the number of trials to maintain an item in working memory (WM delay) in addition to the policy abstraction each task requires, as before. In both the control and restart conditions, participants are not required to hold a task in mind for the duration of the subtask; as such, these conditions were assigned a WM value of 0. In both the delay and dual conditions, participants were required to remember the pre-subtask stimulus across the duration of the subtask (an average of four trials); these conditions were assigned a WM value of 4.

The results for model selection of mixed-effects linear regression with abstraction, WM, ER, and RT are summarized in Table 4. Abstraction was the variable with the highest importance (model-weighted importance: .96), followed by WM, which had an importance value of .93. Abstraction was included in 512 of the 1,024 models considered in an exhaustive search, and had a model-weighted β of −.49, indicating that as abstraction rates increased, participants' selection rates decreased; WM was included in 448 models and had a model-weighted β of −.05, indicating that as WM taxation increased, selection rates decreased.

Model comparison was conducted to determine the degree to which variables with importance values above .5 provided explanatory power. Abstraction alone explained a significant proportion of variance associated with selection rates, $F(1, 72) = 122.4$, $p < .001$. Adding WM to this model explained significantly more variance in selection rates, $\chi^2(2) = 33.2$, $p < .001$, suggesting that WM does provide a contribution independent of abstraction. However, this complementary analysis supported the interpretation that policy abstraction was the best predictor of demand avoidance.

Subjective Ratings

After completing both sessions of the experiment, participants were asked to rate the subjective difficulty of each task condition. Analyses were conducted using repeated measures of ANOVAs (aov() command in R) where dependent variable was subjective rankings and the predictor was policy abstraction level of the task. We observed a significant effect of task type (i.e., abstraction level) on subjective effort ratings, $F(3, 69.97) = 154.9$, $p < .001$, $\eta_p^2 = .89$, and Bonferroni corrected post hoc *t* test verified that subjective ratings for all tasks differed significantly from each other (all *p*s < .001). These results suggest that subjective ratings and objective preferences of effort are aligned when tasks differ in their level of abstraction.

Discussion

Consistent with the first experiment, results from Experiment 2 found that policy abstraction drives demand selections when these decisions concern making selections between different tasks that employ components of WM delay, task switching and a combination

Table 4

Working Memory as Additional Variable: Estimate, Number of Models, Upper and Lower Confidence Interval Bounds, and Importance for Each Variable Input Into glmulti for Consideration

Variable	Estimate	No. of models	CI *LL*	CI *UL*	Importance
Abstraction	−0.4856	512	−0.6252	−0.3460	0.9601
Working memory	−0.0531	448	−0.0832	−0.0230	0.9349
ER	0.1304	512	−0.3045	−0.5653	0.4158
RT	5.52E−06	512	−5.46E−05	6.57E−05	0.2761
Abstraction: RT	6.33E−06	352	−2.19E−05	3.46E−05	0.1997
RT: working memory	−2.09E−06	48	−1.05E−05	6.36E−06	0.0457
Abstraction: working memory	4.75E−06	76	−2.72E−03	3.37E−03	0.0443
Abstraction: ER	−0.0209	240	−0.0010	0.0058	0.0399
ER: RT	1.73E−03	136	−9.37E−06	1.28E−05	0.0375
ER: working memory	1.24E−06	29	−6.64E−03	4.17E−03	0.0268

Note. CI = confidence interval; *LL* = lower limit; *UL* = upper limit; WM = working memory; ER = error rates; RT = response time.

of both. Importantly, this second experiment used different tasks, procedures, and stimuli from the first experiments. For example, whereas Experiment 1 involved a more basic trial-based response selection task, this experiment used a continuous performance procedure. Nonetheless, both tasks could be analyzed in terms of policy abstraction, and this variable was related to task avoidance for both experiments. Thus, this experiment provides evidence of the generalizability of policy abstraction as a task feature that determines avoidance behavior.

Additionally, WM demands of the task might sometimes be correlated with policy abstraction. In order to disentangle the potential influence of WM maintenance from policy abstraction on effort selections, we scored each task as a function of the demands they place on WM during a delay period. The results of this analysis showed that WM maintenance is a significant predictor of task selections although the effect of policy abstraction was still larger. Subjective rankings also showed that participants' sense of difficulty increased with increasing abstraction level of the task. In our final investigation, we leverage this abstraction feature in order to directly test its influence across tasks.

Cross-Study Analysis

Though different in their details, both Experiments 1 and 2 provided evidence that policy abstraction was a significant predictor of effort avoidance. However, these analyses were conducted within each study and there were some differences, particularly in the contributions of and interactions with other variables, such as RT and ER. Importantly, the same participants performed both tasks, allowing us to perform a within-subject cross-study analysis to test whether individual participants completing two very distinct tasks showed the same avoidance behavior as a function of policy abstraction.

Data Analysis

To examine the effect of abstraction on selection rates across both tasks, the same mixed-effects model parameters using glmulti as described above were utilized. Reaction times were scaled by each experiment and task, as in Experiments 1 and 2. The same variables were utilized as described in the prior data analysis sections, and model-weighted variable importance, CIs, and estimate of each variable were calculated using glmulti. Importantly, in order to address the role of number of options across both experiments, we assigned a number of options variable to Experiment 2 data set. Although the number of options in Experiment 2 does not vary across task conditions, each task condition requires a yes/no response. Hence, we assigned all Experiment 2 task conditions 1 number of options. Additionally, we included a nuisance variable for each experiment type (i.e., factor variable with two levels: Experiment 1 vs. Experiment 2) to remove variance associated with separate experiments. Two participants who participated in Experiment 2 but not Experiment 1 were excluded from this analysis, and one participant who was already excluded from both studies, as previously noted, was again excluded, resulting in $N = 16$. Finally, we repeated the analysis by redefining the abstraction variable in the context of the paradigm used in Experiment 2, in order to account for the potential impact of task modality (spatial vs. verbal) on the hierarchy of task representations (see Results 4 in the online supplemental materials).

Results

The model selection results to determine the most important of the candidate variables in explaining the variance in selection rates are presented in Table 5. Abstraction had the highest importance value (model-weighted importance: 1.00), ER had the second-highest importance value (importance: .63), and the number of options had the third-highest importance value (importance: .52). Abstraction rates had a model-weighted β of $-.33$; ER had a model-weighted β of $-.35$, and number of options had a model-weighted β of $-.02$, indicating that as each of these variables increased, selection rates decreased. We conducted model comparison, iteratively adding both variables with an importance level greater than .5. Abstraction significantly explained variance in selection rates, $F(1, 147) = 33.8$, $p < .001$, and adding both ER, $\chi^2(2) = 13.7$, $p = .001$ and number of options, $\chi^2(2) = 27.1$, $p < .001$ significantly improved the variance explained by the predictor variables in selection rates.

Discussion

Consistent with the results of the first two studies, the combined analysis found that abstraction along with ER drove effort avoidance across various types of tasks. We also note that when both experiments were combined, the number of options explained additional

Table 5

Estimate, Number of Models, Upper and Lower Confidence Interval Bounds, and Importance for Each Variable Input Into glmulti for Consideration

Variable	Estimate	No. of models	CI *LL*	CI *UL*	Importance
Abstraction	−0.3296	1000	−0.4389	−0.2204	1.000
ER	−0.3506	768	−0.9829	0.2817	0.6281
Number of options	−0.0158	598	−0.0709	0.0394	0.5203
Abstraction: number of options	0.0139	96	−0.0281	0.0509	0.1479
Number of options: ER	−0.0077	6	−0.0381	0.0227	0.0185
Abstraction: ER	−0.0031	24	−0.0154	0.0092	0.0161
Abstraction: RT	−4.99E−5	64	−0.0003	0.0002	0.0062
Number of options: RT	−2.40E−5	16	−0.0001	0.0001	0.0020
RT	−1.99E−5	18	−0.0001	0.0001	0.0015
ER: RT	−0.0002	4	−0.0009	0.0006	0.0007

Note. CI = confidence interval; *LL* = lower limit; *UL* = upper limit; ER = error rates; RT = response time.

variance in avoidance rates. Hence, the results confirmed the significance of policy abstraction in driving effort costs across and its generality across different task designs.

General Discussion

It is well established that people avoid cognitively challenging tasks (Apps et al., 2015; Kool et al., 2010; Sayalı & Badre, 2019, 2021; Westbrook et al., 2013). Here, we sought to better understand the factors that make a task feel cognitively effortful, and accordingly lead to avoidance of a task. In particular, we tested the hypothesis that tasks involving higher degrees of policy abstraction will incur greater effort costs and will be avoided at higher rates, even while accounting for performance-related factors such as error likelihood and RT. Accordingly, the present experiments combined two tasks manipulating policy abstraction with a DST procedure. Across both studies and in one within-subject, cross-study analysis, we found support for our hypothesis. Higher policy abstraction not only predicted task avoidance, but it did so across tasks that differed in their specific instructions, stimuli, rules, and other details. Thus, we find evidence that policy abstraction is a task-general factor that predicts whether someone will avoid a task.

Importantly, however, tasks with higher order policy necessarily require following more complex rules and so often yield worse performance in terms of longer RT and higher ER. This common confound complicates clean inferences about the effect of policy abstraction on demand avoidance and cognitive effort. Thus, it is important to distinguish whether task differences in policy abstraction or performance drive differences in demand avoidance.

To that end, we selected tasks in which behavioral performance was not collinear with changes in policy abstraction, allowing us to delineate the influence of policy abstraction from that of task performance. We consistently observed that policy abstraction predicted task avoidance better than performance measures of error likelihood and RT. Indeed, people avoided tasks involving greater policy abstraction at higher rates than tasks with lower policy abstraction even in the absence of performance differences between the tasks, such as in the comparison of the control versus delay tasks in Experiment 2. This finding concurs with the observation that cognitive effort is not the same as error avoidance (Feghhi & Rosenbaum, 2021).

Policy abstraction was also a better predictor of demand avoidance than WM delay demands or the number of choice options within each abstraction level (i.e., the breadth of the decision tree), even though this latter manipulation increased task difficulty, affected performance, and influenced subjective effort ratings. Thus, these observations offer strong support for the hypothesis that policy abstraction, in terms of the number of contextual contingencies, influences task avoidance.

Why do people avoid cognitive tasks that involve higher order policy? Our hypothesis assumes that performing tasks with higher order policy abstraction reduces the capacity for multitasking. This diminished capacity is reasonable to assume given that (a) multitasking is itself a form of hierarchical control and so leverages the same capacity-limited system (Koechlin & Hyafil, 2007) and (b) the demands of control over multiple levels of contingency has established behavioral costs (A. G. E. Collins & Frank, 2016), an effect we reproduced in Experiment 1 of the present study. This reduced capacity for multitasking under increasing policy abstraction accordingly increases the opportunity cost of the control-demanding task (Kurzban et al., 2013; Musslick & Cohen, 2021). Consequently, tasks with higher policy abstraction are penalized with effort costs and are avoided at higher rates.

An interesting implication of this study is that factors that reduce policy abstraction should conversely reduce the experience of mental effort and task avoidance. One such factor might be repeated practice of a task that leads to automaticity. In particular, while people may use a hierarchical structure to perform a task initially, with repeated practice they have the opportunity to learn to behave according to a wider but flattened structure (Frank & Badre, 2012; Newell, 1994), and/or they can accumulate instances that combine multiple task features in an integrated nonhierarchical task representation (Logan, 1990). Thus, to the degree that doing so reduces the demand to control behavior at multiple levels of policy abstraction, it may be that such automated tasks, though operating under the same number of contingencies, would no longer be experienced as effortful. However, this hypothesis should be tested directly and contrasted with other changes in task structure or processing that also accompany practice and automation.

The observation that policy abstraction is a determinant of task avoidance is broadly consistent with the proposal that tasks requiring cognitive control are particularly effortful (Shenhav et al., 2013, 2017). Most tasks that demand cognitive control require an internally maintained goal, rule, or other contextual element to provide a control signal over the appropriate course of action. As such, tasks requiring control demand at least some degree of policy abstraction. However, our results add specificity to this view.

For instance, we did not find that distinguishing between more options at a particular level of contingency caused greater avoidance behavior, even though having more options should result in greater choice interference, conflict, and therefore more demands on top-down contextual control. Previous studies reported that people avoid tasks with multiple alternatives or items to keep in mind or number of distractors (Feghhi et al., 2021; Schouppe et al., 2014; Westbrook et al., 2013). Often, policy abstraction is confounded with the number of items in these manipulations. The current work sought to distinguish difficulty from policy abstraction versus the number of options as influences on task avoidance. We found that participants are less sensitive to differences in the number of options between tasks than policy abstraction in their avoidance choices. Though, we note that they did report through their subjective ratings that as the number of options increased, they experienced greater effort. Nonetheless, this did not influence their choices about which task to perform to the same degree as with policy abstraction.

Thus, it may be that certain kinds of control demand, such as those imposed by managing deeper decision trees, are marked as more effortful, relative to others, and are avoided at higher rates. Future studies that consider different control demands, as well as different incentive structures entailed by tasks (Ritz et al., 2022), will be needed to determine the relationship between control, mental effort, and task avoidance. Nonetheless, here we find evidence of at least one unambiguously control-related factor that results in task avoidance, even when controlling for performance confounds, which is in line with an expected value of control and related theories.

Following the above point, our results also indicate that no single task feature drives effort avoidance behavior. Although policy abstraction was the greatest driver of effort decisions in these experiments, multiple factors influenced effort avoidance. Both Experiment 1 and the cross-study analyses found that ER had the second most important influence on decisions. Moreover, the weight of influence each factor has on effort decisions might depend on the task context or the incentive structure of the tasks (Ritz et al., 2022). For example, in the absence of differences in task performance between tasks, effort decisions might more heavily depend on features like policy abstraction, whereas incentivizing accurate behavior or providing explicit performance feedback might weight error likelihood more. Other task structures might place the weight on attentional or other demands for task optimization. Furthermore, the subjective feeling of effort, as distinct from effort avoidance, may likewise derive from multiple task features ranging from task performance to cognitive or physical task difficulty and may depend on situational factors (Janczyk et al., 2022). Future studies should verify the contribution of multiple factors on effort costs and directly test the effect of task context and structure on the determinants of effort avoidance.

Notably, policy abstraction offers an operational way to assess the cognitive effortfulness of a task without relying on the observation of behavior itself. For example, we observed that to the degree that ostensibly different kinds of tasks; WM versus task switching as manipulated by the delay versus restart conditions in Experiment 2; employed greater levels of policy abstraction, they yielded greater levels of effort avoidance. Thus, analyzing the policy abstraction required by a task offers a potential way of predicting the relative costs of different tasks independent of task performance and/or avoidance behavior.

A further implication of the fact that policy abstraction can be assessed independently of performance is that people might be able to assess this variable intuitively during their own planning and preparation for a task. If so, this might allow people to assign an expected effort without having to perform the task. In line with this possibility, we tested the choice between task conditions (F3 vs. R3) that participants had never performed, but that differed in their putative policy abstraction. We found that people nonetheless avoided the task (F3) that was associated with higher policy abstraction. This suggests that participants could predict the cost associated with performing that task, even without directly experiencing it. Whether they did so based on an intuition about the policy abstraction involved, an estimate, their own performance, or other factors should be a topic for future research.

The observation that participants can make their effort decisions without performing a task is also in line with previous studies showing that effort avoidance might be a product of awareness of the effort differences between conditions. For example, in the DST, task instructions do not explicitly state how two options differ from one another in their difficulty or in factors that impact difficulty, like the rate of task switching. Instead, participants learn to avoid the more difficult option from experience (Kool et al., 2010; Sayalı & Badre, 2019, 2021). However, Gold et al. (2015) found that schizophrenia patients failed to avoid options in the DST with higher rates of task switching unless they were explicitly instructed about the task switching difficulty manipulation. This suggests that proper effort prediction might partly be dependent on the awareness of the factors in the task structure that affect difficulty, in addition to learning the value of each option, through experience.

Relatedly, our results are also relevant to cue awareness, as an account of effort avoidance (T. Dunn, Gaspar, & Risko, 2019). This account proposes that effort avoidance may be driven by the salience of the available task cues and the inferences applied to those cues regarding their effortfulness. In our study, the observation that participants were more likely to select the previously unexperienced R3 task over F3 may imply that hierarchical contingencies of a task determine the salience of the task cue, perhaps via affecting its interpretability (see, e.g., Method 1 in the online supplemental materials for task instructions), an assumption our study was not designed to test. However, in the context of both experiments, the predominating effect of policy abstraction on selection rates cannot be entirely explained by cue awareness, at least in the broad sense. For instance, in Experiment 2, all task conditions yield the same number of task cues and these cues switch at the same rate across conditions. Nevertheless, policy abstraction differs among these conditions and selection rates differ accordingly. Likewise, across both experiments, salient cued features about the task other than policy abstraction, such as the number of options within an abstraction level in Experiment 1 and WM delay in Experiment 2, failed to explain effort selections better than policy abstraction. Nevertheless, our data do not speak to whether people are implicitly or explicitly aware of cues of policy abstraction or whether policy abstraction itself could be a metacognitive signal. Thus, cue awareness may still be important, but our results specify that awareness of cues to policy abstraction is important and separable from other factors.

Though we have shown that policy abstraction affects demand avoidance across two studies, there remain limitations in the present work that should be addressed in future research. First, the outcomes here could conceivably have been determined by our particular definition of policy abstraction and task/process analysis, though these are theoretically motivated and consistent with our definitions from previous work (Badre, 2008; Badre & Nee, 2018). We sought to mitigate this limitation by analyzing multiple alternate definitions of policy abstraction (see Results 4 and 5 in the online supplemental materials), and in so doing did not observe qualitative changes to our results. This indicates that the effect of policy abstraction was robust within a range of definitions. This concern is also mitigated by our observation that the effect of policy abstraction generalized across tasks that had several different specific features. Nonetheless, we recognize that there could be other task analyses than those we tested here which might yield different outcomes.

Second, to disentangle the effect of working memory from that of policy abstraction on task selections, we introduced delay as a distinct test of general WM capacity, separate from load. Our reasoning was that if it is the demand on WM maintenance, rather than policy, that was driving effort avoidance, then effort avoidance would be driven by WM demands unrelated to policy abstraction. We did not find evidence that this was the case, however. This observation is inconsistent with a straightforward WM demand account of the effort cost. However, we acknowledge that our study design cannot entirely rule out the possibility of more complex interactions between duration and load, such as that the cognitive demand of maintaining more items over shorter periods might surpass the additive effects of preserving fewer items for longer durations. Thus, to further elucidate these points, future studies might adopt a fully factorial design that crosses duration with load. Such a design could test any interactions and provide a more comprehensive understanding of the interplay between policy abstraction, WM load, and delay. Unpacking these dynamics could lead to a more comprehensive picture of how cognitive load and policy abstraction interact in complex tasks.

Third, despite large effect sizes and a power analysis based on a pilot study, the current study had a relatively small sample size. Future studies should aim at replicating our results in larger samples based on reported effect sizes and generalizing it using an even wider range of tasks.

Finally, we used demand avoidance as our primary objective measure of mental effort. However, we note that this does not directly measure mental effort costs, and so we cannot rule out that other subjective measures or objective cost measures, such as cognitive effort discounting (Westbrook et al., 2013) could observe different results. Additionally, previous research has shown that subjective perception of effort and effort avoidance rates as assessed by DSTs might not be collinear (Sayalı & Badre, 2019) and effort costs may be captured by peripheral indices of cognitive effort such as cardiovascular reactivity (Silvestrini, 2017) or pupil dilation (Bijleveld et al., 2009; Silvetti et al., 2018). The relationship between these indices of effort and demand avoidance rates should be tested in future studies. Thus, our results most directly support an effect of policy abstraction on demand avoidance behavior, whereas its relationship to the subjective experience of mental effort costs is only indirectly supported.

Context

In summary, across two studies and one within-subject, cross-study analysis, we showed that policy abstraction is an important and task-general factor in predicting demand avoidance, over and above task performance. This suggests that people find tasks involving multiple levels of contextual contingency to be more difficulty, aversive, and mentally effortful. Moreover, this factor can account for demand avoidance rates across different cognitive task types, and its cost can be inferred in the absence of direct experience with the task under consideration. To the degree that performing a task with higher order policy compromises multitasking capacity, our findings are also consistent with theories of mental effort that relate the costs of control and mental effort to opportunity costs (Musslick & Cohen, 2021). Thus, our results highlight the importance of task structure and metacognitive strategies, along with expected performance, in a multifactor account of cognitive effort.

References

Alexander, W. H., & Brown, J. W. (2015). Hierarchical error representation: A computational model of anterior cingulate and dorsolateral prefrontal cortex. *Neural Computation*, *27*(11), 2354–2410. https://doi.org/10.1162/NECO_a_00779

Anderson, D. R., & Burnham, K. P. (2002). Avoiding pitfalls when using information-theoretic methods. *The Journal of Wildlife Management*, *66*(3), 912–918. https://doi.org/10.2307/3803155

Apps, M. A. J., Grima, L. L., Manohar, S., & Husain, M. (2015). The role of cognitive effort in subjective reward devaluation and risky decision-making. *Scientific Reports*, *5*(1), Article 16880. https://doi.org/10.1038/srep16880

Badre, D. (2008). Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends in Cognitive Sciences*, *12*(5), 193–200. https://doi.org/10.1016/j.tics.2008.02.004

Badre, D., Bhandari, A., Keglovits, H., & Kikumoto, A. (2021). The dimensionality of neural representations for control. *Current Opinion in Behavioral Sciences*, *38*, 20–28. https://doi.org/10.1016/j.cobeha.2020.07.002

Badre, D., & D'Esposito, M. (2007). Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *Journal of Cognitive Neuroscience*, *19*(12), 2082–2099. https://doi.org/10.1162/jocn.2007.19.12.2082

Badre, D., & D'Esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical? *Nature Reviews Neuroscience*, *10*(9), 659–669. https://doi.org/10.1038/nrn2667

Badre, D., & Nee, D. E. (2018). Frontal cortex and the hierarchical control of behavior. *Trends in Cognitive Sciences*, *22*(2), 170–188. https://doi.org/10.1016/j.tics.2017.11.005

Bijleveld, E., Custers, R., & Aarts, H. (2009). The unconscious eye opener: Pupil dilation reveals strategic recruitment of resources upon presentation of subliminal reward cues. *Psychological Science*, *20*(11), 1313–1315. https://doi.org/10.1111/j.1467-9280.2009.02443.x

Botvinick, M. M. (2007). Multilevel structure in behaviour and in the brain: A model of Fuster's hierarchy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1485), 1615–1626. https://doi.org/10.1098/rstb.2007.2056

Boureau, Y. L., Sokol-Hessner, P., & Daw, N. D. (2015). Deciding how to decide: Self-control and meta-decision making. *Trends in Cognitive Sciences*, *19*(11), 700–710. https://doi.org/10.1016/j.tics.2015.08.013

Brown, J. W., & Braver, T. S. (2005). Learned predictions of error likelihood in the anterior cingulate cortex. *Science*, *307*(5712), 1118–1121. https://doi.org/10.1126/science.1105783

Buckland, S. T., Burnham, K. P., & Augustin, N. H. (1997). Model selection: An integral part of inference. *Biometrics*, *53*(2), 603–618. https://doi.org/10.2307/2533961

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*(1), 116–131. https://doi.org/10.1037/0022-3514.42.1.116

Calcagno, V., & de Mazancourt, C. (2010). glmulti: An R package for easy automated model selection with (generalized) linear models. *Journal of Statistical Software*, *34*(12), 1–29. https://doi.org/10.18637/jss.v034.i12

Cavanagh, J. F., Masters, S. E., Bath, K., & Frank, M. J. (2014). Conflict acts as an implicit cost in reinforcement learning. *Nature Communications*, *5*(1), Article 5394. https://doi.org/10.1038/ncomms6394

Choi, E. Y., Drayna, G. K., & Badre, D. (2018). Evidence for a functional hierarchy of association networks. *Journal of Cognitive Neuroscience*, *30*(5), 722–736. https://doi.org/10.1162/jocn_a_01229

Collins, A. G., & Frank, M. J. (2013). Cognitive control over learning: Creating, clustering, and generalizing task-set structure. *Psychological Review*, *120*(1), 190–229. https://doi.org/10.1037/a0030852

Collins, A. G. E., & Frank, M. J. (2016). Motor demands constrain cognitive rule structures. *PLoS Computational Biology*, *12*(3), Article e1004785. https://doi.org/10.1371/journal.pcbi.1004785

Dayan, P. (2007). Bilinearity, rules, and prefrontal cortex. *Frontiers in Computational Neuroscience*, *1*, Article 73. https://doi.org/10.3389/neuro.10.001.2007

De Santis, F. (2004). Statistical evidence and sample size determination for Bayesian hypothesis testing. *Journal of Statistical Planning and Inference*, *124*(1), 121–144. https://doi.org/10.1016/S0378-3758(03)00198-8

Dunn, T., Gaspar, C., & Risko, E. F. (2019). Cue awareness in avoiding effortful control. *Neuropsychologia*, *123*(1), 77–91. https://doi.org/10.1016/j.neuropsychologia.2018.05.011

Dunn, T. L., Inzlicht, M., & Risko, E. F. (2019). Anticipating cognitive effort: Roles of perceived error-likelihood and time demands. *Psychological Research*, *83*(5), 1033–1056. https://doi.org/10.1007/s00426-017-0943-x

Feghhi, I., Franchak, J. M., & Rosenbaum, D. A. (2021). Towards a common code for difficulty: Navigating a narrow gap is like memorizing an extra digit. *Attention, Perception, & Psychophysics*, *83*(8), 3275–3284. https://doi.org/10.3758/s13414-021-02356-4

Feghhi, I., & Rosenbaum, D. A. (2021). Effort avoidance is not simply error avoidance. *Psychological Research*, *85*(4), 1462–1472. https://doi.org/10.1007/s00426-020-01331-2

Frank, M. J., & Badre, D. (2012). Mechanisms of hierarchical reinforcement learning in corticostriatal circuits 1: Computational analysis. *Cerebral Cortex*, *22*(3), 509–526. https://doi.org/10.1093/cercor/bhr114

Gold, J. M., Kool, W., Botvinick, M. M., Hubzin, L., August, S., & Waltz, J. A. (2015). Cognitive effort avoidance and detection in people with schizophrenia. *Cognitive, Affective, & Behavioral Neuroscience*, *15*(1), 145–154. https://doi.org/10.3758/s13415-014-0308-5

Janczyk, M., Feghhi, I., & Rosenbaum, D. A. (2022). What matters in making demand-based decisions: Time alone or difficulty too? *Psychological Research*, *86*, 1355–1365. https://doi.org/10.1007/s00426-021-01583-6

Johnson, J. B., & Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in Ecology & Evolution*, *19*(2), 101–108. https://doi.org/10.1016/j.tree.2003.10.013

Kleinsorge, T., & Heuer, H. (1999). Hierarchical switching in a multi-dimensional task space. *Psychological Research*, *62*(4), 300–312. https://doi.org/10.1007/s004260050060

Koechlin, E., & Hyafil, A. (2007). Anterior prefrontal function and the limits of human decision-making. *Science*, *318*(5850), 594–598. https://doi.org/10.1126/science.1142995

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General*, *139*(4), 665–682. https://doi.org/10.1037/a0020198

Kurzban, R., Duckworth, A., Kable, J. W., & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behavioral and Brain Sciences*, *36*(6), 661–679. https://doi.org/10.1017/S0140525X12003196

Logan, G. D. (1990). Repetition priming and automaticity: Common underlying mechanisms? *Cognitive Psychology*, *22*(1), 1–35. https://doi.org/10.1016/0010-0285(90)90002-L

McGuire, J. T., & Botvinick, M. M. (2010). Prefrontal cortex, cognitive control, and the registration of decision costs. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(17), 7922–7926. https://doi.org/10.1073/pnas.0910662107

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*(1), 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167

Morey, R., & Rouder, J. (2022). *BayesFactor: Computation of Bayes factors for common designs* (R package version 0.9.12-4.4). https://cran.r-project.org/web/packages/BayesFactor/BayesFactor.pdf

Musslick, S., & Cohen, J. D. (2021). Rationalizing constraints on the capacity for cognitive control. *Trends in Cognitive Sciences*, *25*(9), 757–775. https://doi.org/10.1016/j.tics.2021.06.001

Myerson, J., & Green, L. (1995). Discounting of delayed rewards: Models of individual choice. *Journal of the Experimental Analysis of Behavior*, *64*(3), 263–276. https://doi.org/10.1901/jeab.1995.64-263

Nee, D. E. (2021). Integrative frontal-parietal dynamics supporting cognitive control. *eLife*, *10*, Article e57244. https://doi.org/10.7554/eLife.57244

Nee, D. E., & D'Esposito, M. (2016). The hierarchical organization of the lateral prefrontal cortex. *eLife*, *5*, Article e12112. https://doi.org/10.7554/eLife.12112

Nee, D. E., & D'Esposito, M. (2017). Causal evidence for lateral prefrontal cortex dynamics supporting cognitive control. *eLife*, *6*, Article e28040. https://doi.org/10.7554/eLife.28040

Newell, A. (1994). *Unified theories of cognition*. Harvard University Press.

Niv, Y., Daw, N. D., Joel, D., & Dayan, P. (2007). Tonic dopamine: Opportunity costs and the control of response vigor. *Psychopharmacology*, *191*(3), 507–520. https://doi.org/10.1007/s00213-006-0502-4

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D., Conway, A., Cowan, N., Donkin, C., Farrell, S., Hitch, G. J., Hurlstone, M. J., Ma, W. J., Morey, C. C., Nee, D. E., Schweppe, J., Vergauwe, E., & Ward, G. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin*, *144*(9), 885–958. https://doi.org/10.1037/bul0000153

Otto, A. R., & Daw, N. D. (2019). The opportunity cost of time modulates cognitive effort. *Neuropsychologia*, *123*, 92–105. https://doi.org/10.1016/j.neuropsychologia.2018.05.006

Rac-Lubashevsky, R., & Frank, M. J. (2021). Analogous computations in working memory input, output and motor gating: Electrophysiological and computational modeling evidence. *PLoS Computational Biology*, *17*(6), Article e1008971. https://doi.org/10.1371/journal.pcbi.1008971

Ranti, C., Chatham, C. H., & Badre, D. (2015). Parallel temporal dynamics in hierarchical cognitive control. *Cognition*, *142*, 205–229. https://doi.org/10.1016/j.cognition.2015.05.003

Ritz, H., Leng, X., & Shenhav, A. (2022). Cognitive control as a multivariate optimization problem. *Journal of Cognitive Neuroscience*, *34*(4), 569–591. https://doi.org/10.1162/jocn_a_01822

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

Sayalı, C., & Badre, D. (2021). Neural systems underlying the learning of cognitive effort costs. *Cognitive, Affective, & Behavioral Neuroscience*, *21*(4), 698–716. https://doi.org/10.3758/s13415-021-00893-x

Schouppe, N., Demanet, J., Boehler, C. N., Ridderinkhof, K. R., & Notebaert, W. (2014). The role of the striatum in effort-based decision-making in the absence of reward. *Journal of Neuroscience*, *34*(6), 2148–2154. https://doi.org/10.1523/JNEUROSCI.1214-13.2014

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron*, *79*(2), 217–240. https://doi.org/10.1016/j.neuron.2013.07.007

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience*, *40*(1), 99–124. https://doi.org/10.1146/annurev-neuro-072116-031526

Silvestrini, N. (2017). Psychological and neural mechanisms associated with effort-related cardiovascular reactivity and cognitive control: An integrative approach. *International Journal of Psychophysiology*, *119*, 11–18. https://doi.org/10.1016/j.ijpsycho.2016.12.009

Silvetti, M., Vassena, E., Abrahamse, E., & Verguts, T. (2018). Dorsal anterior cingulate-brainstem ensemble as a reinforcement meta-learner. *PLoS Computational Biology*, *14*(8), Article e1006370. https://doi.org/10.1371/journal.pcbi.1006370

Weiss, R. (1997). Bayesian sample size calculations for hypothesis testing. *Journal of the Royal Statistical Society. Series D: The Statistician*, *46*(2), 185–191. https://doi.org/10.1111/1467-9884.00075

Westbrook, A., Kester, D., & Braver, T. S. (2013). What is the subjective cost of cognitive effort? Load, trait, and aging effects revealed by economic preference. *PLoS ONE*, *8*(7), Article e68210. https://doi.org/10.1371/journal.pone.0068210

Received June 14, 2022
Revision received May 9, 2023
Accepted May 18, 2023 ■