Optimal decision-making under task uncertainty: a computational basis for cognitive stability versus flexibility

Seth Madlon-Kay (seth.madlonkay@duke.edu)
Department of Neurobiology, Duke University
Durham, NC 27708

John Pearson (john.pearson@duke.edu)
Department of Neurobiology, Duke University
Durham, NC 27708

Tobias Egner (tobias.egner@duke.edu)
Department of Psychology and Neuroscience, Duke University
Durham, NC 27708

Abstract

Cognitive control is thought to regulate the conflict between stability—maintaining the current task in the face of distraction—and flexibility—switching to a new task of greater priority. However, evidence conflicts regarding when and to what extent stability and flexibility trade-off. A normative theory of flexibility and stability may help clarify when and why we should expect such trade-offs to occur. Towards such a theory, we model task-switching as a problem of decision-making under uncertainty, in which the decision-maker must simultaneously infer both the identity of a stimulus and the task governing the correct response to that stimulus. We find that optimal behavior is either extremely stable or extremely flexible, but not both, indicating a normative basis for a trade-off between the two. However, we also show that a sub-optimal but more realistic decision-maker exhibits behavior between these two extremes, and more closely resembles experimental data.

Keywords: cognitive control; decision making; computation; drift-diffusion models

Introduction

Suppose you are writing a report on your recent research for an imminent deadline. While you work, your co-workers gossip loudly. You ignore the gossip and focus on your report. You have exhibited cognitive stability: instead of reacting to an irrelevant stimulus, you have continued executing the current task. Then, one of your co-workers says that the building is on fire and everyone must evacuate immediately. You stop working on the report and leave. You have exhibited cognitive flexibility: you put aside your previous task and switched to a new one when it was adaptive to do so. Extant accounts of cognitive stability and flexibility typically assume that the two are opposed. The more strongly you focus on your report (high stability), the harder it is to switch away from it to fleeing the fire (low flexibility). While many empirical findings support this supposition (Dreisbach & Goschke, 2004; Dreisbach & Fröber, 2019), this support is not unanimous, and some recent findings have suggested that stability and flexibility can, at least under some circumstances, vary independently (Geddert & Egner, 2022).

When the empirical record conflicts, a computational approach may clarify. Is it obligatory that stability and flexibility trade off against each other? Are specific assumptions required to produce such a trade-off, or is such a trade-off required of adaptive behavior? A theory that addresses these questions could help us understand when and why we observe stability-flexibility tradeoffs in human behavior. However, existing computational theories in this domain tend to assume that such trade-offs exist, building a conflict between stability and flexibility into the architecture of the models themselves (Ueltzhöffer, Armbruster-Genç, & Fiebach, 2015; Musslick, Jang, Shvartsman, Shenhav, & Cohen, 2018; Musslick & Cohen, 2021). While such accounts provide valuable insights into neural and cognitive systems that may exhibit such trade-offs, they do not address questions of if or when such trade-offs should actually occur. In contrast, here we take a decision-theoretic approach, asking whether a trade-off between stability and flexibility emerges naturally from normative principles of optimal decision-making. We find that such a trade-off can emerge from decision-making under uncertainty: if both the stimulus and the task are unknown, and must be concurrently inferred from the environment, then stability and flexibility manifest as opposing strategies for resolving uncertainty.

Quantifying stability and flexibility

Before proceeding, we should specify precisely what we mean by "stability" and "flexibility." We operationalize "stability" as the slowing of reaction times that occurs when two tasks disagree on the correct response to a single stimulus. In multitasking experiments where humans alternate between two tasks, they respond faster to stimuli for which both tasks dictate the same action; such stimuli are termed "congruent". Conversely, for "incongruent" stimuli where the two tasks demand different actions, responses are slower and less accurate (MacLeod, 1991; Mante, Sussillo, Shenoy, & Newsome, 2013). The difference in reaction times between incongruent and congruent trials, or the "congruency effect", is our measure of stability, in that a smaller congruency effect indicates a more stable focus on the current task.

We similarly operationalize "flexibility" as the slowing of reaction times which occurs when the correct task changes from one trial to the next. In multitasking experiments where

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

participants are cued on each trial to perform a particular task, responses are faster on trials where the cued task is a "repeat" of the previous trial's task; conversely, responses are slower on "switch" trials where the task differs from that of the previous trial (Rogers & Monsell, 1995; Monsell, Sumner, & Waters, 2003). The difference in reaction times between switch and repeat trials, or the "switch cost", is our measure of flexibility, whereby a smaller switch cost indicates a greater degree of flexibility.

For our decision-theoretic analysis of stability and flexibility, we model a multi-tasking experiment which permits both stimulus (in)congruency and task switching.

Multitasking as sequential hypothesis testing

In order to perform a decision-theoretic analysis of a multitasking experiment, we build on a long tradition in psychology and neuroscience linking binary choices to statistical inference. Specifically, we extend the sequential hypothesis testing (SHT) framework, widely used to model binary decisions, to describe optimal inference under multitasking.

Single task model

We begin by describing a single-task version of our model, which we then extend to describe multi-tasking. Consider a simple perceptual discrimination task. A task-*relevant* stimulus R is presented, which can take one of two values: $R \in \{r_A, r_B\}$. An agent must then choose between two simple hypotheses: $H_A : R = r_A$ versus $H_B : R = r_B$. At each time point t the agent observes a noisy piece of binary evidence, $x_{r,t} \in \{-1, 1\}$, pertaining to the stimulus:

$$p(x_{r,t} = 1|R) = \begin{cases} \sigma_r & \text{if } R = r_A \\ 1 - \sigma_r & \text{if } R = r_B \end{cases}, \qquad \sigma_r \geq 0.5. \quad (1)$$

To choose between H_A and H_B, the decision-maker need only track a running sum y_r of the evidence they have observed thus far:

$$y_{r,t+1} = y_{r,t} + x_{r,t}, \qquad y_{r,1} = 0. \quad (2)$$

At each time step t, the agent chooses an action $a_t \in \{a_A, a_B, a_R\}$: actions a_A or a_B accept hypotheses A and B, respectively, ending the trial; and action a_R advances t by one step and samples one more piece of evidence. If the agent accepts a hypothesis, their expected payoff is the probability that the hypothesis is true; for example, the payoff for a_A is

$$\begin{aligned} u(a_A, y_r) := p(H_A|y_r) &= \text{logit}^{-1}(y_r \eta_r) \\ &= \frac{1}{1 + \exp -y_r \eta_r}, \\ \eta_r = \log \frac{\sigma_r}{1 - \sigma_r}&. \end{aligned} \quad (3)$$

If the agent instead chooses action a_R and samples more evidence from the stimulus, they pay a fixed sampling cost $u(a_R, y_r) = \kappa$, $\kappa < 0$. This sampling cost forces the agent to trade off speed and accuracy, with larger magnitudes of κ pushing the agent to favor speed. The challenge for a decision-maker is to know for what values of y_r they should continue sampling, and at what values they should accept a hypothesis.

The optimal policy in this setting has a simple form: continue to sample until the accumulated evidence y_r passes a threshold, at which point choose the favored hypothesis (Wald & Wolfowitz, 1948). Readers may recognize this as a discrete-time version of the drift-diffusion model (DDM), widely used to model binary decisions (Ratcliff, 1978; Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006); the term σ_r acts as the drift rate, determining how rapidly and reliably evidence accumulates.

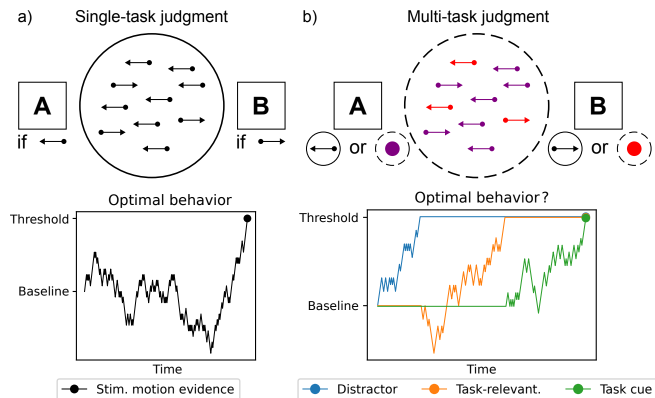

Figure 1: Modeling multitasking as sequential hypothesis testing. **a)** A single-task judgment, such as motion discrimination, can be modeled as the sequential testing of two simple hypotheses, A or B, each corresponding to a direction of motion. Optimal behavior corresponds to a drift-diffusion process where evidence for preferring A over B accumulates until a threshold is reached. **b)** A multitask judgment. Hypotheses A and B correspond to combination of one of the two stimulus dimensions (color and motion) and the task cue (for illustration purposes, a dashed versus solid aperture). A naive strategy is shown of sampling each source of evidence sequentially up to a fixed threshold. This strategy is likely inefficient, but the optimal strategy is not known.

Multitask model

Next, we extend our formulation to accommodate multitasking, where the optimal policy will be more complicated. Consider a stimulus with two attributes, one of which is relevant (and which we continue to label R), the other being an irrelevant distractor which we label D. Which attribute is relevant and which is irrelevant depends on the current task. We assume that the distractor also takes on two values $D \in \{d_A, d_B\}$, favoring hypothesis A or B, respectively, and that the task must be inferred from a *task cue* $C \in \{c_R, c_D\}$, which, *a priori*, can favor either the task-relevant or the distractor dimension of the stimulus. We assume that the task cue is presented at the same time as the task stimulus. On each time step, in addition to $x_{r,t}$, there exist pieces of evidence $x_{d,t} \in \{-1,1\}$ and $x_{c,t} \in \{-1,1\}$ pertaining to the true identity of D and C:

$$\begin{aligned} p(x_{d,t}=1|D) &= \begin{cases} \sigma_d & \text{if } D = d_A \\ 1-\sigma_d & \text{if } D = d_B \end{cases}, \qquad \sigma_d = \sigma_r, \\ p(x_{c,t}=1|C) &= \begin{cases} \sigma_c & \text{if } C = c_R \\ 1-\sigma_c & \text{if } C = c_D \end{cases}, \qquad \sigma_c \geq 0.5. \end{aligned} \tag{4}$$

Note that we assume that, as the agent lacks *a priori* knowledge of the task cue, both the task-relevant dimension R and the distractor dimension D have the same drift rate; we will refer to their shared drift rate as σ_s.

At each time point the agent can terminate the trial by choosing a_A or a_B, or instead choose to gather more evidence. We assume our agent has *selective attention* and can only sample one source of evidence at a time; they do this by choosing a_R, a_D, or a_C, which sample the task-relevant dimension, distractor dimension, or task cue respectively. The accumulated evidence y for each of R, D, and C only advances when the agent chooses to sample the corresponding evidence source:

$$y_{j,t+1} = \begin{cases} y_{j,t} + x_{j,t} & \text{if } a_t = a_j \\ y_{j,t} & \text{if } a_t \neq a_j \end{cases}, \qquad \forall j \in \{R,D,C\}. \tag{5}$$

Finally, H_A is correct if $R = r_A$ and $C = c_r$, *or* if $D = d_A$ and $C = C_d$. Accordingly, the expected payoff for accepting H_A (or H_B) now depends on the probabilities of all three unknown aspects of the multitask experiment:

$$\begin{aligned} u(a_A, \mathbf{y}) := p(H_A|\mathbf{y}) = \text{logit}^{-1}(y_r\eta_r) \times \text{logit}^{-1}(y_c\eta_c) + \\ \text{logit}^{-1}(y_d\eta_d) \times (1 - \text{logit}^{-1}(y_c\eta_c)). \end{aligned} \tag{6}$$

How should an optimal decision-maker choose which pieces of evidence to sample? There is no simple solution for the optimal policy available to us here. However, given values for σ_s, σ_c, and the sampling cost κ, we can solve numerically for the optimal policy using dynamic programming techniques (Sutton & Barto, 2018). Specifically, we used value iteration to recursively determine the expected future payoff from each state, repeating value updates until expected payoffs converged. Then, in order to characterize the nature of the optimal policies, we simulated agents performing a multitask trial, choosing the action with the maximum expected future payoff at each time point, until they accepted a hypothesis. Unless otherwise noted, all reported statistics are averages across 10,000 simulations of an optimal decision-maker. To make computation tractable, we restricted the range of the y values to lie between -30 and 30.

Results

Instability in optimal behavior

First we ask whether optimal policies can exhibit cognitive instability (Fig. 2). In the multitasking paradigm used here, instability is measured by the congruency effect: whether choices are impaired when the task-relevant stimulus R and distractor D favor different responses (incongruent stimulus).

We find that the stability of optimal behavior depends critically on the relative drift rates of the stimulus, σ_s, and the task cue, σ_c. When σ_s is the same as σ_c, the optimal policy is to sample evidence from the task cue early in the trial, then to transition to sampling evidence from the task-relevant stimulus dimension; the distractor stimulus dimension is rarely, if ever, sampled. As a consequence, the (in)congruency of the distractor is of no consequence, and reaction times are identical to congruent and incongruent stimuli. Accordingly, behavior is highly stable.

When σ_c is comparatively low, the optimal policy shows qualitative differences. Instead of sampling the task cue first, optimal behavior involves sampling both the task-relevant and task-irrelevant stimulus dimensions at similar rates. Then, if the agent finds that the stimulus is congruent, a response can be chosen without any evidence regarding the task cue. However, if the agent instead infers that the stimulus is incongruent, then they transition to sampling the task cue. As a result, reaction times show a large congruency effect, a hallmark of cognitive instability.

Flexibility in optimal behavior

We next ask whether optimal policies can exhibit cognitive flexibility (Fig. 3). In the multitasking paradigm used here, flexibility is measured by switch costs: whether choices are slower when the current task differs from the previous task. Our model has no explicit representation of trial-to-trial transitions, so we implement "pseudo" switch and repeat trials by lowering or raising the prior probability of the task cue, $p(C = c_R)$. Specifically, we set $p(C = c_R) = 0.8$ for repeat trials, and $p(C = c_R) = 0.2$ for switch trials. In other words, we assume that the previous task acts as a cue for the current task, with the agent believing that the task which occurred previously is more likely to occur again. Assumptions of this sort are common in, e.g., dynamic belief models of sequential effects across a variety of domains (A. J. Yu & Cohen, 2008; Nguyen, Josić, & Kilpatrick, 2019), including task-switching in particular (Jiang, Wagner, & Egner, 2018). The stimulus dimensions R and D are assumed to be congruent.

As before, we find that the flexibility of optimal behavior depends critically on the relative drift rates σ_s, and σ_c. When σ_c matches σ_s, the optimal policy is again to sample the task cue, then proceed to sample the relevant stimulus dimension. On "repeat" trials, the starting point for the task cue evidence y_c is elevated in the direction of drift, so the repeat facilitates faster responses; conversely, on "switch" trials, y_c is shifted away from the direction of drift, leading to slower responses. As a consequence, optimal behavior exhibits switch costs and looks "inflexible".

In contrast, when σ_c is low the optimal agent chooses to sample the stimulus dimension favored by the prior first, followed by the other stimulus dimension. The agent rarely samples the task cue; accordingly, the impact of the switch versus

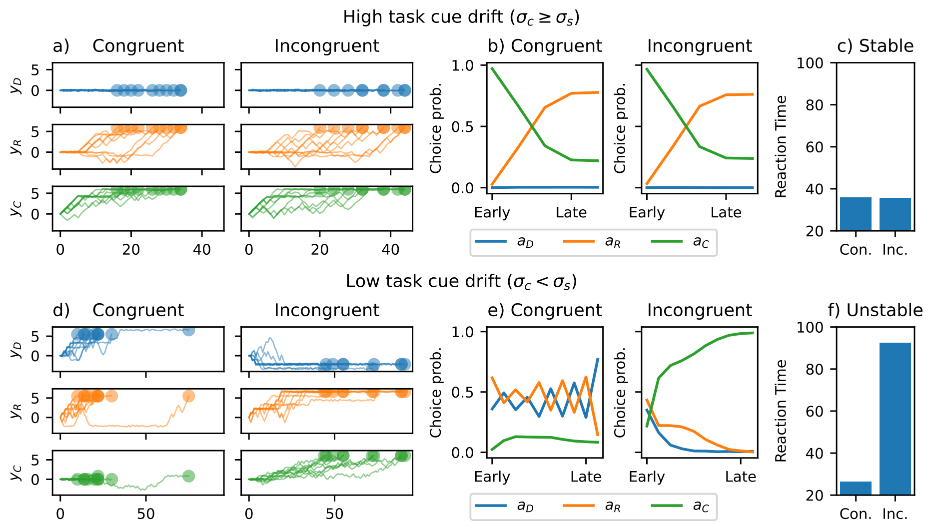

Figure 2: Optimal decision-making can be stable (top) or unstable (bottom). **a-c)**: Behavior when task cue and stimulus have the same drift rate, $\sigma_c = \sigma_s = 0.7$. **a)** Example diffusion paths showing how accumulated evidence y evolves over the course of a trial for each aspect of the multitask: the distractor stimulus D (top), the task-relevant stimulus R (middle), and task cue C (bottom). Circles indicate when each diffusion path terminates. We compare example diffusion paths depending on whether the distractor is congruent with the relevant stimulus, $(D,R) = (d_A, r_A)$, or incongruent, $(D,R) = (d_B, r_A)$. **b)** The probability of that an agent will choose to sample the distractor (action a_D), the task-relevant stimulus, (a_R), or the task cue (a_C), at different epochs over the course of a trial. Each simulated diffusion path is binned into 5 sequential epochs; choice rates are first calculated in each epoch for each simulation separately, then averaged within epochs across simulations such that, *e.g.*, the last epoch is aligned with the end of trial regardless of trial length. **c)** Average reaction times (equivalently, trial length) for congruent versus incongruent stimuli. In this environment, the agent never samples the distractor stimulus, so behavior is stable and there is no congruency effect. **d-f)** Behavior when $(\sigma_c, \sigma_s) = (0.6, 0.75)$. Here, the agent samples both the relevant and distractor stimulus dimensions first, leading to high congruency effects.

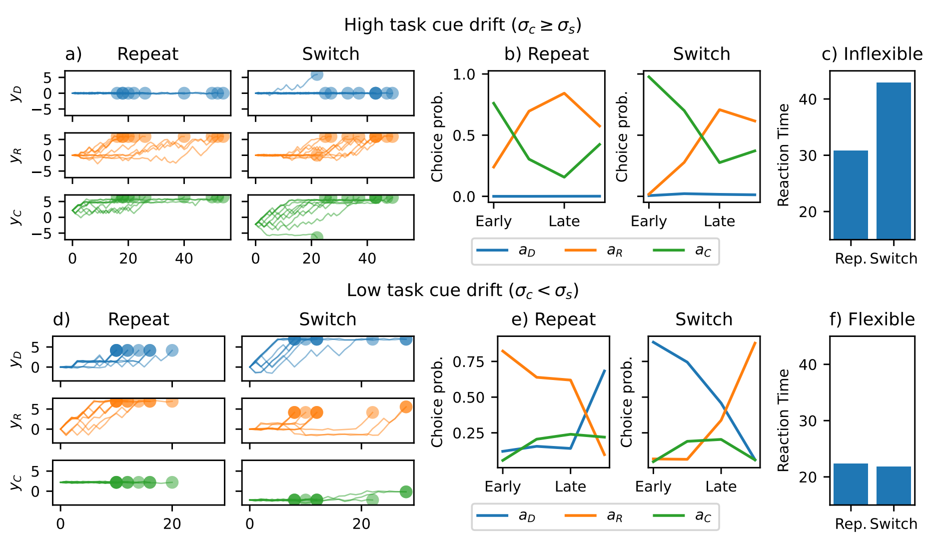

Figure 3: Optimal decision-making can be flexible (bottom) or inflexible (top). **a-c)**: Behavior when task cue and stimulus have the same drift rate, $\sigma_c = \sigma_s = 0.7$. Because the previous trial provides information for the current task, the baseline for y_C is increased for repeat trials and decreased for switch trials. This leads to switch costs (inflexibility) because the agent chooses to sample the task cue first. **d-f)** Behavior when $(\sigma_c, \sigma_s) = (0.6, 0.8)$. The agent samples both stimulus dimensions instead of the task cue. Accordingly, the starting point of y_c does not matter, and there is no switch cost.

repeat has little effect, leading to behavior that appears "flexible" and robust to switch costs.

An abrupt transition between flexibility and stability

Thus far we have shown that a high σ_c versus σ_s leads to behavior that is stable but inflexible, while low σ_c (high σ_s) leads to optimal behavior that is unstable but flexible. This would appear to support a trade-off between stability and flexibility; however, we have so far shown only two hand-picked sets of σ_c and σ_s values.

To determine whether a stability-flexibility tradeoff holds more generally across the range of potential parameters, we systematically vary σ_c and σ_s across a 9×9 grid of values ranging from very high (0.95) to very low (0.55). For each combination of σ_s and σ_c we estimate the relative congruency effect $(RT_{inc} - RT_{con})/RT_{con}$, and the relative switch cost $(RT_{switch} - RT_{rep})/RT_{rep}$ (Fig. 4).

The results show a sharp transition where, as σ_s exceeds σ_c, congruency effects emerge abruptly from non-existent to rapidly growing, leading to incongruent trials that are an order of magnitude slower than congruent trials. Switch costs are more variable and lower in magnitude than congruency effects, but also show a sharp transition from low to high switch costs in the same region. As a consequence, we see an apparent "ridge" where nearby environments show starkly opposed patterns of behavior, as illustrated in Figure 4b with environments F and S. However, switch costs diminish as σ_c increases, likely because the increasing evidence accumulation rate of the task cue renders the information provided by the previous trial irrelevant.

Finally, examining more closely the relationship between switch costs and congruency effects in state S, we notice that while congruency effects clearly dominate, we also observe a switch-by-congruency interaction whereby congruency effects become worse on switch trials. This pattern is sometimes observed in experimental data (Rogers & Monsell, 1995; Kiesel et al., 2010); here it arises because the task cue takes longer to resolve on switch trials.

Realistic behavior from a suboptimal decision-maker

The optimal behaviors shown above exhibit at least two features which contrast with typical human behavior in multitasking experiments. First, optimal agents show a sharp shift from extreme flexibility (high congruency effects without switch costs) to extreme stability (high switch costs without congruency effects). They also show far greater congruency effects than switch costs. Human behavior, by contrast, typically shows both switch costs and congruency effects together, and in similar orders of magnitude (Geddert & Egner, 2022). Moreover, in optimal agents, any non-zero congruency effects require that the task cue be harder to resolve than the stimulus. In typical multitasking experiments, by contrast, the stimuli and the task cue are of similar visual clarity and salience, so it is unlikely that this requirement is satisfied.

We next asked whether more human-like performance could be obtained in our model by implementing a more realistic action selection mechanism. In our previous results, the optimal decision-maker always picks the action with the highest value, even when multiple actions have extremely similar payoffs. A real decision-maker, by contrast, is unlikely have accurate enough knowledge of the environment to resolve very small differences in expected payoffs. In practice, knowledge of the environment is imperfect, and achieving good performance involves some amount of exploration and trial-and-error. Because of this, most practical decision-making systems involve some amount of randomness in their choices (Sutton & Barto, 2018). One mechanism for achieving this is by introducing a softmax decision-rule, which is biased towards actions with higher expected payouts but leaves some probability of choosing each action; the softmax rule is frequently used in computational models of value-based decision-making in humans and animals (Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006).

We find that when stimulus and task cue drift rates are similar, modifying the decision-maker to use a softmax choice rule results in a shift from extreme stability to a more balanced policy exhibiting intermediate levels of both stability and flexibility (Fig. 5). Systematically varying the amount of noise in the softmax choice shows that as variability increases, congruency effects emerge without eliminating the switch costs exhibited by the optimal agent.

Discussion

We began this manuscript by asking whether the trade-off between stability and flexibility is obligatory. While our investigations do not give a definitive answer, we believe they provide a useful and clarifying perspective on the question, and we give our own thoughts on the matter here.

In a decision-theoretic framework, we find that stability and flexibility correspond to alternative strategies for resolving uncertainty. A "stable" strategy resolves the task cue first, and only once the task is certain goes on to resolve the relevant stimulus dimension; a "flexible" strategy instead resolves stimulus (in)congruency first, then resolves the task cue only if necessary. The most extreme versions of these strategies are incompatible; if one exclusively prioritizes the task cue, one cannot also exclusively prioritize the (in)congruency of the stimulus. In this sense, some kind of trade-off is indeed obligatory.

However, "stability" and "flexibility" are not the only strategies an agent could use, and neither is a decision-maker obligated to slavishly adhere to one strategy or another. Though optimal agents in our setting do, on *average*, look extremely flexible or extremely stable, in any *individual trial* the randomness of the diffusion path may lead to unexpected behavior. Looking closely at individual diffusion paths in Figs 2 and 3, for example, one can see examples where the agent has come to an incorrect conclusion regarding some aspect of the decision task, and is led down a garden path of contradic-

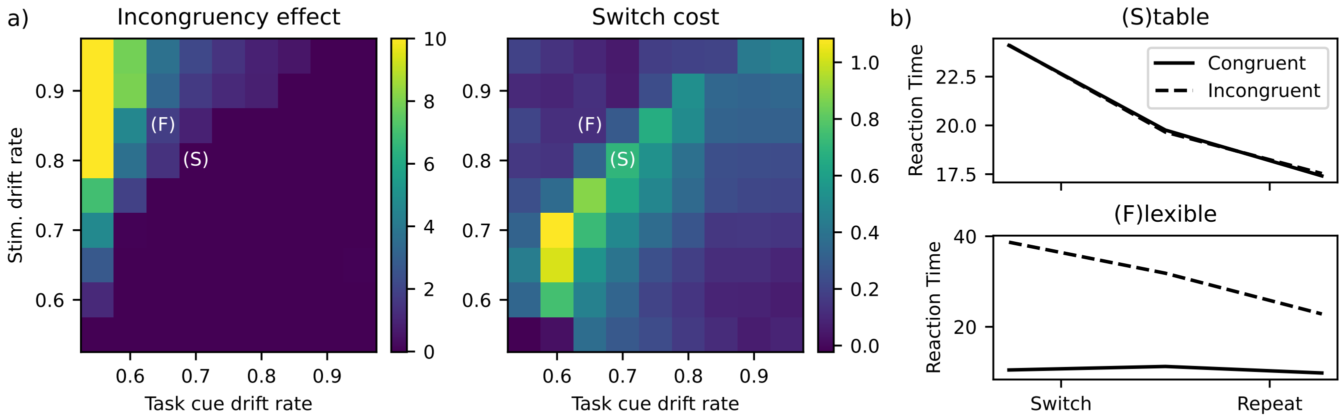

Figure 4: Stability and flexibility trade-off. **a)** We estimate the congruency effects (left) and the switch costs (right) under optimal behavior as a function of the stimulus (both task-relevant and distractor) drift rate σ_s and the task cue drift rate σ_c. Comparing the congruency effect and switch costs maps indicates a sharp transition between congruency effects dominating and switch costs dominating. **b)** Two nearby environments, F and S, on either side of the transition showing opposite behavior patterns. F has parameters $(\sigma_s, \sigma_c) = (0.65, 0.85)$, while S has parameters $(\sigma_s, \sigma_c) = (0.7, 0.8)$.

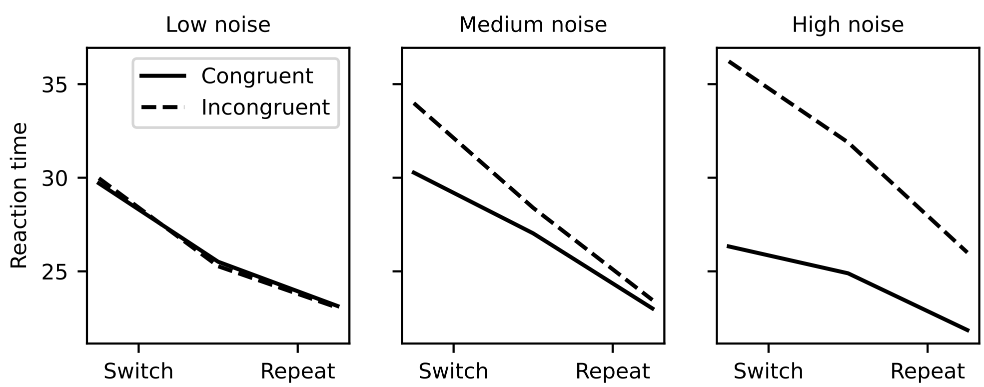

Figure 5: Switch costs and congruency effects in agents using a softmax choice rule, with varying levels of decision noise, at $\sigma_c = \sigma_s = 0.7$. The low, medium, and high noise agents used inverse temperatures of 10^5, 10^3, and 100 respectively.

tory evidence forcing them to deviate from strict flexibility or stability. Furthermore, whether extreme stability or extreme flexibility are optimal can depend on very small differences in the underlying multitask environment; in Figure 4 we saw that even the small move from (F) to (S) induced dramatic shifts in optimal policy. In any realistic environment, an agent likely could not depend on knowing for certain which extreme strategy were optimal.

If extreme flexibility and stability are not practically and consistently achievable, we may expect more realistic decision-makers to end up somewhere in the middle. Indeed, when an agent adopts a softmax decision rule that forces them away from extremes, their behavior does appear to match at least the crudely stylized facts of the experimental literature: switch costs and congruency effects co-occur in roughly similar magnitudes. However, the middle between two extremes is big and varied place and may contain strategies that are unrelated to the distinction between flexibility and stability. To put it concisely, we propose that flexibility and stability do necessarily trade-off at their extreme manifestations, but that those extremes may not always be relevant. To what extent the trade-off holds in the space between those extremes is yet unknown.

The decision-theoretic model of multi-tasking we developed here is poorly suited to exploring that middle space, as it predisposes agents to ruthlessly optimize for hyper-specific circumstances. To probe this space, we may look to the literatures in cognitive neuroscience regarding how people make decisions across unknown and volatile environments, such as in how people learn and explore complex maps (Botvinick, Weinstein, Solway, & Barto, 2015; Behrens et al., 2018) or track constantly evolving temporal dynamics (Ryali, Reddy, & Yu, 2018; L. Q. Yu, Wilson, & Nassar, 2021; Wen, Geddert, Madlon-Kay, & Egner, 2023). How does such open-ended learning interact with the need to adaptively prioritize certain tasks? Further research is, as ever, still necessary.

Acknowledgments

The authors would like to thank Raphael Geddert, who directly inspired this project with his own research on and generous conversations regarding cognitive stability and flexibility; Juniper Walker-Kay, who indirectly inspired this research by demonstrating the development of stability and flexibility in young children; Ashley Walker for her unfailing support and encouragement; and the National Institute of Mental Health (R01MH133550) for funding this research.

References

Behrens, T. E. J., Muller, T. H., Whittington, J. C. R., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018). What is a cognitive map? Organizing knowledge for flexible behavior. *Neuron*, *100*(2), 490–509. doi: 10.1016/j.neuron.2018.10.002

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, *113*(4), 700–765. doi: 10.1037/0033-295X.113.4.700

Botvinick, M., Weinstein, A., Solway, A., & Barto, A. (2015). Reinforcement learning, efficient coding, and the statistics of natural tasks. *Current Opinion in Behavioral Sciences*, *5*, 71–77. doi: 10.1016/j.cobeha.2015.08.009

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*(7095), 876–9. doi: 10.1038/nature04766

Dreisbach, G., & Fröber, K. (2019). On how to be flexible (or not): Modulation of the stability-flexibility balance. *Current Directions in Psychological Science*, *28*(1), 3–9. doi: 10.1177/0963721418800030

Dreisbach, G., & Goschke, T. (2004). How positive affect modulates cognitive control: Reduced perseveration at the cost of increased distractibility. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 343–353. doi: 10.1037/0278-7393.30.2.343

Geddert, R., & Egner, T. (2022). No need to choose: Independent regulation of cognitive stability and flexibility challenges the stability-flexibility trade-off. *Journal of Experimental Psychology: General*, *151*(12), 3009–3027. doi: 10.1037/xge0001241

Jiang, J., Wagner, A. D., & Egner, T. (2018). Integrated externally and internally generated task predictions jointly guide cognitive control in prefrontal cortex. *eLife*, *7*, e39497. doi: 10.7554/eLife.39497

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching—a review. *Psychological Bulletin*, *136*(5), 849–874. doi: 10.1037/a0019842

MacLeod, C. (1991). Half a century of reseach on the stroop effect: An integrative review. *Psychological Bulletin*, *109*(2), 163–203. doi: 10.1037/0033-2909.109.2.163

Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, *503*(7474), 78–84. doi: 10.1038/nature12742

Monsell, S., Sumner, P., & Waters, H. (2003). Task-set reconfiguration with predictable and unpredictable task switches. *Memory & Cognition*, *31*(3), 327–342. doi: 10.3758/BF03194391

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

Musslick, S., Jang, S. J., Shvartsman, M., Shenhav, A., & Cohen, J. (2018). Constraints associated with cognitive control and the stability-flexibility dilemma. In *Proceedings of the 40th Annual Meeting of the Cognitive Science Society.*

Nguyen, K. P., Josić, K., & Kilpatrick, Z. P. (2019). Optimizing sequential decisions in the drift–diffusion model. *Journal of Mathematical Psychology*, *88*, 32–47. doi: 10.1016/j.jmp.2018.11.001

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. doi: 10.1037/0033-295X.85.2.59

Rogers, R. D., & Monsell, S. (1995). Costs of a predictible switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, *124*(2), 207–231. doi: 10.1037/0096-3445.124.2.207

Ryali, C., Reddy, G., & Yu, A. J. (2018). Demystifying excessively volatile human learning: A Bayesian persistent prior and a neural approximation. In *Advances in Neural Information Processing Systems 31* (pp. 2781–2790). Curran Associates, Inc.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.

Ueltzhöffer, K., Armbruster-Genç, D. J. N., & Fiebach, C. J. (2015). Stochastic dynamics underlying cognitive stability and flexibility. *PLOS Computational Biology*, *11*(6), e1004331. doi: 10.1371/journal.pcbi.1004331

Wald, A., & Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics*, *19*(3), 326–339. doi: 10.1214/aoms/1177730197

Wen, T., Geddert, R. M., Madlon-Kay, S., & Egner, T. (2023). Transfer of learned cognitive flexibility to novel stimuli and task sets. *Psychological Science*, *34*(4), 435–454. doi: 10.1177/09567976221141854

Yu, A. J., & Cohen, J. D. (2008). Sequential effects: Superstition or rational behavior? In *Advances in Neural Information Processing Systems* (Vol. 21). Curran Associates, Inc.

Yu, L. Q., Wilson, R. C., & Nassar, M. R. (2021). Adaptive learning is structure learning in time. *Neuroscience & Biobehavioral Reviews*, *128*, 270–281. doi: 10.1016/j.neubiorev.2021.06.024