# Quantifying the cost of context sensitivity in decision making

**Shuze Liu (shuzeliu@fas.harvard.edu)**
PhD Program in Neuroscience, Harvard University, Cambridge, MA 02138 USA

**Samuel J. Gershman (gershman@fas.harvard.edu)**
Department of Psychology and Center for Brain Science, Harvard University, Cambridge, MA 02138 USA

**Bilal A. Bari (bbari@mgh.harvard.edu)**
Massachusetts General Hospital, Boston, MA 02114 USA

## Abstract

It is well known that context-dependent decisions incur mental costs. While previous research has sought to formalize these costs at various levels of analysis, we still lack basic insight into the nature of mental costs, including the underlying cognitive resources being consumed. Moreover, many computational models assume that mental costs scale linearly with the cognitive resource being used, an assumption of convenience that has yet to be systematically tested. To address these gaps, we build on rate-distortion theory by formalizing an information-theoretic notion of mental costs. Specifically, we define the cost of policies—the mappings from states to actions—as a function of the mutual information between states and actions, the *policy complexity*. Across four decision-making experiments featuring diverse task manipulations, we find that this mental cost formulation offers a parsimonious description of how humans adaptively adjust their policy complexity across different tasks. Notably, a quadratic mental cost formulation, where increases in policy complexity incur supralinear costs, provides the best fit. These findings highlight the meta-cognitive ability of humans to account for mental costs when forming decision strategies, and pave the way towards a domain-general quantification of mental effort.

**Keywords:** rate-distortion theory; policy compression; action selection; motivational effects; mental effort; mental costs

## Introduction

Mental effort significantly impacts decision making: people avoid tasks with high cognitive demands (Kool et al., 2010) and exert more effort for more reward (Krebs & Woldorff, 2017). Despite these empirical findings, it remains unclear why mental effort is costly. One recent perspective is resource rationality—the idea that the brain must adapt to environmental demands and internal resource limitations (Lieder & Griffiths, 2020). These resource limitations are thought to give rise to the mental effort of cognition (Kool & Botvinick, 2018). Even so, the problem of mental effort remains ill-constrained, and it is unclear whether there are generalizable formulations of mental costs (Shenhav et al., 2017).

We shed light on the nature of mental costs in contextual decisions through rate-distortion theory and rational inattention (Maćkowiak et al., 2023). This builds upon the idea that the mind is an information processing system that can be understood via its inputs and outputs (Marr, 2010; Simon, 1978). We use the policy compression framework, which defines the mental cost of a policy in terms of its *policy complexity*. If the policy's cost exceeds what the agent is willing to pay, it must compress its policy by making it less state-dependent, which reduces the mental cost.

Here, we quantify the mental cost of policy complexity and test the assumption that mental costs scale linearly with the resource being used (Bhui et al., 2021). We conducted four human experiments manipulating intertrial intervals (ITIs), stimulus set sizes, and reward magnitudes. We fit models with different cost formulations to predict empirical policy complexity, and confirmed human sensitivity to relevant mental costs. We also found that human cost functions are *supralinear* in policy complexity, which has implications for the neural instantiation of such costs. These findings represent a promising step toward a domain-general quantification of mental effort in human decision making.

## Methods

### Theory: Policy compression

The brain has evolved to function under myriad cognitive resource constraints; here we focus on the transmission of information. We model an agent executing a policy, $\pi(a|s)$, a probabilistic mapping from states $s$ to actions $a$. For a resource-rational agent, we formalize the cognitive resource as the mutual information between states and actions, $I^{\pi}(S;A)$, or *policy complexity* (Gershman, 2020; Parush et al., 2011; Sims, 2016). We focus on $I^{\pi}(S;A)$ due to Shannon's noisy channel theorem, which states that the minimum expected number of bits to transmit a signal across a noisy channel without error is equal to $I^{\pi}(S;A)$ (Figure 1A). High-complexity policies require more bits to execute and thus recruit more resources. To investigate how humans determine an optimal allocation of resources (their $I^{\pi}(S;A)$), it would be informative to derive the maximum attainable reward at any given $I^{\pi}(S;A)$. To do so, we must find the optimal policy, $\pi^* = \text{argmax}_{\pi} V^{\pi}$ subject to $I^{\pi}(S;A) \leq C$ for every $C$, where $V^{\pi}$ is the expected reward under $\pi$. This problem has the following Lagrangian form:

$$\pi^*(a|s) = \underset{\pi}{\text{argmax}}\, \beta V^{\pi} - I^{\pi}(S;A) + \sum_{s} \lambda(s) \left( \sum_{a} \pi(a|s) - 1 \right) \tag{1}$$

where $\beta, \lambda(s)$ are Lagrangian multipliers to enforce $I(S;A) \leq C$ and policy normalization.[1] Solving the Lagrangian leads to

[1] While Equation 1 contains a linear $I^{\pi}(S;A)$ term, we only introduced it to numerically derive the optimal policy at each $I^{\pi}(S;A)$ level, which enabled us to trace out task-specific reward-complexity frontiers (e.g., Figure 1D). We do not commit to its implied assumption that human mental costs are linear in $I^{\pi}(S;A)$, and will systematically compare linear and nonlinear cost formulations later on.

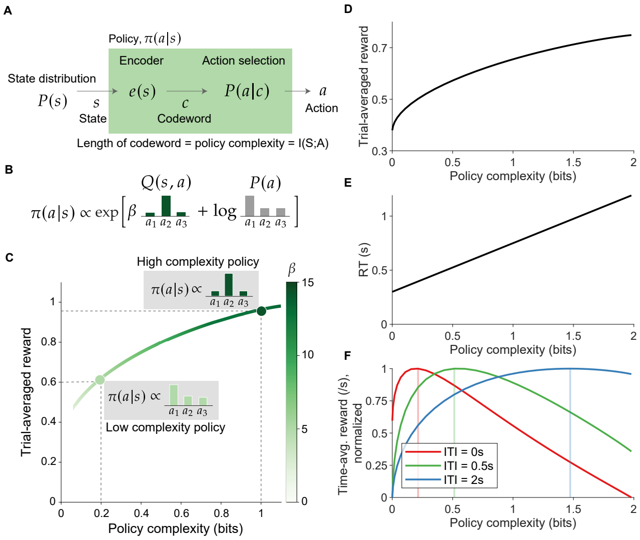

Figure 1: **The policy compression framework.** **(A)** The policy as a communication channel. A state distribution $P(s)$ generates states $s$ that are encoded into memory, yielding a codeword $c$. The codeword is then mapped onto an action $a$ according to $P(a|c)$. Together, encoding and action selection produce the policy $\pi(a|s)$. **(B)** The optimal policy includes a state-dependent term, $Q(s,a)$, and a state-independent term, $\log P(a)$. The $\log P(a)$ term biases choices towards actions frequently chosen across all states. The $\beta$ parameter determines the relative contribution of $Q(s,a)$ and $\log P(a)$, controlling the policy's state-dependence. **(C)** The $\beta$ parameter increases monotonically with policy complexity. We highlight two optimal policies at different policy complexity levels. The optimal policies trace out the reward-complexity frontier, which delimits achievable performance for a given policy complexity. **(D)** Reward-complexity frontier for Experiment 1. **(E)** Proposed linear relationship between RT and policy complexity. **(F)** Time-averaged reward as a function of policy complexity for each ITI under the linear RT relationship in (E); the policy complexity that maximizes time-averaged reward for each condition is highlighted (vertical lines). Panels A-C adapted from Lai and Gershman (2024); Panels D-F adapted from Liu et al. (2024).

the solution:

$$\pi^*(a|s) \propto \exp(\beta Q(s,a) + \log P^*(a)) \tag{2}$$

where $Q(s,a)$ is the expected reward for taking action $a$ in state $s$ and $P^*(a) = \sum_s \pi^*(a|s)p(s)$ is the optimal marginal action distribution. The optimal policy is similar to the softmax decision rule (Sutton & Barto, 2018), but additionally features influence from $P^*(a)$, which biases the optimal policy towards frequently taken actions (Figure 1B). The Lagrange multiplier, $\beta$, is analogous to the inverse temperature parameter and changes based on the desired policy complexity level: $\beta^{-1} = \frac{dV^\pi}{dI^\pi(S;A)}$. It is large at high policy complexity and small at low policy complexity. By varying $\beta$ and calculating the optimal policy, we can trace out the reward-complexity frontier, which delimits the maximal trial-averaged reward obtainable at any given policy complexity level (Figure 1C,D). In general, high-complexity policies yield more reward per trial than low-complexity policies. Moreover, low-complexity policies are dominated by the $\log P^*(a)$ term, a form of state-independent perseveration (Lai & Gershman, 2021).

The reward-complexity frontier prescribes only that agents should fall somewhere along it, without specifying an exact $I^\pi(S;A)$ level they should select. If we assume the agent has a fixed resource budget, its $I^\pi(S;A)$ should never change across tasks. If the agent aims to maximize trial-averaged reward and has sufficient resources, its $I^\pi(S;A)$ should always be the maximum allowable (e.g., 2 bits in Figure 1D). The empirically inaccurate predictions above arise because they ignore the time costs of $I^\pi(S;A)$. To understand why an agent would adaptively adjust its $I^\pi(S;A)$ across non-maximal levels, let us assume states are represented as codewords through entropy coding, a canonical example being the Huffman code (Huffman, 1952). The Huffman code corresponds to a binary tree in which leaf nodes correspond to decoded states, where more complex state descriptions necessitate more leaf nodes and therefore more bits. If we assume bits are inspected at a constant rate, then high-complexity policies require longer readout times to reveal the decoded action, necessitating longer response times (RTs) (Hick, 1952). Moreover, given that bits are inspected at a constant rate, the trial-averaged RT should be a linear function of $I^\pi(S;A)$ / average description length, with some offset to reflect motor delay (Figure 1E).

To model human decision making as sensitive to these time costs, we assume subjects attempt to maximize *time-averaged* reward (Balci et al., 2011; Drugowitsch et al., 2015):

$$V^\pi_{\text{time}}(I(S;A)) = \frac{V^\pi(I(S;A))}{\text{RT}(I(S;A)) + \text{ITI}} \tag{3}$$

where $V^\pi_{\text{time}}(I(S;A))$ is the time-averaged reward. $V^\pi(I(S;A))$ is a function of policy complexity through the derivation of the optimal policy[2] (i.e., the reward-complexity frontier in Figure 1D) and $\text{RT}(I(S;A))$ is a function of policy complexity through the assumption of a linear relationship between RT and policy complexity (Figure 1E). This yields the relationship in Figure 1F, where we varied the ITI. To maximize time-averaged reward, humans should decrease policy complexity when ITIs are short; although these policies result in less trial-averaged reward, they increase time-averaged reward because they allow agents to perform more actions due to smaller decoding time cost. The theory also predicts that RTs should grow as a function of the number of possible states, a type of set-size effect. This is because larger sets require higher policy complexity to maximize time-averaged reward, which in turn demands longer decoding time. Finally, the theory predicts that greater reward magnitudes should increase policy complexity by increasing the numerator, $V^\pi(I(S;A))$.

[2] Due to our study's focus on how humans determine their $I(S;A)$, we assume that they always use the optimal policy at their chosen $I(S;A)$ level. How humans find this policy is a topic for future work.

## General task description

Experiments 1-2 were from Liu et al. (2024). Here we augment the new Experiments 3-4, which generalize the theory to extreme cognitive loads and a different domain (reward magnitude changes). Participants did not overlap across experiments and gave informed consent. All statistical tests on behavioral variables were preregistered; the mental cost modeling is post-hoc. Data and code are available at https://github.com/LSZ2001/policycompression_mentalcostmodeling.

## Experiment 1

**Procedure.** Each participant completed three blocks of trials with ITIs of 0s, 0.5s, and 2s. The block order was randomized across participants. Participants were informed of the ITI of each block. Participants were informed that they would receive a bonus proportional to their performance for each block. There were four possible states (images) and four available actions. Each stimulus was assigned a unique optimal action (Figure 2B). Participants were informed that the stimulus-action mapping was held fixed across all blocks.

On each trial, participants were presented with one image (state) and responded by pressing one of several possible keyboard keys (actions; Figure 2A). Stimulus presentation was counterbalanced within runs of 8 trials, where the stimulus presentation order was randomized within each run and each of the four images appeared exactly twice per run. We did this to ensure a uniform state distribution $P(s)$, allowing us better estimate policy complexity. Reward delivery was binary and probabilistic: each state was associated with one optimal action (Figure 2B). After making a response, participants were given immediate feedback for 0.3s—either a green border around the image to indicate reward or a gray border to indicate no reward. A fixation cross then appeared throughout the ITI. Each block lasted until 3 minutes elapsed and the current run of trials finished. Participants could track the remaining time and reward earned during the block, displayed as red and green bars respectively. At the end of each block, they were provided with feedback on the total reward they earned in that block (Figure 2C).

Participants completed three 1-minute training blocks, one for each ITI condition, to familiarize themselves with the task and learn the mapping from stimulus to response. These data were not analyzed. Participants then completed the three 3-minute blocks where ITI was varied, as mentioned above.

**Participants.** 100 participants (37 women, 61 men, 1 non-binary, 1 prefer not to say) were recruited. We selected the sample size based on the lowest estimated effect size (Cohen's $d = 0.312$) of interest, according to estimates from a separate group of $N = 48$ pilot participants (data excluded from final analysis). The behavioral analyses were preregistered at https://aspredicted.org/blind.php?x=VF2_NH6. We excluded 3 participants for having an average RT for any block exceeding 5 seconds, leaving data from 97 participants (35 women, 60 men, 1 non-binary, 1 prefer not to say) for subsequent analyses.

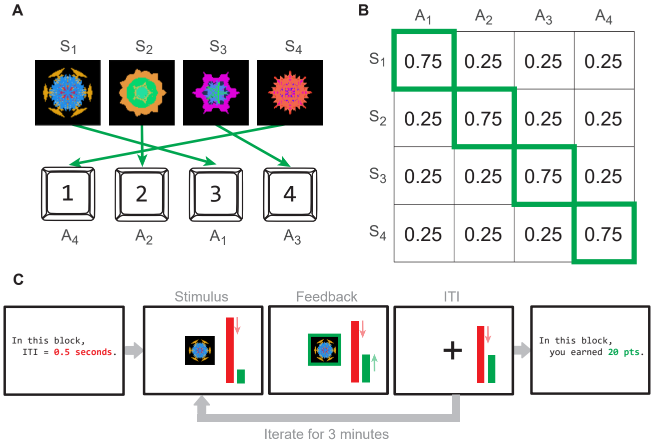
A
$S_1$
$S_2$
$S_3$
$S_4$
1
2
3
4
$A_4$
$A_2$
$A_1$
$A_3$
B
$A_1$ $A_2$ $A_3$ $A_4$
$S_1$ 0.75 0.25 0.25 0.25
$S_2$ 0.25 0.75 0.25 0.25
$S_3$ 0.25 0.25 0.75 0.25
$S_4$ 0.25 0.25 0.25 0.75
C
In this block, ITI = 0.5 seconds.
Stimulus
Feedback
ITI
+
In this block, you earned 20 pts.
Iterate for 3 minutes

Figure 2: **Experiment design.** Here we illustrate Experiment 1, as other experiments feature similar designs. **(A)** The four possible states (images) and the corresponding optimal actions (key presses). The mapping between images and keys was randomized across participants. **(B)** Reward probability for each state-action pair. The optimal action for each state is indicated by a green border. **(C)** On every trial, the participant observes an image (state) and responds by pressing a key (action). They can track the cumulative reward (green bar) and time remaining (red bar) in the block. Reward feedback is provided as an image border, whose color indicates the reward magnitude. Participants then see a fixation cross during the ITI before proceeding to the next trial. Participants are informed of the block's ITI before starting.

## Experiment 2

**Procedure.** The three test blocks had stimuli set sizes of 2, 4, and 6 stimuli respectively, and their order was randomized across participants. Each set size used unique images to make each set-size manipulation as independent as possible. The action set size was fixed at 6 across all set-size conditions. We used ITI = 2s for each block. Participants were informed of the ITI and set size of each block.

Each stimulus was associated with a unique optimal action. Like Experiment 1, optimal actions yielded reward with probability 0.75 and suboptimal actions yielded reward with probability 0.25. Stimuli were randomized and presented in counterbalanced runs of 8, 8, and 10 trials for set sizes 2, 4, and 6 respectively (each stimulus therefore appeared 4 times, 2 times, and 2 times respectively within each run).

For each set-size condition, participants first completed three training blocks with ITI = 0s, 0.5s, and 2s, similar to Experiments 1 and 2. We did this to encourage learning and minimize the length of training. To ensure similar learning across set-size conditions, we presented each stimulus 48 times during training (24 for ITI = 0s, 16 for ITI = 0.5s, and 8 for ITI = 2s) rather than training for a fixed time duration. Participants were told that the mapping from stimuli to actions remained fixed between training and test blocks. After completing the three training blocks, participants proceeded to the 3-minute test block of the same set-size. The structure, visual display, and duration of blocks were identical to Experiment 1.

**Participants.** 101 participants (54 women, 44 men, 2 non-binary, 1 prefer not to say) were recruited. We selected the sample size based on the lowest estimated effect size (Cohen's $d = 0.459$) of interest, according to analyses of a separate group of $N = 48$ pilot participants (data excluded from final analysis). The preregistration is at https://aspredicted.org/ZSW_HFY. The inclusion criterion was identical to Experiment 1. 99 participants (53 women, 43 men, 2 non-binary, 1 prefer not to say) were included for analyses.

## Experiment 3

**Procedure.** The procedures are mostly identical to Experiment 2, with the following changes: 1) The three 3-minute test blocks had set sizes of 2, 6, and 7 to impose extreme cognitive loads. 2) The number of available actions was equal to the set size of that block. 3) Unlike the training in Experiment 2, there are three 1-minute training blocks before each test block sharing the same set size, with ITI 0s, 0.5s, and 2s.

**Participants.** 157 participants (70 women, 85 men, 1 non-binary, 1 prefer not to say) were recruited. We selected the sample size based on typical sample sizes for policy compression experiments. The preregistration is at https://aspredicted.org/bpp2-djvs.pdf. The inclusion criterion was identical to Experiment 1. 153 participants (68 women, 83 men, 1 non-binary, 1 prefer not to say) were included.

## Experiment 4

**Procedure.** Each participant completed two test blocks of trials with different reward magnitudes, specified in U.S. cents (¢) provided as bonuses. The block order was randomized across participants. In each block, there were four possible states (images) and four available actions, where each state was assigned a unique optimal action (Figure 2A). Each block used unique images, requiring participants to relearn the state-action mappings for every block.

Each test block contained 100 trials. On each trial, participants saw one image (state) and responded by pressing one of four keyboard keys (actions). The order of state presentation was randomized and each state was presented for 25 trials to ensure a uniform state distribution. If the action taken was suboptimal for the presented state, the participant received 0.024¢. If the action was optimal, the participant received 0.024¢ with 0.2 probability or a larger reward with 0.8 probability (Figure 2B). For the high reward magnitude block, the larger reward was 2.5¢. For the low reward magnitude block, the larger reward was 0.025¢. On every trial, participants needed to respond in 1 second; otherwise, they lost -0.5¢ and were immediately redirected to the next trial. The final bonus was the summed reward over test blocks.

After making a response or failing to respond in 1 second, participants were given feedback for 0.5s—either a dark green image border to indicate 2.5¢ reward, a light green border to indicate 0.025¢ reward, a gray border to indicate 0.024¢ reward, or a red border to indicate -0.5¢ loss. A fixation cross appeared during the ITI, with an adaptive duration such that the time spent per trial was exactly 1.5 seconds.

Before each test block, participants additionally completed one training block with 60 trials, where the images used and the state-action mapping are identical to the upcoming test block to facilitate learning. These data were not analyzed.

**Participants.** 121 participants (61 women, 60 men) were recruited. We selected the sample size based on the lowest estimated effect size (Cohen's d = 0.301) of interest, according to estimates from a separate group of N = 28 pilot participants (data excluded from final analysis). The preregistration is at https://aspredicted.org/mtwr-4rm2.pdf. We excluded 2 participants for not responding for more than or equal to 20 test block trials, leaving data from 119 subjects (60 women, 59 men) for subsequent analyses.

## Statistical analyses and modeling

We estimated policy complexity for each participant in each condition using the Hutter estimator (Hutter, 2001). All tests were one-sided paired $t$-tests in the preregistered directions: ITI = 0s versus 2s for Experiment 1, set size 2 versus the largest set size for Experiment 2-3, and the low versus high reward conditions for Experiment 4. It is well known that exerting cognitive control during action selection taxes mental effort (Kool et al., 2010; Krebs & Woldorff, 2017; Shenhav et al., 2017; Umemoto & Holroyd, 2015). We therefore postulated that policy complexity should incur a mental cost, where more complex policies are more cognitively demanding and, therefore, more subjectively effortful. We model subjects as optimizing policy complexity (i.e., picking a policy of the desired policy complexity) as follows:

$$I(S;A)^* = \underset{I(S;A)}{\operatorname{argmax}} \frac{V^\pi(I(S;A)) - \text{MentalCost}(I(S;A))}{\text{RT}(I(S;A)) + \text{ITI}} \quad (4)$$

where $I^*$ is the optimal policy complexity level, $V^\pi(I(S;A))$ is the task-specific reward-complexity frontier, and $\text{MentalCost}(I(S;A))$ and $\text{RT}(I(S;A))$ are the agent-specific relationships between mental cost / average RT and policy complexity (which we assume are monotonically increasing). To characterize $\text{RT}(I(S;A))$, we fit a linear mixed-effects (LME) model separately for each experiment to predict the average RTs of participants in each task condition. The LME model contained fixed and participant-specific random effects for intercepts and policy complexity.

For Experiment 4 only, the task structure ensured reward rate was independent of response time (i.e., trials lasted a fixed duration, regardless of response times). Therefore, optimizing time-averaged reward vs trial-averaged reward were identical. We hence assumed agents were optimizing trial-averaged reward when selecting their policy complexity:

$$I(S;A)^* = \underset{I(S;A)}{\operatorname{argmax}} V^\pi(I(S;A)) - \text{MentalCost}(I(S;A)) \quad (5)$$

To characterize MentalCost($I(S;A)$). we fit three models:

$$\text{MentalCost}(I(S;A)) = \begin{cases} 0, & \text{NoCost} \\ \theta_1 I(S;A), & \text{LinearCost} \\ \theta_1 I(S;A) + \theta_2 I(S;A)^2, & \text{QuadCost} \end{cases} \tag{6}$$

which formalize no mental cost, linear cost in $I(S;A)$, and quadratic cost. Notably, the LinearCost model is equivalent to rational inattention models in economics (Gershman & Bhui, 2020), which similarly apply rate-distortion theory but treat $\beta$ as a mental cost sensitivity parameter, instead of a Lagrangian multiplier for constrained optimization. We fit models via the No U-Turn Sampler (NUTS) in Numpyro, report the leave-one-out expected log predictive densities (ELPD-LOO) for model comparison, and visualize posterior predictive means.

For Experiment 4, the low reward condition created difficulties in model fitting, as the range of possible total reward spanned less than 1¢, meaning rational agents should exert 0 policy complexity in this condition.[3] This is not what we observed (Figure 3, Row 2) and we addressed this discrepancy by assuming a subjective utility for responding accurately. We made an ad-hoc assumption and transformed the subjective reward for correct responses to 75% that of the high reward condition, which produced qualitatively good fits.

# Results

## Behavioral results

For each of the four experiments, participants lay close to the optimal reward-complexity frontier (Figure 3 Row 1). Consistent with theoretical predictions, subjects adopted policies with higher complexity with longer ITIs, with larger set sizes, and when the reward magnitude was high (Row 2). Note that these qualitative predictions came from optimizing Equation 4 (Equation 5 for Experiment 4). We next sought quantitative predictions, and optimized Equations 4 and 5 assuming no mental cost. The difference between this optimal policy complexity, assuming no mental cost, and empirically-estimated policy complexity is shown in Row 3. Here, we see a systematic leftward bias, indicating that subjects are consistently adopting policies with lower complexity than what is optimal for each task condition. This suggests that policy complexity incurs a mental cost, dissuading subjects from adopting the optimal policy complexity for each task. LME modeling of the relationship between policy complexity and RT achieved visibly good fits and revealed positive slopes for most participants, consistent with the idea that average RTs should linearly increase with policy complexity. Model comparison supported the LME's linearity assumption (Liu et al., 2024).

## Quantitative modeling of mental effort

Following the success of the RT($I(S;A)$) LME fits, we proceeded to characterize the relationship between mental cost and policy complexity. We did this by fitting each of the mental cost models in Equation 6 to participant behavior.

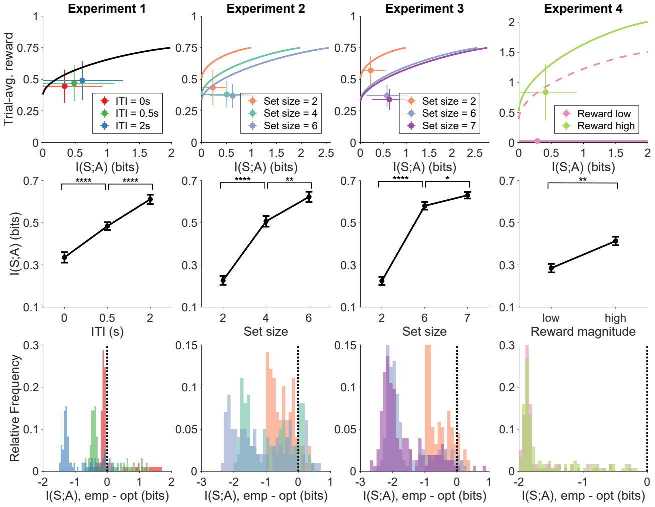

Figure 3: **Humans systematically exhibit less than optimal policy complexity.** Columns 1-4 correspond to Experiments 1-4. Colors denote task conditions. **Row 1:** Reward-complexity frontiers for each experiment, overlaid with human $I(S;A)$ and trial-averaged reward. **Row 2:** Mean±SEM of human $I(S;A)$ across conditions. **Row 3:** Difference between optimal and empirical $I(S;A)$ across conditions. Experiment 1-2 results are adapted from Liu et al. (2024). For Experiment 4, the dotted pink line denotes the reward-complexity frontier after an ad hoc reward transformation for the low-reward condition, used for mental cost modeling.

The NoCost model severely overestimated human policy complexity (Figure 4 Row 1; shown as densities left of the vertical line). In contrast, mental cost consideration in the LinearCost and Quadcost models produced more accurate estimates of human policy complexity across all experiments (Row 1; QuadCost shown as densities to the right of the vertical line). The model fits revealed positive and monotonically increasing mental costs over policy complexity, consistent with intuition (Row 2). Model comparison consistently preferred the QuadCost model across experiments (Table 1), with all ELPD-LOO differences between model pairs significant at the $\alpha = 0.005$ level. These results support the idea that policy complexity incurs a supralinear mental cost.

Table 1: ELPD-LOO of models, summed over participants.

| Exp. | NoCost | LinearCost | QuadCost |
|---|---|---|---|
| 1 | −403.7 | 133.9 | 161.8 |
| 2 | −417.6 | −141.6 | −104.8 |
| 3 | −862.0 | −256.3 | −222.4 |
| 4 | −520.4 | −276.1 | −149.3 |

# Discussion

Across four datasets spanning three distinct task manipulations, we identified a systematic underallocation of cognitive

[3] Experiments 1-3 had binary rewards and did not have this issue.

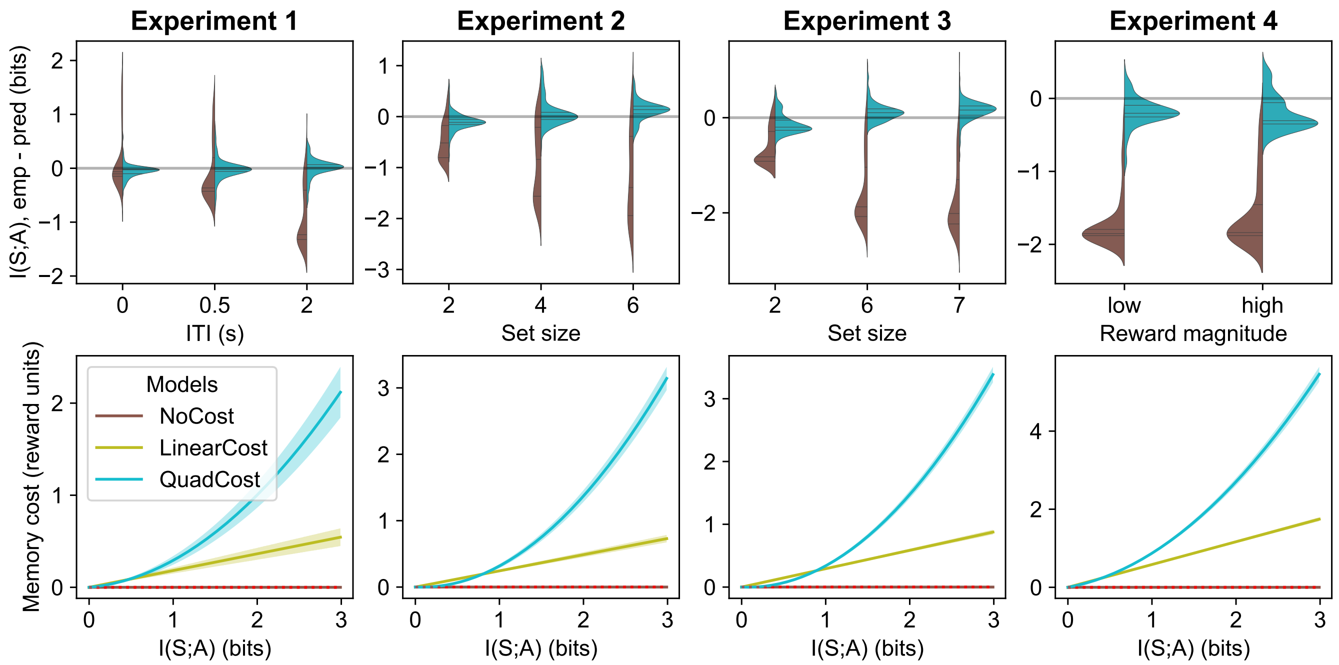

Figure 4: **Mental cost modeling results.** Columns 1-4 correspond to Experiments 1-4. Colors denote models. **Row 1:** Violin plots of the difference between predicted and empirical $I(S;A)$. Densities to the left and right of the vertical line correspond to the NoCost and QuadCost models. **Row 2:** Mean±SEM of fitted MentalCost$(I(S;A))$ functions across participants.

resources. In contrast, models that incorporate mental costs could better explain human decisions. These results confirm that humans meta-cognitively account for their mental effort during contextual decisions, akin to rational inattention (Gershman & Burke, 2023) and rational meta-reasoning accounts (Lieder & Griffiths, 2017). In our experiments, this mental cost reflects the burden of memorizing state-action mappings. These findings raise additional process-level questions about how humans dynamically determine their policy complexity level (e.g., memory encoding strength) and optimize their memory representations over time—an area that requires further theory development and computational modeling.

Our key contribution here is revealing human sensitivity to a specific mental cost formulation—policy complexity—in contextual decision-making, and assessing the cost function's functional form. Such cost sensitivity is independent from time cost considerations, and humans determine their $I^{\pi}(S;A)$ according to both. Model comparison favored a quadratic mental cost in policy complexity, which differs from typical linear assumptions (Gershman & Bhui, 2020; Prat-Carrabin & Woodford, 2024). Interestingly, neuroscience studies have similarly suggested that the metabolic cost of information flow across synapses is supralinear (Laughlin, 2001). Because individual synapses convey little information due to noise, neurons rely on many synapses; however, many synapses transmit identical signals, increasing the cost per bit. The success of the QuadCost model raises questions about past attempts to quantify a hard limit to human memory (Miller, 1956). While such a hard limit certainly exists (the brain contains finite neurons), empirically identifying this limit is rendered difficult by the continuous mental cost of memory. The QuadCost model implies that each marginal increase in policy complexity incurs greater and greater costs; hence humans may choose not to pay increasing costs under increasingly memory-intensive tasks, even if they possess the capacity to encode information at a higher bit rate. However, the resulting "soft ceiling" is mainly a theoretical argument, as the observed increase in policy complexity from set size 6 to 7 suggests that participants have not reached this ceiling. To better understand this nonlinear relationship, future studies should explore non-memory-intensive tasks and develop neurobiological models consistent with such supralinear information costs (Badre et al., 2021; Laughlin et al., 1998).

One advantage of our abstract, information-theoretic formulation is that it can be readily applied to various task setups, parallel to the diverse applications of rate-distortion theory. This means future studies can attempt to define the functional form of mental costs in diverse domains including perception, visual working memory, generalization, and intertemporal choice (Bhui et al., 2021; Jakob & Gershman, 2023; Sims, 2015; Sims et al., 2012). In this way, the universality of an information-theoretic formulation of mental costs can be studied directly. In the future, we hope to develop a more rigorous and comprehensive account of the mental cost of policy complexity, so as to deepen our understanding of mental effort and their instantiations in human behavior.

## Acknowledgments

We are grateful to Jan Drugowitsch for helpful discussions. This work was supported by a Harvard Brain Science Initiative Bipolar Disorder Seed Grant, grants R25MH094612 and R25MH135837 from the National Institute of Mental Health, the National Science Foundation under Grant No. DRL-2024462, the Harvey Fellowship, and the Kempner Institute for the Study of Natural and Artificial Intelligence.

## References

Badre, D., Bhandari, A., Keglovits, H., & Kikumoto, A. (2021). The dimensionality of neural representations for control. *Current Opinion in Behavioral Sciences*, *38*, 20–28.

Balci, F., Simen, P., Niyogi, R., Saxe, A., Hughes, J. A., Holmes, P., & Cohen, J. D. (2011). Acquisition of decision making criteria: Reward rate ultimately beats accuracy. *Attention, Perception, & Psychophysics*, *73*, 640–657.

Bhui, R., Lai, L., & Gershman, S. J. (2021). Resource-rational decision making. *Current Opinion in Behavioral Sciences*, *41*, 15–21.

Drugowitsch, J., DeAngelis, G. C., Angelaki, D. E., & Pouget, A. (2015). Tuning the speed-accuracy trade-off to maximize reward rate in multisensory decision-making (T. Behrens, Ed.). *eLife*, *4*.

Gershman, S. J. (2020). Origin of perseveration in the trade-off between reward and complexity. *Cognition*, *204*, 104394.

Gershman, S. J., & Bhui, R. (2020). Rationally inattentive intertemporal choice. *Nature communications*, *11*(1), 3365.

Gershman, S. J., & Burke, T. (2023). Mental control of uncertainty. *Cognitive, Affective, & Behavioral Neuroscience*, *23*(3), 465–475.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of experimental psychology*, *4*(1), 11–26.

Huffman, D. A. (1952). A method for the construction of minimum-redundancy codes. *Proceedings of the IRE*, *40*(9), 1098–1101.

Hutter, M. (2001). Distribution of mutual information. *Advances in neural information processing systems*, *14*.

Jakob, A. M., & Gershman, S. J. (2023). Rate-distortion theory of neural coding and its implications for working memory. *Elife*, *12*, e79450.

Kool, W., & Botvinick, M. (2018). Mental labour. *Nature human behaviour*, *2*(12), 899–908.

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.

Krebs, R. M., & Woldorff, M. G. (2017). Cognitive control and reward. *The Wiley handbook of cognitive control*, 422–439.

Lai, L., & Gershman, S. J. (2021). Policy compression: An information bottleneck in action selection. In *Psychology of learning and motivation* (pp. 195–232, Vol. 74). Elsevier.

Lai, L., & Gershman, S. J. (2024). Human decision making balances reward maximization and policy compression. *PLOS Computational Biology*, *20*(4), e1012057.

Laughlin, S. B. (2001). Energy as a constraint on the coding and processing of sensory information. *Current opinion in neurobiology*, *11*(4), 475–480.

Laughlin, S. B., de Ruyter van Steveninck, R. R., & Anderson, J. C. (1998). The metabolic cost of neural information. *Nature neuroscience*, *1*(1), 36–41.

Lieder, F., & Griffiths, T. L. (2017). Strategy selection as rational metareasoning. *Psychological review*, *124*(6), 762.

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and brain sciences*, *43*, e1.

Liu, S., Lai, L., Gershman, S. J., & Bari, B. A. (2024). Time and memory costs jointly determine a speed-accuracy trade-off and set-size effects.

Maćkowiak, B., Matějka, F., & Wiederholt, M. (2023). Rational inattention: A review. *Journal of Economic Literature*, *61*(1), 226–273.

Marr, D. (2010). *Vision: A computational investigation into the human representation and processing of visual information*. MIT press.

Miller, G. A. (1956). The magic number seven plus or minus two: Some limits on our capacity for processing information. *Psychological review*, *63*, 91–97.

Parush, N., Tishby, N., & Bergman, H. (2011). Dopaminergic balance between reward maximization and policy complexity. *Frontiers in Systems Neuroscience*, *5*, 22.

Prat-Carrabin, A., & Woodford, M. (2024). Endogenous precision of the number sense. *bioRxiv*, 2024–03.

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*, 99–124.

Simon, H. A. (1978). *Information-processing theory of human problem solving*. Erlbaum Hillsdale, NJ.

Sims, C. R. (2015). The cost of misremembering: Inferring the loss function in visual working memory. *Journal of vision*, *15*(3), 2–2.

Sims, C. R. (2016). Rate–distortion theory and human perception. *Cognition*, *152*, 181–198.

Sims, C. R., Jacobs, R. A., & Knill, D. C. (2012). An ideal observer analysis of visual working memory. *Psychological review*, *119*(4), 807.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Umemoto, A., & Holroyd, C. B. (2015). Task-specific effects of reward on task switching. *Psychological research*, *79*, 698–707.