Research Article

Psychological Science
1–20

DOI: 10.1177/09567976221141854
www.psychologicalscience.org/PS

Transfer of Learned Cognitive Flexibility to Novel Stimuli and Task Sets

Tanya Wen[1], Raphael M. Geddert[1], Seth Madlon-Kay[2], and Tobias Egner[1,3]

[1]Center for Cognitive Neuroscience, Duke University; [2]Department of Biostatistics and Bioinformatics, Duke University School of Medicine; and [3]Department of Psychology and Neuroscience, Duke University

Abstract

Adaptive behavior requires learning about the structure of one's environment to derive optimal action policies, and previous studies have documented transfer of such structural knowledge to bias choices in new environments. Here, we asked whether people could also acquire and transfer more abstract knowledge across different task environments, specifically expectations about cognitive control demands. Over three experiments, participants (Amazon Mechanical Turk workers; N = ~80 adults per group) performed a probabilistic card-sorting task in environments of either a low or high volatility of task rule changes (requiring low or high cognitive flexibility, respectively) before transitioning to a medium-volatility environment. Using reinforcement-learning modeling, we consistently found that previous exposure to high task rule volatilities led to faster adaptation to rule changes in the subsequent transfer phase. These transfers of expectations about cognitive flexibility demands were both task independent (Experiment 2) and stimulus independent (Experiment 3), thus demonstrating the formation and generalization of environmental structure knowledge to guide cognitive control.

Keywords

cognitive flexibility, reinforcement learning, generalization, task switching, meta-flexibility, open data, open materials

Received 5/5/22; Revision accepted 11/3/22

Adaptive behavior requires us to identify and keep in mind the currently relevant "rules of the game"—that is, which responses to which stimuli likely lead to desirable outcomes (also known as task sets; Monsell, 2003). Moreover, given that the world is ever-changing, optimal regulation of task sets involves resolving the tradeoff of needing to implement the current task set and shielding it from distraction (cognitive stability) versus being ready to update (or switch) task sets in response to changing environmental contingencies (cognitive flexibility; Goschke, 2003; Nassar & Troiani, 2020). Importantly, neither stability nor flexibility is inherently beneficial; rather, it is the ability to dynamically adapt one's flexibility level to suit varying environmental demands, referred to as meta-flexibility, that facilitates optimal cognition (Goschke, 2013).

To adjust cognitive flexibility in an optimal manner, one must infer which task sets to use at a given time by observing environmental statistics (Behrens et al., 2007; Yu et al., 2020), such as associations between stimuli, responses, and outcomes (Niv, 2019). The process of learning these associations can be characterized by reinforcement-learning models (Barraclough et al., 2004; Lee et al., 2012; Sutton & Barto, 1998). In the context of task switching, the value to be learned is the likelihood that a given task set is currently relevant. The level of cognitive flexibility that a learner exhibits can be described by their learning rate, which determines the degree to which recent feedback updates their beliefs. Previous studies have shown that people's learning rates are typically low during periods of environmental stability and high during periods of volatility (Behrens et al., 2007; Browning et al., 2015; Jiang et al., 2014, 2015; Massi et al., 2018), although those studies

Corresponding Author:
Tanya Wen, Duke University, Center for Cognitive Neuroscience
Email: tanya.wen@duke.edu

have looked at the direct learning of stimulus-reward associations or proportions of stimulus types. To our knowledge, it has not been tested whether environmental volatility can similarly influence the learning of higher order rules such as task sets.

Moreover, adapting cognitive flexibility is impossible in a previously unobserved environment. How, then, do people set their cognitive flexibility in new situations? We posit that successfully matching cognitive flexibility levels to varying demand contexts could be mediated by learning and transferring knowledge about the demand structure of previous environments to novel environments. For example, while a novel task is learned, it may be beneficial to exploit relevant information acquired in the past (Kemp et al., 2010; Mark et al., 2020; Yu et al., 2020). Previous studies have demonstrated that structural knowledge of an environment in the form of cognitive maps of stimulus associations (Mark et al., 2020) and correlated bandit arms (Baram et al., 2021; Schulz et al., 2020) can foster transferable expectations about the structure of new environments. However, to the best of our knowledge, it has not been tested whether learning parameters driving cognitive control processes, such as task set updating, can be transferred to different contexts.

We here combined these two prior insights—volatility learning and structure transfer—to create a novel test of the acquisition and transfer of cognitive control policies, specifically, one's level of cognitive flexibility or switch readiness. We hypothesized that, first, cognitive flexibility is adjusted in response to environmental volatility and, second, that a level of cognitive flexibility learned in one environment can be transferred to another. Observing such transfer of learned cognitive flexibility would be a novel finding in the fields of decision-making and cognitive control.

To this end, the current study investigated whether participants learning to update task sets faster or slower (i.e., at different learning rates) in one context transfer their expectations to another context. Specifically, we conducted three experiments employing a probabilistic version of the Wisconsin Card Sorting Task (Berg, 1948; Van Eylen et al., 2011), wherein two groups of participants were initially exposed to either a low- or high-volatility learning environment, with seldom versus frequent rule changes, respectively. Next, participants from both groups switched to the same medium-volatility transfer environment, which had an intermediate rate of rule changes. Reinforcement-learning models were fitted to participants' rule-choice behavior to quantify the rates of rule updating (i.e., the learning rates) of the two groups in both task phases. We predicted that participants who encountered the low-volatility condition would have lower learning rates than participants who experienced the high-volatility condition and that these learning rates would generalize to the transfer phase. Across the three experiments, we systematically decreased the task and stimulus overlap to investigate whether the similarity between the learning and transfer phases influenced learning rate transfer.

Statement of Relevance

Many psychologists are interested in training people in generalizable cognitive skills, but learning often does not transfer outside of the specific trained task. One prominent target skill is cognitive flexibility, the ability to switch between different tasks, because many clinical populations have trouble with being flexible. In this study, we investigated a new way of training people in flexibility by having them learn about how frequently task rules change in one context and then testing whether this knowledge about the rate of rule changes would be used in other contexts. We found that participants learned and transferred expectations about rule changes, even to contexts where both the tasks and stimuli were completely different from the training context. This shows that people can extract statistical information about how variable their environment is and then use that information to guide how flexible they are in other environments. This may help develop new cognitive training regimes.

Method

In three experiments, we examined whether participants could acquire and transfer knowledge about cognitive flexibility demands across different contexts. In all experiments, participants were split into two groups (low volatility and high volatility) that completed a learning phase in which the task sets switched less or more frequently, respectively. Next, we tested them in a medium-volatility environment transfer phase in which the switch rate of task rules was the same for both groups. Our main question was whether expectations about the frequency of task-set updating acquired in the learning phase would generalize to the subsequent transfer phase. This study was reviewed and approved by the Duke Campus Institutional Review Board.

Procedure

Figure 1 illustrates two sample trials of the task paradigm in the learning phase of all three experiments. On each trial, three cards arranged in a pyramid were

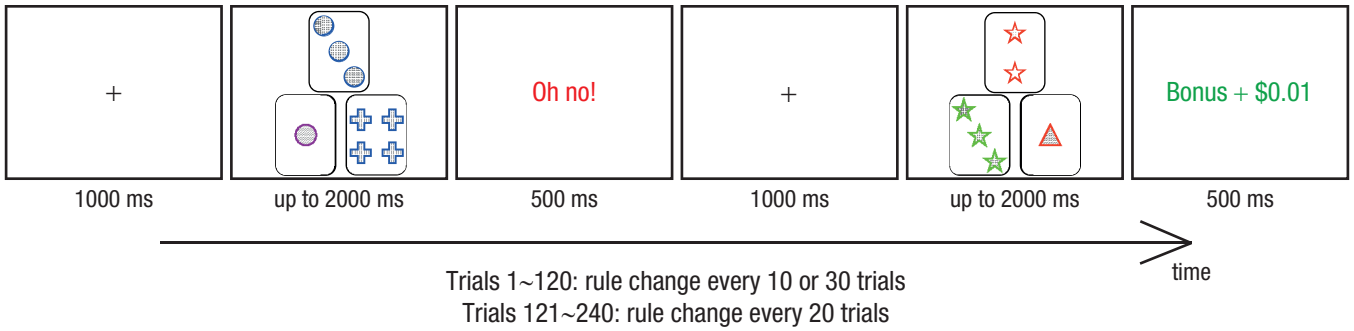

Fig. 1. Illustration of the task paradigm. Each trial began with a fixation period, followed by a display of the reference card (top) and two choice cards (bottom) that required a participant response, followed by feedback. Participants were asked to match the correct choice card with the reference card according to the dimension (i.e., color, shape, filling, or number) that they believed to be the currently relevant matching rule. In the example above, participants had to sort cards according to color or shape. In the first half of the experiment (learning phase), the sorting rule changed every 30 trials for participants in the low-volatility group, and every 10 trials for participants in the high-volatility group. In the second half of the experiment (transfer phase), the sorting rule changed every 20 trials for both groups.

simultaneously presented on the screen, randomly chosen with the following constraints. The card on the top served as the reference card, and the cards at the bottom were choice cards. One of the two choice cards shared the same value as the reference card in one dimension (e.g., shape) but had different values in all the other dimensions (e.g., color, filling, number). The other choice card shared the same value as the reference card in a second dimension (e.g., color) and was different in the three other dimensions (e.g., shape, filling, number). Additionally, there were no shared values on any dimensions between the two choice cards. Only two of the four dimensions, randomly assigned for each participant, were relevant as possible matching rules during the experiment. The two relevant dimensions were explicitly explained to the participant and practiced (see below) prior to the experiment. Only one of the two dimensions was the valid matching rule at any one time, and the valid rule changed over time. It was the participants' goal to figure out, via trial-and-error learning, which matching rule was currently valid on a given trial.

Each trial began with a 1-s fixation period. Then, participants were asked to match the reference card to the correct choice card on the basis of the dimension that they believed to be the currently valid matching rule, using the "z" or "m" button to indicate the left or right choice card, respectively. The cards remained on the screen for up to 2 s or until participants made a response. If participants did not respond in time, they would receive the feedback "Too slow!" and would be asked to press the spacebar to begin the next trial. Otherwise, they would be given feedback of either "Oh no!" or "Bonus + $0.01" for 500 ms. The feedback validity was 80%, which was achieved by switching valid to invalid feedback on 20% of all trials. This means that participants had an 80% chance of receiving positive feedback (and a 20% chance of negative feedback) on correct responses, and vice versa for incorrect responses. The trials with invalid feedback were predetermined pseudorandomly with the constraint of no more than two consecutive instances of invalid feedback. We adopted 80% validity of feedback because it is a typical value for probabilistic reversal learning paradigms (e.g., Behrens et al., 2007; Costa et al., 2015, 2016). Participants were informed of the 80% feedback validity before the experiment. The correct sorting dimension stayed the same for a fixed number of trials before changing to the other dimension, but participants were not explicitly informed about the frequency of rule changes.

Before starting the main experiment, participants were asked to perform a practice task consisting of 40 trials, with the sorting rule changing after 20 trials. The practice task was similar to the main experiment, except that the sorting rule was explicitly displayed on the screen. Participants had to achieve at least 90% accuracy on the practice task to move on to the main experiment. In the main experiment, both the low- and high-volatility groups completed a total of 240 trials. In the first half of the experiment (the learning phase), the sorting rule changed every 30 trials for participants in the low-volatility group and every 10 trials for participants in the high-volatility group. Note that an inherent property of volatility is that recent feedback becomes more informative in more volatile conditions, as feedback on the current trial is more diagnostic of the current rule when the rule changes frequently. In the second half of the experiment (the transfer phase), the sorting rule changed every 20 trials for both groups; task rule volatility (and the informational value of recent feedback) was equal between the two groups during this phase. In Experiment 1, the stimuli and task sets remained the same during the transfer phase as in

the learning phase; in Experiment 2, the stimuli remained the same, but task sets were novel; and in Experiment 3, both stimuli and task sets were novel. There were no explicit instructions informing participants about the currently relevant rule; participants always had to rely on the feedback to figure out the currently relevant sorting rule to maximize their earnings.

Behavioral analyses

For each experiment, we compared the accuracies of the low- and high-volatility groups, and an accurate trial is defined as responding according to the correct sorting rule, regardless of feedback. We then split the data into learning and transfer phases, calculated the accuracies for each phase, and entered participants' mean accuracy values into a Phase (learning vs. transfer) × Volatility (low vs. high) analysis of variance (ANOVA). This was to ensure that any group differences in reinforcement-learning model parameters were not confounded by differences in overall accuracy.

We hypothesized that participants in the low-volatility groups would be slower to switch tasks in response to a rule change than the high-volatility participants. To test this directly, we examined the probability of participants choosing the currently (or previously) active rule on trials before (and after) the periodic rule change point. For both learning and transfer phases, we averaged individuals' choice probability from −5 to +5 trials around the change point. Choice probability was entered into a Phase (learning vs. transfer) × Boundary (before vs. after) × Volatility (low vs. high) ANOVA. Because we hypothesized a priori that the high-volatility group would be quicker to switch after a rule change, we also compared volatility effects separately in each phase and time bin in planned follow-up analyses.

Reinforcement-learning modeling

We fitted reinforcement-learning models (Sutton & Barto, 1998) to the choice behavior to estimate learning rates for the low- and high-volatility groups in the learning and transfer phases of the three experiments, using Markov chain Monte Carlo (MCMC) sampling via the "stan" function from the RStan package (Stan Development Team, 2020) in the R programming environment (R Core Team, 2022). We ran four MCMC chains for 1,000 samples, discarding the first 150 as a warm-up.

Our first model was a standard hierarchical reinforcement-learning (RW-RL) model (Rescorla & Wagner, 1972), fitted to the rule choice behavior. A hierarchical model was used because it takes into account the within-subjects error of each subject's parameter estimate, unlike in the classic approach of comparing the mean value of each parameter for each condition after estimating point estimates for each subject (Daw, 2011). The model consisted first of a *Q*-learning model (Watkins & Dayan, 1992), whereby value estimates for each rule are updated over time on the basis of feedback. Specifically, after an individual i on trial t chooses a matching rule, $C_{i,t} \in \{1, 2\}$ (e.g., color or shape), and feedback is received for that choice, $R_{i,t} \in \{0, 1\}$ (0 if negative feedback and 1 if positive feedback), the value estimate of that rule, $V(C)$, is updated according to the following:

$$V_{i,t+1}(C_{i,t}) = V_{i,t}(C_{i,t}) + \alpha_i \left(R_{i,t} - V_{i,t}(C_{i,t})\right), \quad (1)$$

where α_i is each individual's learning rate. The first trial of each experiment, $V_{i,1}$, as well as the first trial of the transfer phase in Experiments 2 and 3, $V_{i,121}$, were initialized with a separate starting utility, with the prior distribution of $N(0.5, 0.5)$. This was not done for the first trial of the transfer phase of Experiment 1 because in that experiment, there was no change in stimuli or task between the learning and transfer phases.

To estimate the distribution of learning rates across experiments, conditions, and individuals, we estimated a multilevel model with three levels of hierarchy. The top level of the hierarchy described how the average learning rate varied across different conditions, whereas the middle level described how learning rates varied among individuals within a condition. Finally, the bottom level, described by Equations 1, 4, and 7, modeled how individuals learned from feedback over the course of the task and probabilistically generated their choices. The advantage of using a single hierarchical model across all experiments and conditions is to pool information across conditions, resulting in less noisy estimates and reducing overfitting to individual conditions (Gelman, Hill, & Yajima, 2012).

At the top level of the hierarchy, we assumed that the average learning rates for each condition were generated by a mixed-effects general linear model:

$$\begin{aligned} \mu_\eta(p, g, e) &= \phi_\eta + \lambda_\eta(p, g, e) \\ \lambda_\eta(p, g, e) &\sim N(0, \tau_\eta^2) \\ \phi_\eta &\sim N(0, 1) \\ \tau_\eta &\sim N^+(0, 1) \\ \mu_\alpha(p, g, e) &= logit^{-1}\left(\mu_\eta(p, g, e)\right). \end{aligned} \quad (2)$$

Each condition was defined by a combination of a phase p (learning or transfer), a group g (high volatility or low volatility), and an experiment e (Experiment 1, 2, or 3). The hyperparameter ϕ_η is the population average learning rate for all subjects across all conditions. The condition-level random effects $\lambda_\eta(p,g,e)$ determine how far the average of each condition (i.e., phase p for each group g in each experiment e) is from the average of the population mean, with the variance τ^2_η governing the overall variability across conditions.

Next, for the middle level of the hierarchy, we modeled the learning rates of individual participants as arising from a mixed-effects general linear model:

$$\begin{aligned}\eta_{i,p} &= \mu_\eta(p,g,e) + \gamma_{i,p}\\ \gamma_{i,p} &\sim N\left(0,\sigma^2_\eta\right)\\ \sigma_\eta &\sim N^+(0,1)\\ \alpha_{i,p} &= logit^{-1}\left(\eta_{i,p}\right),\end{aligned} \tag{3}$$

where the random effects $\gamma_{i,p}$ determine how far each individual i is from the average for their condition, $\mu_\eta(p,g,e)$, with σ^2_η governing the overall variability across subjects within each condition.

We examined the contrasts between conditions of interest. In particular, for each experiment and phase, we compared whether there were any differences in learning rate between the low- and high-volatility groups. We further examined whether this differed across experiments.

We also sought to examine whether the effects of learning rates were differentially driven by positive feedback (rewarded) versus negative feedback (unrewarded) trials. Intuitively, the negative feedback trials would be expected to drive rule switches because participants would presumably recognize that their currently applied rule was incorrect. We therefore fitted a second model, which we call the two-rates reinforcement-learning (2R-RL) model, in which learning rate was fitted separately for positive (+) and negative (−) reward (r) feedback trials (Donahue & Lee, 2015). The value for each rule, $V\left(C\right)$, is here updated according to the following:

$$V_{i,t+1}\left(C_{i,t}\right) = \begin{cases} V_{i,t}\left(C_{i,t}\right) + \alpha_{+,i}\left(R_{i,t} - V_{i,t}\left(C_{i,t}\right)\right) if\ R_{i,t} = 1\\ V_{i,t}\left(C_{i,t}\right) + \alpha_{-,i}\left(R_{i,t} - V_{i,t}\left(C_{i,t}\right)\right) if\ R_{i,t} = 0, \end{cases} \tag{4}$$

where $\alpha_{+,i}$ is each individual's learning rate for positive feedback trials, and $\alpha_{-,i}$ is each individual's learning rate for negative feedback trials. Similar to the RW-RL model, a hierarchical general linear model was used to estimate the mean effects of each condition:

$$\begin{aligned}\mu_{r,\eta}(p,g,e) &= \phi_{r,\eta} + \lambda_{r,\eta}(p,g,e)\\ \lambda_{r,\eta}\left(p,g,e\right) &\sim N\left(0,\tau^2_{r,\eta}\right)\\ \sigma_{r,\eta} &\sim N^+(0,1)\\ \mu_{r,\alpha}(p,g,e) &= logit^{-1}\left(\mu_{r,\eta}(p,g,e)\right),\end{aligned} \tag{5}$$

where $\phi_{r,\eta}$ are hyperparameters representing the population mean learning rates for the feedback level r for all subjects across the two phases and three experiments. The random effects $\lambda_{r,\eta}(p,g,e)$ represent the deviations of each condition from the population means, and $\tau^2_{r,\eta}$ governs the overall variability in each condition.

Another hierarchical general linear model was used to model individual learning rates:

$$\begin{aligned}\eta_{r,i,p} &= \mu_{r,\eta}(p,g,e) + \gamma_{r,i,p}\\ \gamma_{r,i,p} &\sim N\left(0,\sigma^2_{r,\eta}\right)\\ \sigma^2_{r,\eta} &\sim N^+(0,1)\\ \alpha_{R_{i,t},i,p} &= logit^{-1}\left(\eta_{R_{i,t},i,p}\right),\end{aligned} \tag{6}$$

where $\mu_{r,\eta}(p,g,e)$ are hyperparameters representing the mean learning rates for positive and negative feedback trials for each condition, and $\gamma_{r,i,p}$ is the deviation of each individual from the condition means, with $\sigma^2_{r,\eta}$ variability.

Finally, we also examined how the two volatility groups differed in terms of their action policies. For both the RW-RL and the 2R-RL models, we assumed that subjects chose rules probabilistically on the basis of the value estimates according to a softmax distribution (Daw, 2011). Thus, choice probabilities of selecting each rule (e.g., color or shape) for each trial were computed as follows:

$$p_{i,t}\left(chooseC_{i,t}\right) = \frac{e^{\beta_i V_{i,t}\left(C_{i,t}\right)}}{\sum_{j=1}^{2} e^{\beta_i V_{i,t}\left(C_{i,j}\right)}}. \tag{7}$$

Here, β_i is a hyperparameter known as the inverse temperature, which represents how sensitive choice probabilities are to differences in choice value (Katahira, 2015). β_i values were calculated for each subject similar to η_i, with a hyperparameter representing the population's mean ϕ_β and how many standard deviations, $\lambda_\beta(p,g,e)$, each condition deviated from their group's mean and then another hyperparameter $\mu_\beta(p,g,e)$

representing each condition's mean and the deviations of each individual, $\gamma_{i,p}$, from the condition mean:

$$\mu_\beta(p,g,e) = \phi_\beta + \lambda_\beta(p,g,e)$$
$$\gamma_{\mu_\beta}(p,g,e) \sim N\left(0, \tau_\beta^2\right)$$
$$\tau_\beta \sim N^+(5,5). \quad (8)$$

$$\beta_{i,p} = \mu_\beta(p,g,e) + \gamma_{i,p}$$
$$\gamma_{i,p} \sim N\left(0, \sigma_\beta^2\right)$$
$$\sigma_\beta \sim N^+(0,5). \quad (9)$$

Whereas we report results from both the RW-RL and 2R-RL models, using them to characterize a general overall learning rate ($\alpha_{i,p}$) as well as separate learning rates for positive feedback ($\alpha_{+,i,p}$) and negative feedback ($\alpha_{-,i,p}$) trials, we compared model fits between the two models using the leave-one-out information criterion (Vehtari et al., 2017) and found that the 2R-RL model fit the data better. The expected log pointwise predictive density difference between the two models was −70.1, and its standard error difference was 30.6. This suggests that in the current experiments, participants did indeed have different learning rates for positive and negative feedback trials.

In the following analyses, we compared the posterior distribution of parameter estimates for learning rates and inverse temperature, in the learning phase and transfer phase, across the three experiments. We report $\hat{\delta}$, representing the mean difference between conditions in the model. In analyses with multiple factors, the results are reported in the format of an ANOVA. That is, we report the main effect of a factor by comparing the means of each level of that factor, averaging over all other factors. In the case of interactions, we examined the difference of differences between levels in each factor. We also report credible interval (CrI), which is the Bayesian equivalent of a confidence interval (with a slightly different technical interpretation). All CrIs reported are central 95% intervals of the posterior differences. Additionally, given our a priori expectation that the learning rate in the high-volatility group could only be the same or higher than the low-volatility group in the transfer phase, for tests comparing the two volatility groups, we also report $p(\hat{\delta} < 0)$, which is the proportion of the posterior difference that falls below zero (corresponding to the logic of a one-tailed p value).

Parameter estimates for the RW-RL model are summarized in Table 1, and for the 2R-RL model in Table 2. Our main analyses focused on learning rates, however. We report the results of inverse temperatures in the Supplemental Material available online.

Finally, we performed a parameter recovery analysis that demonstrates that our models can faithfully reproduce parameter estimates from simulated data. We demonstrate that our models identify differences in learning rates between groups when differences exist, and not when differences do not exist, because our results critically depend on trusting in the group learning rate differences determined by our models. The details of the parameter recovery analysis are reported in the Supplemental Material.

Exploring mechanisms of volatility learning

We closely examined choice data across the three experiments to explore possible computational mechanisms behind the learning and transfer of volatility expectations. Firstly, although we fitted behavior using a simple reinforcement-learning model with no explicit representation of the periodic structure of the task, it is possible that participants did learn (at least approximately) that the task consisted of alternating blocks of predictable length and used this to anticipate rule changes. To look for evidence that participants anticipated switch points, we compared the probability of performing the previous rule on the first trial after the rule change with the probability of performing the current rule one to five trials prior to the rule change. Next, we asked whether we could identify temporal dynamics of volatility learning that might shed light on why transfer occurred. As participants learned the volatility over time, we would expect them to change in how quickly their behavior adapts after a rule change; accordingly, we analyzed their mean accuracy in the one to five trials after each rule change, which could be used as a proxy for their learning rate at that point in the task.

To simulate the learning and transfer of environmental volatility across task phases, we simulated an agent that combined the basic reinforcement-learning method used to model behavior with a "meta" learner that tracked volatility via the variance of the prediction errors. To generate choices and track the value of each rule, we had the agent use Equations 7 and 1 respectively, with initial starting utility and parameter values α and β. The optimal rule gave a reward of one with 80% probability, compared with 20% for the suboptimal rule. The optimal rule switched every 10 (high volatility) or 30 (low volatility) trials for the first 120 trials, and then every 20 trials for the last 120 trials. After each

Table 1. Mean Parameter Estimates From the Standard Hierarchical Reinforcement-Learning Model

Parameters	Model fitted parameter						
	ϕ_η	τ_η		ϕ_β	τ_β		
	0.47 (0.09)	0.26 (0.08)		4.17 (0.21)	0.65 (0.19)		
	Transformed parameter						Generated quantity
Experiment and parameter	μ_η	σ^2_η	λ_η	μ_β	σ^2_β	λ_β	μ_α
Experiment 1		0.61 (0.03)			1.42 (0.07)		
Learning phase							
Low volatility	0.23 (0.11)		–0.96 (0.55)	4.73 (0.25)		0.91 (0.52)	0.56 (0.03)
High volatility	0.59 (0.12)		0.47 (0.56)	4.20 (0.25)		0.05 (0.48)	0.64 (0.03)
Transfer phase							
Low volatility	0.16 (0.11)		–1.24 (0.55)	4.73 (0.26)		0.91 (0.53)	0.54 (0.03)
High volatility	0.74 (0.12)		1.06 (0.57)	4.32 (0.26)		0.25 (0.50)	0.68 (0.03)
Experiment 2							
Learning phase							
Low volatility	0.23 (0.12)		–0.97 (0.55)	4.44 (0.24)		0.44 (0.47)	0.56 (0.03)
High volatility	0.71 (0.13)		0.93 (0.57)	3.83 (0.25)		–0.55 (0.49)	0.67 (0.03)
Transfer phase							
Low volatility	0.52 (0.11)		0.19 (0.53)	3.71 (0.23)		–0.74 (0.50)	0.63 (0.03)
High volatility	0.69 (0.11)		0.88 (0.55)	3.96 (0.23)		–0.33 (0.46)	0.67 (0.03)
Experiment 3							
Learning phase							
Low volatility	0.36 (0.11)		–0.45 (0.53)	4.79 (0.26)		1.01 (0.54)	0.59 (0.03)
High volatility	0.54 (0.11)		0.29 (0.52)	4.55 (0.25)		0.63 (0.49)	0.63 (0.03)
Transfer phase							
Low volatility	0.33 (0.11)		–0.58 (0.54)	3.37 (0.24)		–1.29 (0.51)	0.58 (0.03)
High volatility	0.61 (0.12)		0.54 (0.54)	3.26 (0.23)		–1.47 (0.55)	0.65 (0.03)

Note: Values in parentheses are standard deviations.

reward, the volatility υ was updated according to the following equation:

$$\upsilon_{t+1} = \upsilon_t + \frac{\tau}{1+\omega t}\left(\delta^2 - \upsilon\right), \tag{10}$$

where δ is the reward prediction error from Equation 1, and the decay parameters τ and ω were both fixed at 0.1. The initial expected volatility, υ_0, was set to 0.25 to match the variance of a Bernoulli random variable with probability 0.5.

Table 2. Mean Parameter Estimates From the Two-Rates Reinforcement-Learning Model

Parameters	Model fitted parameter					
	$\phi_{\eta+}$	$\tau_{\eta+}$	$\phi_{\eta-}$	$\tau_{\eta-}$	ϕ_{β}	τ_{β}
	0.83 (0.24)	0.69 (0.17)	0.40 (0.11)	0.32 (0.10)	4.39 (0.27)	0.80 (0.21)

Experiment and parameter	Transformed parameter									Generated quantity	
	$\phi_{+,\eta}$	$\sigma^2_{+,\eta}$	$\lambda_{+,\eta}$	$\phi_{-,\eta}$	$\sigma^2_{-,\eta}$	$\lambda_{-,\eta}$	μ_{β}	σ^2_{β}	$\gamma_{\mu_{\beta}}$	$\mu_{+,\alpha}$	$\mu_{-,\alpha}$
Experiment 1		1.46 (0.12)			0.59 (0.05)			1.28 (0.08)			
Learning phase											
Low volatility	0.40 (0.27)		−0.63 (0.47)	0.28 (0.12)		−0.40 (0.46)	4.97 (0.26)		0.77 (0.45)	0.60 (0.06)	0.57 (0.03)
High volatility	0.90 (0.33)		0.10 (0.53)	0.65 (0.14)		0.81 (0.52)	4.26 (0.28)		−0.18 (0.46)	0.71 (0.07)	0.66 (0.03)
Transfer phase											
Low volatility	0.64 (0.26)		−0.28 (0.45)	0.10 (0.12)		−1.01 (0.51)	4.78 (0.25)		0.50 (0.43)	0.65 (0.06)	0.53 (0.03)
High volatility	1.59 (0.32)		1.14 (0.54)	0.60 (0.13)		0.62 (0.49)	4.30 (0.25)		−0.12 (0.43)	0.83 (0.05)	0.64 (0.03)
Experiment 2											
Learning phase											
Low volatility	−0.30 (0.24)		−1.72 (0.56)	0.47 (0.11)		0.23 (0.48)	5.38 (0.29)		1.30 (0.53)	0.43 (0.06)	0.62 (0.03)
High volatility	0.46 (0.30)		−0.56 (0.51)	0.91 (0.15)		1.67 (0.56)	4.15 (0.27)		−0.31 (0.44)	0.61 (0.07)	0.71 (0.03)
Transfer phase											
Low volatility	1.61 (0.26)		1.19 (0.54)	0.11 (0.12)		−0.99 (0.53)	3.80 (0.22)		−0.79 (0.46)	0.83 (0.04)	0.53 (0.03)
High volatility	1.36 (0.29)		0.80 (0.51)	0.56 (0.12)		0.49 (0.48)	4.07 (0.24)		−0.42 (0.41)	0.79 (0.05)	0.64 (0.03)
Experiment 3											
Learning phase											
Low volatility	0.23 (0.27)		−0.90 (0.49)	0.46 (0.12)		0.21 (0.48)	5.31 (0.29)		1.22 (0.52)	0.56 (0.06)	0.61 (0.03)
High volatility	1.07 (0.29)		0.36 (0.46)	0.36 (0.12)		−0.15 (0.48)	4.69 (0.25)		0.41 (0.45)	0.74 (0.06)	0.59 (0.03)
Transfer phase											
Low volatility	1.05 (0.27)		0.34 (0.47)	0.00 (0.13)		−1.35 (0.55)	3.50 (0.23)		−1.19 (0.49)	0.74 (0.05)	0.50 (0.03)
High volatility	1.29 (0.28)		0.69 (0.47)	0.42 (0.13)		0.06 (0.49)	3.32 (0.22)		−1.42 (0.49)	0.78 (0.05)	0.60 (0.03)

Note: Values in parentheses are standard deviations.

Experiment 1

Experiment 1 examined whether prior exposure to low-volatility versus high-volatility rule-switching environments biased people's propensity to infer rule changes in response to negative feedback in subsequent medium-volatility environments. Here, we explored the transfer of learning rates between initial and subsequent environments that differed solely in terms of rule change volatility, with task stimuli and categorization rules held constant.

Method

Participants. Because of a lack of comparable prior studies, we could not base our target sample size on an

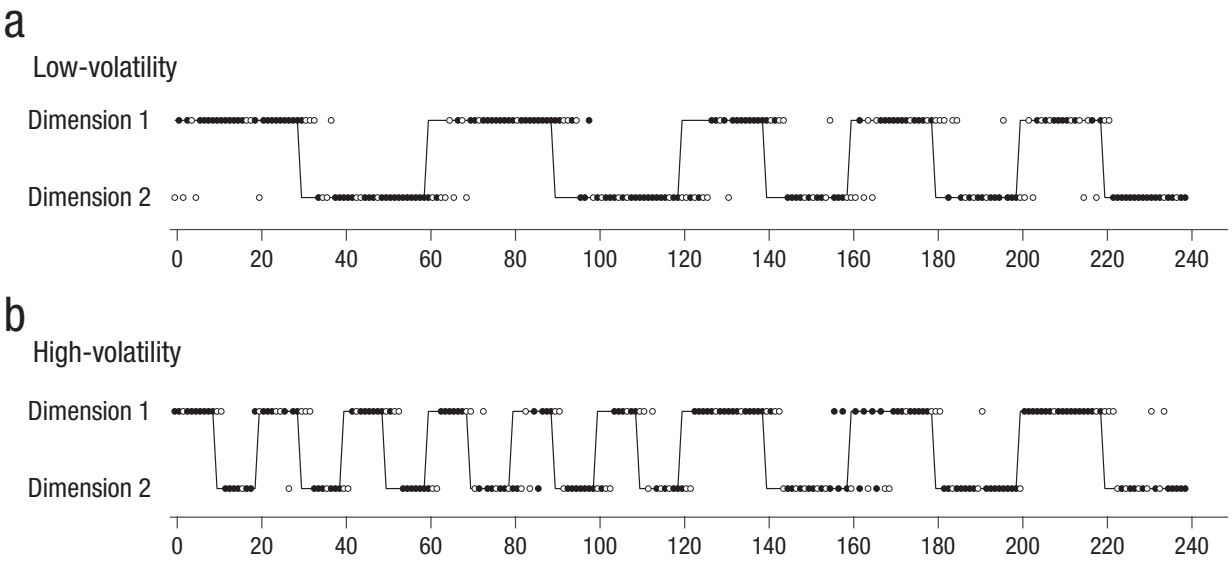

Fig. 2. Dimension rule sequences and a representative participant from (a) the low-volatility group and (b) the high-volatility group. On each trial, participants chose a card on the basis of their belief of the currently valid dimensional matching rule, here called Dimension 1 or 2 (circles). They received positive feedback (filled circles) when sorting according to the correct dimension (black line) 80% of the time and for incorrect choices 20% of the time, and they received negative feedback (open circles) for correct choices 20% of the time and for incorrect choices 80% of the time.

empirical effect size. We therefore opted for a relatively large target sample size (N = ~80). Eighty-eight participants were recruited from Amazon Mechanical Turk (MTurk); each participant was randomly assigned to one of two experimental groups. Participants were compensated at a base rate of $2.50 plus any additional bonuses (M = $1.86, SD = $0.10) earned during the experiment. Thirteen participants were excluded from the analysis because of overall accuracy lower than 65%, leaving a final sample size of 75. The low-volatility group had 39 participants (22 male, 15 female, two did not wish to reply; age: range = 26–56 years, M = 36.69 years, SD = 8.81), and the high-volatility group had 36 participants (24 male, 11 female, one did not wish to reply; age: range = 22–60 years, M = 39.44 years, SD = 10.40).

Stimuli. Task stimuli consisted of 256 unique "cards" with one to four display items consisting of a specific shape (circle, triangle, plus, or star) in a particular color (blue, green, red, or purple) and a particular filling (checkered, dots, wave, or grid). We refer to these card properties as "dimensions" (number, shape, color, filling) that can take particular "values" (1, 2, 3, or 4 for the number dimension). Each trial involved a display of three such cards. See Figure 1 for sample stimuli.

Procedure. In the first half of the experiment, participants had to sort cards according to two of the four dimensions (e.g., color and shape; randomly assigned across participants). Sorting rules alternated every 30 trials for the low-volatility group and every 10 trials for the high-volatility group. In the second half (transfer phase) of the experiment, sorting rules alternated every 20 trials in the transfer phase. There was no explicit separation between the first half and second half of the experiment.

Results

Behavior. Figure 2 illustrates the rule sequence and choice data from a representative participant from each group. In spite of the 80%-validity probabilistic feedback, participants were able to track the correct rule most of the time (low-volatility group: M = 79.28%, SD = 5.76%; high-volatility group: M = 73.19%, SD = 4.13%). A Phase (learning vs. transfer) × Volatility (low vs. high) ANOVA showed a significant main effect of phase, $F(1, 73) = 39.82$, $p < .001$, and group, $F(1, 73) = 27.27$, $p < .001$, as well as a Phase × Volatility interaction, $F(1, 73) = 99.74$, $p < .001$. The main effect of phase was driven by participants having higher accuracy for the transfer phase compared with the learning phase ($t = 3.56$, $p < .001$), presumably because of a practice effect. The main effect of volatility was driven by the low-volatility group having overall higher accuracy than the high-volatility group ($t = 5.22$, $p < .001$), and this difference was significant in the learning phase ($t = 11.03$, $p < .001$) but not in the transfer phase ($t = -0.99$, $p = .33$). This effect was expected, given the greater number of (error-inducing) rule reversals in the high-volatility group's learning phase.

Figure 3 illustrates participants' choice probability as a function of time. We examined whether participants were more likely to switch task rules after the change

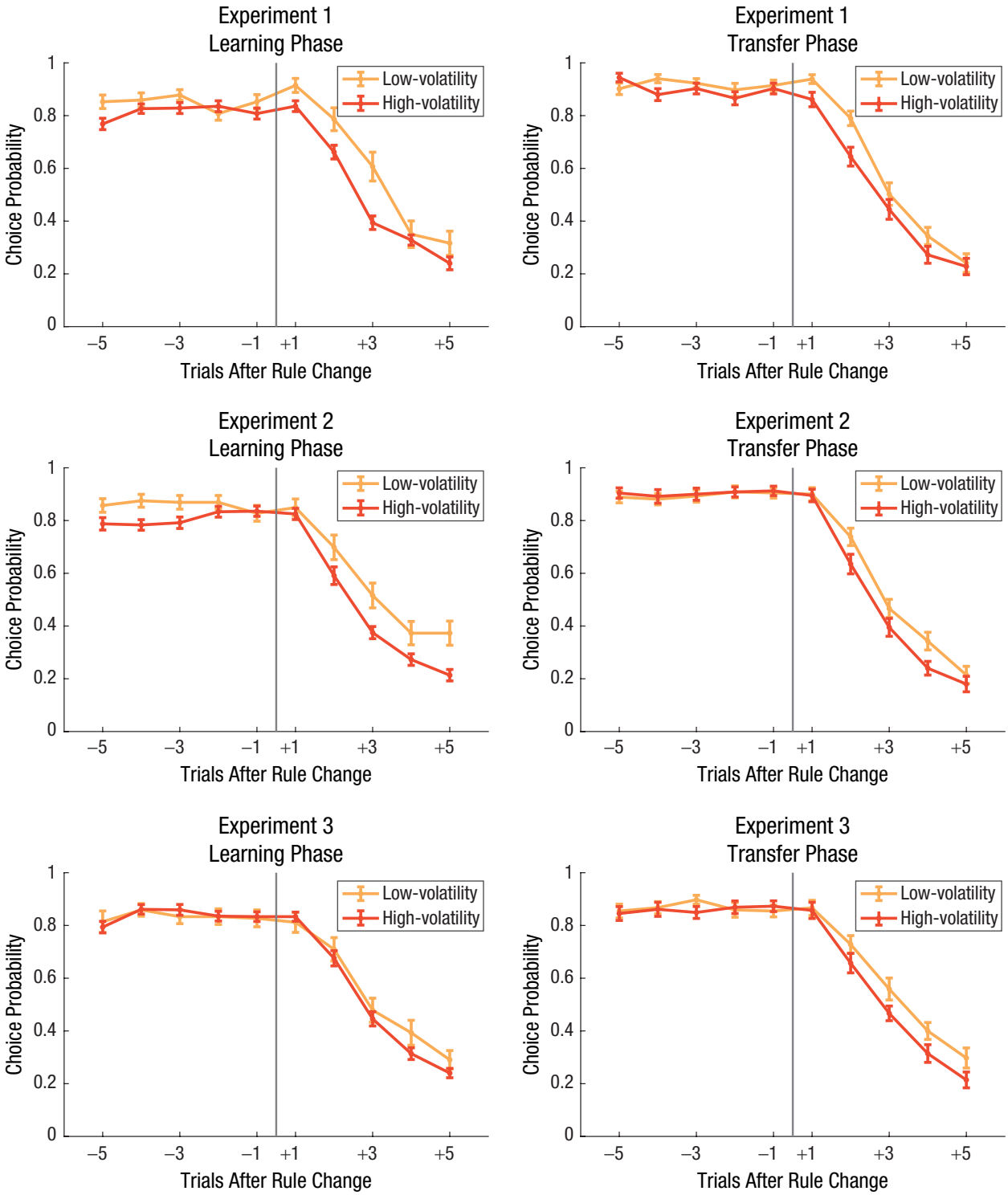

Fig. 3. Choice probability (participants' tendency to choose the rule that was active prior to the rule change) as a function of trial before and after the rule change point for the learning and transfer phases of each experiment. Error bars represent standard errors.

point. A Phase (learning vs. transfer) × Boundary (before vs. after) × Volatility (low vs. high) ANOVA showed a significant main effect of phase, $F(1, 73) = 6.76$, $p = .01$, a main effect of boundary, $F(1, 73) = 600.71$, $p < .001$, and a main effect of volatility, $F(1, 73) = 12.57$, $p < .001$. There was a Phase × Boundary interaction, $F(1, 73) = 20.77$, $p < .001$, and a Boundary × Volatility interaction, $F(1, 73) = 5.09$, $p = .03$. No other interaction was significant. We next conducted separate Boundary (before vs. after) × Volatility (low vs. high) ANOVAs on the learning and transfer phases. In the learning phase, we found a main effect of boundary, $F(1, 73) = 225.21$, $p < .001$, driven by a higher probability of performing the current task rule prior to the rule change. There was also a main effect of volatility, $F(1, 73) = 10.73$, $p < .01$, driven by the low-volatility group being overall more likely to perform the initial task rule. There was no Boundary × Volatility interaction, $F(1, 73) = 2.98$, $p = .09$. *t* tests between volatility groups showed that prior to the rule change, the low- and high-volatility groups were equally likely to perform the current task rule ($t = 1.81$, $p = .07$). However, the low-volatility group was more likely to persist with the previous rule after the change point ($t = 2.93$, $p < .01$). In the transfer phase, we found a main effect of boundary, $F(1, 73) = 693.90$, $p < .001$, driven by a higher probability of performing the current task rule prior to the rule change. There was also a main effect of volatility, $F(1, 73) = 6.21$, $p = .02$, which was driven by the low-volatility group being overall more likely to perform the initial task rule. There was a marginally significant Boundary × Volatility interaction, $F(1, 73) = 3.86$, $p = .05$. *t* tests between volatility groups showed that prior to the rule change, the low- and high-volatility groups were equally likely to perform the current task rule ($t = 0.92$, $p = .36$). However, the low-volatility group was more likely to persist with the previous rule after the change point ($t = 2.67$, $p < .01$).

Reinforcement-learning models. Learning rate parameter estimates are visualized in Figure 4. As a confirmatory analysis, we first tested whether learning rates in the high-volatility group were higher than in the low-volatility group during the learning phase, where the matching rule switched every 10 trials compared with 30 trials. As expected, the RW-RL model showed that learning rates for participants in the high-volatility group were significantly larger than for those in the low-volatility group, $\hat{\delta} = 0.08$, 95% CrI = [0.01, 0.16], $p(\hat{\delta} < 0) = .01$. Similarly, in the 2R-RL model, we also found a main effect of volatility; the high-volatility group exhibited higher learning rates than the low-volatility group, $\hat{\delta} = 0.10$, 95% CrI = [0.004, 0.19], $p(\hat{\delta} < 0) = .02$. There was no main effect of feedback ($\hat{\delta} = -0.04$, 95% CrI = [−0.14, 0.07]) and no Feedback × Volatility interaction ($\hat{\delta} = -0.01$, 95% CrI = [−0.11, 0.09]).

Our main interest centered on the learning rates during the transfer phase. We found that according to the RW-RL model, the high-volatility group continued to show a higher learning rate than the low-volatility group during this phase, $\hat{\delta} = 0.14$, 95% CrI = [0.06, 0.21], $p(\hat{\delta} < 0) < .001$. The 2R-RL model showed a main effect of volatility, $\hat{\delta} = 0.15$, 95% CrI = [0.07, 0.23], $p(\hat{\delta} < 0) < .001$, driven by higher learning rates for the high-volatility group; there was also a main effect of feedback ($\hat{\delta} = -0.15$, 95% CrI = [−0.24, −0.07]), driven by the learning rates being higher for positive feedback trials, and no Feedback × Volatility interaction ($\hat{\delta} = -0.03$, 95% CrI = [−0.11, 0.05]). These results document that volatility-driven learning rates acquired during the first half of the task generalized to the second half transfer phase, where volatility was equated between groups.

Our results showed that participants adapted their rule-switching strategies to the volatility of the task environment, reflected in faster rule switches around task boundaries as well as in a higher learning rate in participants in the high-volatility compared with the low-volatility environment. Importantly, pre-exposure to high-volatility compared with low-volatility environments led to a higher learning rate in a subsequent medium-volatility environment. In other words, learned expectations about the level of cognitive flexibility required in the environment endured over time.

Experiment 2

Experiment 2 tested whether learning rates would transfer when the sets of rules changed between the learning and transfer phases (while the stimuli remained the same). Specifically, we probed whether exposure to low- or high-volatility learning environments involving two of four possible task rules (e.g., shape and color matching) would bias the learning rate in subsequent medium-volatility environments with the other two possible task rules (i.e., number and filling matching). Obtaining transfer under these conditions would indicate that the expectations that are being transferred are independent of the specific task rules, thus representing a form of meta-learning.

Method

Participants. Ninety-four participants were recruited from MTurk, and each participant was randomly assigned to one of the two volatility groups. Participants were compensated at a base rate of $2.50 plus any additional bonuses ($M = \$1.86$, $SD = \$0.10$) earned during the experiment. Twelve participants were excluded from the analysis because of overall accuracy lower than 65%, leaving a final sample size of 82. The low-volatility group

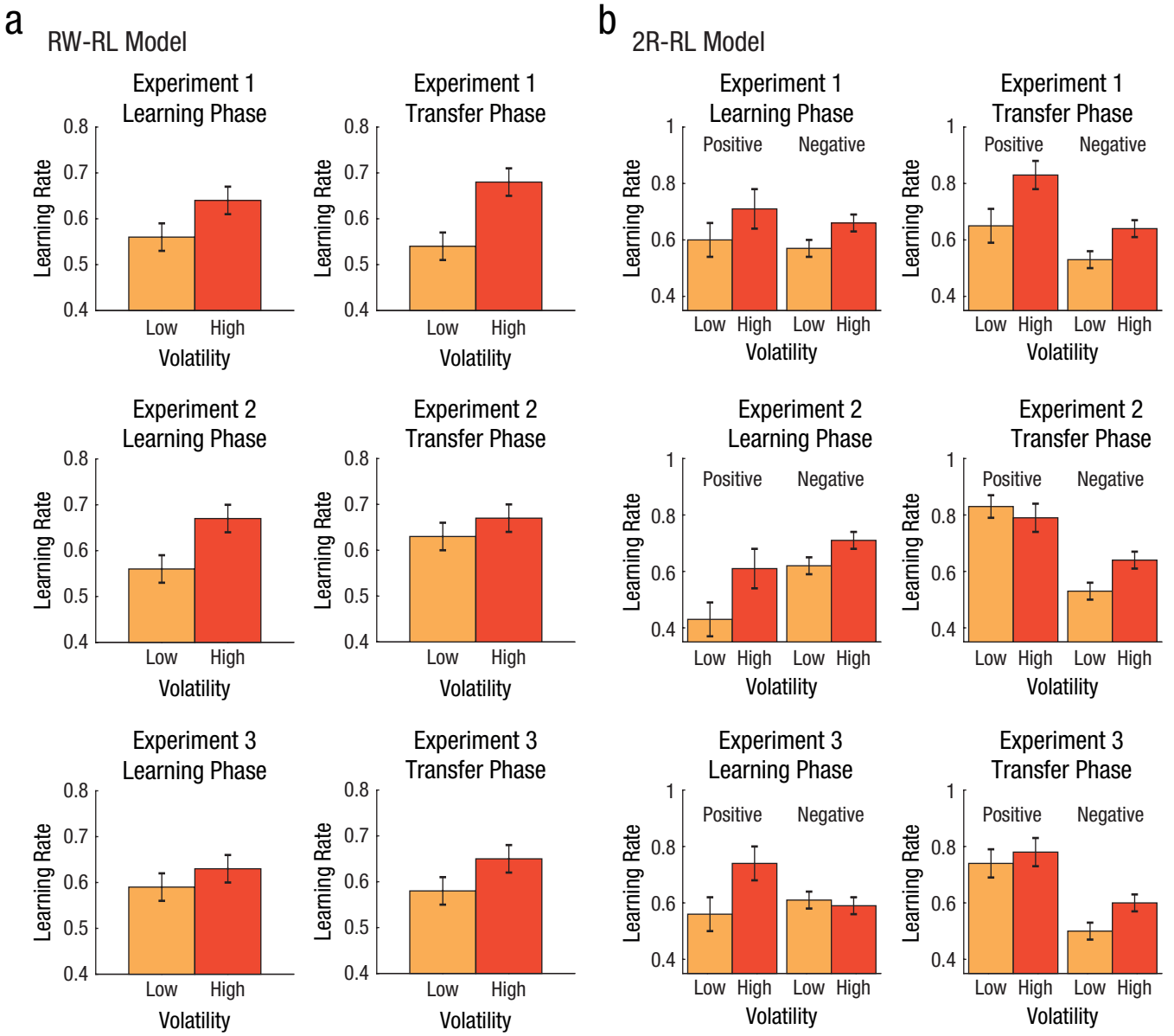

Fig. 4. Group level (high volatility vs. low volatility) learning rate parameter estimates from (a) the standard hierarchical reinforcement-learning (RW-RL) model and (b) the two-rates reinforcement-learning (2R-RL) model for the learning and transfer phases in all three experiments. The error bars reflect the fitted group-level standard deviation estimates.

had 42 participants (20 male, 22 female; age: range = 23–70 years, M = 39.64 years, SD = 11.80), and the high-volatility group had 40 participants (22 male, 18 female; age: range = 24–76 years, M = 38.80 years, SD = 11.36).

Stimuli. The stimuli were the same as in Experiment 1.

Procedure. As in Experiment 1, in the first half of the experiment, participants had to sort cards according to two of the four dimensions (e.g., color and shape; randomly assigned across participants). However, unlike Experiment 1, before the start of the second half (transfer phase) of the experiment, participants were taken to another instruction screen and informed that they would now be sorting cards according to the other two dimensions that were previously irrelevant in the first half (e.g., filling and number). There was no practice for the transfer phase, and it started as soon as participants indicated that they were ready.

Results

Behavior. Participants were able to perform the task reasonably well (low-volatility group: M = 79.22%, SD = 5.66%;

high-volatility group: $M = 73.58\%$, $SD = 4.45\%$). The Phase (learning vs. transfer) × Volatility (low vs. high) ANOVA showed a significant main effect of phase, $F(1, 80) = 22.80$, $p < .001$, a main effect of volatility, $F(1, 80) = 24.89$, $p < .001$, and a Phase × Volatility interaction, $F(1, 80) = 40.93$, $p < .001$. As in Experiment 1, the main effect of phase was driven by participants having higher accuracy for the transfer phase compared with the learning phase ($t = 4.99$, $p < .001$), presumably because of generic task practice effects. The main effect of volatility was driven by the low-volatility group having higher accuracy than the high-volatility group ($t = 3.71$, $p < .001$). The interaction was again driven by the low-volatility group having higher accuracy compared with the high-volatility group in the learning phase ($t = 8.39$, $p < .001$) but not in the transfer phase ($t = -0.54$, $p = .59$).

We next examined participants' choice probabilities before and after the rule change point. The Phase (learning vs. transfer) × Boundary (before vs. after) × Volatility (low vs. high) ANOVA showed a significant effect of phase, $F(1, 80) = 5.67$, $p = .02$, a main effect of boundary, $F(1, 80) = 837.97$, $p < .001$, and a main effect of volatility, $F(1, 80) = 11.52$, $p = .001$. There was a Phase × Boundary interaction, $F(1, 80) = 10.59$, $p < .01$, a Phase × Volatility interaction, $F(1, 80) = 4.55$, $p = .04$, and a Boundary × Volatility interaction, $F(1, 80) = 6.23$, $p = .01$. The Phase × Boundary × Volatility interaction was not significant, $F(1, 80) = 0.16$, $p = .69$. We next conducted separate Boundary (before vs. after) × Volatility (low vs. high) ANOVAs on the learning and transfer phases. In the learning phase, we found a main effect of boundary, $F(1, 80) = 287.38$, $p < .001$, driven by a higher probability of performing the current task rule prior to the rule change. There was a main effect of volatility, $F(1, 80) = 12.45$, $p < .001$, driven by the low-volatility group being overall more likely to perform the initial task rule. There was no Boundary × Volatility interaction, $F(1, 80) = 1.93$, $p = .17$. t tests between volatility groups showed that prior to the rule change, the low-volatility group was more likely to perform the current task rule ($t = 2.37$, $p = .02$). Additionally, the low-volatility group was more likely to persist with the previous rule after the change point ($t = 3.01$, $p < .01$). In the transfer phase, results showed a main effect of boundary, $F(1, 80) = 766.71$, $p < .001$, driven by a higher probability of performing the current task rule prior to the rule change. There was no main effect of volatility, $F(1, 80) = 2.65$, $p = .11$. There was a significant Boundary × Volatility interaction, $F(1, 80) = 6.16$, $p = .02$. t tests between volatility groups showed that prior to the rule change, the low- and high-volatility groups were equally likely to perform the current task rule ($t = -0.38$, $p = .71$). However, the low-volatility group was more likely to persist with the previous rule after the change point ($t = 2.75$, $p < .01$).

Reinforcement-learning models. In the learning phase, the high-volatility group had a higher learning rate than the low-volatility group, as estimated by the RW-RL model, $\hat{\delta} = 0.11$, 95% CrI = [0.03, 0.19], $p(\hat{\delta} < 0) < .01$. Learning rates from the 2R-RL model also showed a main effect of volatility, $\hat{\delta} = 0.14$, 95% CrI = [0.05, 0.23], $p(\hat{\delta} < 0) = .001$, which was again driven by the high-volatility group having a higher learning rate. We additionally found a main effect of feedback ($\hat{\delta} = 0.15$, 95% CrI = [0.04, 0.25]) because of higher learning rates for the negative feedback trials. There was no Feedback × Volatility interaction ($\hat{\delta} = -0.04$, 95% CrI = [−0.14, 0.05]).

In the transfer phase, the RW-RL model showed no significant difference of learning rates between groups, $\hat{\delta} = 0.04$, 95% CrI = [−0.03, 0.11], $p(\hat{\delta} < 0) = .13$, although the high-volatility group showed a numerically higher learning rate. The 2R-RL model showed no main effect of volatility, $\hat{\delta} = 0.04$, 95% CrI = [−0.04, 0.10], $p(\hat{\delta} < 0) = .15$, but there was a main effect of feedback ($\hat{\delta} = -0.23$, 95% CrI = [−0.30, −0.15]), which was driven by higher learning rates for positive feedback trials. Critically, results also showed a significant Feedback × Volatility interaction ($\hat{\delta} = 0.07$, 95% CrI = [0.01, 0.14]). The interaction was driven by the high-volatility group showing a higher learning rate compared with the low-volatility group for negative feedback trials, $\hat{\delta} = 0.11$, 95% CrI = [0.03, 0.19], $p(\hat{\delta} < 0) < .01$, but not for positive feedback trials, $\hat{\delta} = -0.04$, 95% CrI = [−0.15, 0.07], $p(\hat{\delta} < 0) = .74$. Thus, learning rates acquired during the learning phase generalized to rule-switching performance in the transfer phase despite a change in specific task rules between phases, but only for negative feedback.

Even though in Experiment 2, the task rules that participants were switching between changed from the learning phase to the transfer phase, we observed evidence for transfer of rule-learning rates. These results suggest that participants do not transfer a specific association between particular task rules and change point estimates in the present paradigm but, rather, that they form and transfer a more abstract expectation of the volatility of the rules governing the environment, as reflected in the learning rate. This transfer was expressed primarily in response to negative feedback rather than to positive feedback, in line with the assumption that negative feedback trials in particular cause participants to switch rules.

Experiment 3

In Experiment 3, we provided a test of "far transfer" by investigating whether prior experiences of low- or high-volatility environments would bias the tendency to shift sets in subsequent medium-volatility environments with both novel rules and novel stimuli.

Method

Participants. One hundred one participants were recruited from MTurk, and each participant was randomly assigned to one of two volatility groups. Participants were compensated at a base rate of $2.50 plus any additional bonuses (M = $1.82, SD = $0.09) earned during the experiment. Twenty participants were excluded from the analysis because of overall accuracy lower than 65%, leaving a final sample size of 81. The low-volatility group had 39 participants (26 male, 12 female, one did not wish to reply; age: range = 27–67 years, M = 41.41 years, SD = 10.94), and the high-volatility group had 42 participants (23 male, 18 female, one did not wish to reply; age: range = 21–56 years, M = 36.29 years, SD = 7.88).

Stimuli. We used the same stimuli as in Experiments 1 and 2 for the learning phase of Experiment 3. To test whether the rule learning rates could transfer to other tasks with novel stimuli, we used face stimuli taken from the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015) for the Experiment 3 transfer phase. A total of 64 emotion-neutral faces (16 Asian male, 16 Asian female, 16 Caucasian male, and 16 Caucasian female) were used.

Procedure. In the first half of the experiment (the learning phase), participants had to sort cards according to two of the four dimensions (e.g., color and shape; randomly assigned across participants). Before the start of the second half (transfer phase) of the experiment, participants were taken to another instruction screen and informed that they would now be sorting face images according to either gender (male vs. female) or race (Asian vs. Caucasian). As in the card-matching task, on each trial three faces were displayed arranged in a pyramid, with the face on the top serving as the reference face, and the faces at the bottom as choice faces. Each of the two choice faces shared only one matching domain (gender or race) with the reference face. There was no practice for the transfer phase, and it started as soon as participants indicated that they were ready.

Results

Behavior. Both groups were able to perform the task reasonably well (low-volatility group: M = 74.75%, SD = 5.66%; high-volatility group: M = 71.96%, SD = 4.81%). The Phase (learning vs. transfer) × Volatility (low vs. high) ANOVA showed a significant main effect of phase, $F(1, 79) = 11.38$, $p = .001$, a main effect of volatility, $F(1, 79) = 5.74$, $p = .02$, and a Phase × Volatility interaction, $F(1, 79) = 16.28$, $p < .001$. As in the prior two experiments, the main effect of phase was driven by higher accuracy for the transfer phase ($t = 3.36$, $p < .001$). In line with previous results, mean accuracy was higher for the low-volatility group in the learning phase ($t = 4.45$, $p < .001$) but not in the transfer phase ($t = -1.24$, $p = .22$).

Examination of participants' rule choice probability before and after the rule change using a Phase (learning vs. transfer) × Boundary (before vs. after) × Volatility (low vs. high) ANOVA showed a significant effect of boundary, $F(1, 79) = 685.66$, $p < .001$. There were no main effects of phase, $F(1, 79) = 3.56$, $p = .06$, or volatility, $F(1, 79) = 2.58$, $p = .11$. There was a Boundary × Volatility interaction, $F(1, 79) = 4.10$, $p < .05$. None of the other interactions were significant—all Fs$(1, 79) < 0.84$; all ps $> .36$. We conducted separate Boundary (before vs. after) × Volatility (low vs. high) ANOVAs on the learning and transfer phases. In the learning phase, we found a main effect of boundary, $F(1, 79) = 350.51$, $p < .001$, but no main effect of group, $F(1, 79) = 0.59$, $p = .45$, and no Boundary × Volatility interaction, $F(1, 79) = 1.28$, $p = .26$. Direct comparisons between volatility groups showed no difference in before ($t = -0.14$, $p = .89$) or after ($t = 1.18$, $p = .24$) the rule change point. In the transfer phase, we found a main effect of boundary, $F(1, 79) = 350.15$, $p < .001$, but no main effect of group, $F(1, 79) = 3.45$, $p = .07$, or Boundary × Volatility interaction, $F(1, 79) = 3.07$, $p = .08$. t tests between volatility groups showed that there were no differences between the low- and high-volatility groups ($t = 0.31$, $p = .75$) before the change point; however, the low-volatility group was significantly more likely to perform the previous task rule after the change point ($t = 2.27$, $p = .03$).

Reinforcement-learning models. In the learning phase, the RW-RL model found no significant difference between learning rates in the two volatility groups, $\hat{\delta} = 0.04$, 95% CrI = [–0.03, 0.12], $p(\hat{\delta} < 0) = .12$, although the high-volatility group had a numerically higher learning rate compared with the low-volatility group. The 2R-RL model showed that the high-volatility group had higher learning rates than the low-volatility group, $\hat{\delta} = 0.08$, 95% CrI = [–0.01, 0.17], $p(\hat{\delta} < 0) = .03$. There was no main effect of feedback ($\hat{\delta} = -0.05$, 95% CrI = [–0.15, 0.05]). However, there was a significant Feedback × Volatility interaction ($\hat{\delta} = -0.10$, 95% CrI = [–0.19, –0.01]). Post hoc analysis suggests that the interaction was driven by the high-volatility group showing higher learning rates than the low-volatility group for positive feedback, $\hat{\delta} = 0.18$, 95% CrI = [0.02, 0.34], $p(\hat{\delta} < 0) = .01$, but not negative feedback, $\hat{\delta} = -0.02$, 95% CrI = [–0.10, 0.05], $p(\hat{\delta} < 0) = .74$.

In the transfer phase, the RW-RL model had higher learning rates in the high-volatility group compared with the low-volatility group, $\hat{\delta} = 0.07$, 95% CrI = [–0.01, 0.15], $p(\hat{\delta} < 0) = .04$. Similarly, in the 2R-RL model, the

high-volatility group had higher learning rates than the low-volatility group, $\hat{\delta} = 0.07$, 95% CrI = [−0.003, 0.15], $p(\hat{\delta} < 0) = .03$. We found a main effect for feedback ($\hat{\delta} = -0.21$, 95% CrI = [−0.29, −0.12]), driven by higher learning rates during positive feedback trials. We found no Feedback × Volatility interaction ($\hat{\delta} = 0.03$, 95% CrI = [−0.05, 0.11]). Although this interaction was not significant, on the basis of our previous findings we further examined the volatility effect in positive and negative feedback separately and found that the volatility effect was significant in the negative feedback trials, $\hat{\delta} = 0.10$, 95% CrI = [0.02, 0.19], $p(\hat{\delta} < 0) < .01$, but not positive feedback trials, $\hat{\delta} = 0.04$, 95% CrI = [−0.09, 0.18], $p(\hat{\delta} < 0) = .25$. Thus, learning rates acquired during the learning phase generalized to rule-switching performance in the transfer phase, in spite of a change in both the stimulus materials and task rules between phases.

In sum, we observed rule- and stimulus-independent "far transfer" of rule-learning rates, in particular for negative feedback trials. In a final analysis, we sought to directly compare the degree of learning rate transfer between experiments, testing whether transfer differed quantitatively as a function of whether rules and stimuli remained the same (Experiment 1), whether rules changed (Experiment 2), or whether both rules and stimuli changed between the learning and transfer phases.

Extended Analysis Across Experiments

Cross-experiment transfer effect comparison

We compared the transfer phase learning rates across the three experiments. Results from the RW-RL model showed a main effect of volatility, $\hat{\delta} = 0.08$, 95% CrI = [0.03, 0.12], $p(\hat{\delta} < 0) < .001$; the high-volatility groups had higher learning rates than the low-volatility groups. There were no pairwise differences in overall learning rates across the three experiments (max $|\hat{\delta}| = 0.04$, 95% CrI = [−0.09, 0.01]). There were no interactions between volatility and experiments (max $\hat{\delta} = 0.05$, 95% CrI = [−0.001, 0.10]).

With learning rates from the 2R-RL model, we found a main effect of volatility, $\hat{\delta} = 0.09$, 95% CrI = [0.04, 0.13], $p(\hat{\delta} < 0) < .001$; the high-volatility groups had higher learning rates than the low-volatility groups. There was also a main effect of feedback ($\hat{\delta} = 0.20$, 95% CrI = [0.14, 0.25]), driven by higher learning rates during negative feedback trials. There were no differences in mean learning rates between pairwise comparisons across experiments (max $\hat{\delta} = 0.4$, 95% CrI = [−0.01, 0.09]). There was no Volatility × Feedback × Experiment interaction (max $\hat{\delta} = 0.05$, 95% CrI = [−0.0004, 0.10]).

These results echo the previous experiments in providing evidence for a transfer of task rule-learning rates, even though in Experiment 3, this involved applying new rules to new stimuli (far transfer). Our results again also suggest that this transfer occurs primarily for negative rather than positive feedback trials. A comparison of the three experiments showed that the degree of this transfer did not differ between experiments and was therefore unaffected by the similarity between learning and transfer tasks.

Mechanisms of volatility learning

To probe whether participants explicitly anticipated switch points (e.g., by counting trials), we compared the probability of performing the previous rule at the first trial after the rule change with the probability of performing the current rule one to five trials prior to the rule change. As illustrated in Figure 3, in all experiments and conditions, the probability of picking the preswitch rule remains stable near 0.8 through the first trial after the rule switch. Across all experiments and groups, none of the first trials after the rule change showed a decreased probability of choosing the preswitch rule compared with the one to five trials prior to the rule change (max $t = 0.60$, $p = .55$). Therefore, only after participants have received feedback did rule choice probabilities substantially change. This suggests that participants do not anticipate the task switches ahead of time or count the number of trials. Rather, they likely used a reactive strategy that relies on negative feedback. Note that whereas participants in the high-volatility condition weighted recent feedback comparatively more heavily, participants in both conditions appear to have used a reactive, feedback-driven strategy, which is consistent with the simple model-free reinforcement-learning mechanism we have used to describe behavior. Second, to gain insight into the mechanism of transfer between the learning and transfer phases, we examined the time course of volatility learning over the course of the task. The mean accuracy in the one to five trials after each rule change as a function of rule change experience is plotted in Figure 5a. The time course of learning shows that behavior was adjusted only in the beginning stages of the learning phase but remained stable when environmental volatility changed at the onset of the transfer phase. Specifically, in the high-volatility group, postchange accuracy increased over the first three rule changes (third change vs. first change; $t = 2.63$, $p < .01$) but then remained stable over the rest of the task, including after the beginning of the transfer phase. The low-volatility group, on the other hand, began with initial postchange accuracy similar to the high-volatility group ($t = 0.51$,

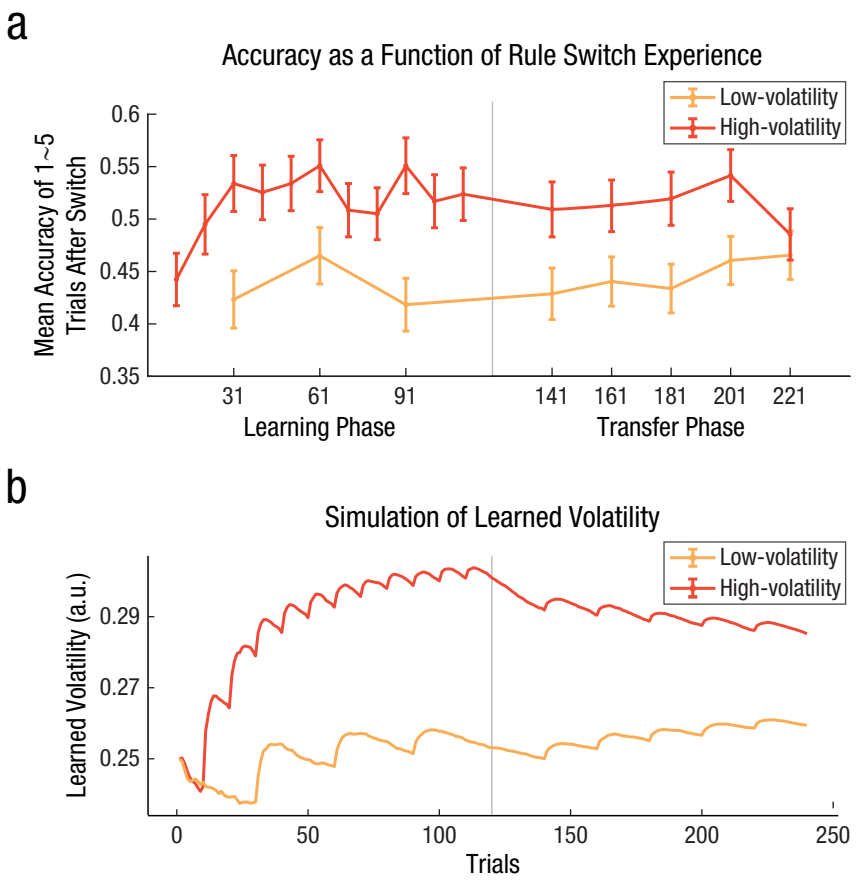

Fig. 5. Top panel (a) shows accuracy as a function of rule change experience. Each dot represents a rule change point. Bottom panel (b) shows the learned volatility of simulated agents with decaying volatility learning rates. Lines represent the average of 500 simulations.

p = .61) but never significantly increased in accuracy with subsequent rule changes (max t = 1.41, p > .15). This indicates that an unexpected volatility level is quickly learned at the beginning of the task, yet when the volatility level changes later in the transfer phase, learning happens slowly if at all. Thus, one possibility is that transfer of volatility expectations occurs because of the failure to update volatility expectations when volatility changes.

On the basis of the above two observations, our data are consistent with a learning model that learns volatility by tracking prediction error magnitudes but updates its belief about environmental volatility with a "meta" learning rate that decays over time. To demonstrate the plausibility of this mechanism for the transfer effects observed in our task, we simulated a learning agent with a decaying volatility learning rate performing the same task as the participants. The agent's learned volatility over the course of the task (Fig. 5b) mirrors the learning trajectories of the participants (Fig. 5a), with the learned volatility rising rapidly in the high-volatility condition and then dropping slowly in the transfer phase, while remaining stable throughout the low-volatility condition (for details on the simulation, see the Method section in the introduction). Decaying learning rates are a standard algorithmic policy in modern machine learning, allowing agents to more readily find global minima in value landscapes (Duchi et al., 2011; Jacobs, 1988). In humans, it has been shown that learning rates decrease with the scale of prediction errors (Pearce & Hall, 1980); learning stabilizes over time when the environment is stable. Thus, people might reduce their learning rate of volatility learning under the expectation that volatility itself will remain stable, resulting in an inability to adjust to changes in volatility when they occur.

General Discussion

The current study examined whether participants acquire and transfer expectations about cognitive flexibility demands across different contexts. We tested whether learning to change task sets more or less frequently in one context would affect learning rates in subsequent contexts. We replicated previous findings

showing that participants adjusted their learning rates according to environment volatility (Behrens et al., 2007; Massi et al., 2018), with high-volatility environments leading to higher learning rates, and extended that finding from stimulus-specific associations to stimulus-independent rules. Crucially, we further found that the inductive biases acquired during the learning phase affected learning rates in a subsequent transfer phase (Experiment 1) and generalized to novel task rules (Experiment 2) and novel rules and stimuli (Experiment 3). This was reflected by an overall higher learning rate in the transfer phase for participants previously exposed to a high-volatility environment compared with those previously exposed to a low-volatility environment, which was mainly driven by learning from negative feedback (unrewarded) trials. Taken together, this demonstrates that people form and transfer an abstract, stimulus- and task-independent expectation of the volatility of the rules governing their environment, expressed in a more or less cognitively flexible rule-updating strategy. To the best of our knowledge, this is the first demonstration of people's ability to extract and reuse cognitive control learning parameters that transcend specific stimuli and tasks.

Previous behavioral studies have shown that participants can strategically adapt their readiness to switch tasks in line with changes in contextual switch likelihood (reviewed by Braem & Egner, 2018). For instance, when the frequency of cued task switches is manipulated over different blocks of trials, participants exhibit smaller switch costs in blocks where switches are frequent compared with when they are rare (e.g., Chiu & Egner, 2017; Dreisbach & Haider, 2006; Leboe et al., 2008; Monsell & Mizon, 2006; Siqi-Liu & Egner, 2020). However, unlike in the present experiments, this change in switch readiness seems to be limited to "biased" task sets that are associated with more frequent switch or repeat trials and does not generalize to intermingled "unbiased" task sets where the likelihood of switching versus repeating a task is equal (Siqi-Liu & Egner, 2020). This suggests that meta-flexibility in cued task switching is task-set or stimulus specific, rather than being due to participants developing a more global flexible cognitive strategy or processing mode that would promote switching in general (i.e., to *any* other task). This lack of transfer to other tasks also fits with cognitive training studies that demonstrated that switch costs decreased over training when the same tasks were used, but substantial costs reemerged when participants were given new tasks to switch between (Sabah et al., 2019, 2021). A possible explanation underlying this dearth of transfer in prior studies could be that frequent forced (cued) switching motivates participants to keep the multiple relevant task sets in working memory, which would ease the switching between the respective tasks but may not necessarily transfer to new tasks (Dreisbach & Fröber, 2019; Fröber et al., 2021).

How was transfer of switch readiness achieved in the present study? We posit that this is likely because of reliance on a different mechanism for modulating meta-flexibility. One proposal is that one's set point on the cognitive stability-flexibility continuum can be conceptualized in terms of an "updating threshold"—the ease with which new task rule information is allowed to enter working memory (Dreisbach & Fröber, 2019; Goschke, 2003, 2013). Thus, one way to increase flexibility may be to lower this updating threshold, which in turn could increase flexibility in a more generalizable fashion (Dreisbach & Fröber, 2019). Our experiments lacked explicit instructions for when to switch; instead, people had to discover the underlying rules on the basis of environmental feedback amid uncertainty (Behrens et al., 2007; Niv et al., 2015; Van Eylen et al., 2011). Self-initiated switches without explicit cueing are thought to require a higher degree of disengagement from the previous task to perform the switch (Manly et al., 2002; Van Eylen et al., 2011), making it less likely for participants to keep the alternative task in working memory. Studies that examined the influence of forced-choice task switching on voluntary task switches found that increasing the proportion of forced choices, in particular in combination with high switch rates, increases voluntary task-switching rates (Chiu et al., 2020; Fröber & Dreisbach, 2017). Other studies showed that by rewarding switches in a prior cued task-switching phase, it is possible to increase subsequent voluntary task-switching behavior (Braem, 2017), suggesting that cognitive flexibility is susceptible to its recent reinforcement-learning history. Taken together, in conditions in which experience is used to learn task models (Niv, 2019), meta-flexibility may be achieved by altering one's updating threshold or the rate at which this threshold is reached (both would be observed as a difference in reinforcement-learning rate), which in turn may promote transferable effects.

According to our volatility learning simulation, the current transfer effects may reflect the frequency structure during the learning phase, leading participants to continue with their previous expectations, even when the environmental volatility changes (Sabah et al., 2019). Previous studies have shown that inferring hidden underlying structural forms such as the relationships between stimuli, periodicities, or cognitive maps can enable rapid generalization of behavior to new environments (Behrens et al., 2018; Halford et al., 1998; Kemp et al., 2010; Mark et al., 2020). For instance, in the Mark et al. (2020) study, two groups of participants learned either hexagonal or community structure graphs and then learned a new graph with either the same or the alternate structure. The authors found that the experience with the first graph

shaped prior expectations over the underlying structure on the second graph, shown by improved task performance in the group that had the correct prior structural knowledge. By analogy, it is likely that in the current study, the temporal structure of the card-sorting tasks was generalized over task rules and stimuli using similar mechanisms of applying previously learned task adaptations, in this case, the updating threshold or learning rate (Baram et al., 2021). Consequently, the frequency of the switches encountered during the learning phase drove expectations and switch readiness in response to negative feedback during the transfer phase. This supports that adaptations to the abstract structure of the learning phase create an inductive bias that affects how participants make environmental inferences in the transfer phase.

Finally, it is important to consider the generalizability of our findings. The participant population consisted of U.S.-based adult MTurk workers. This online sample tends to be more diverse and representative of the U.S. population than college student samples, and it is known to replicate standard laboratory effects very reliably (Buhrmester et al., 2018; Crump et al., 2013). Although we would not expect cross-cultural differences in these effects, we cannot rule them out. An interesting question with respect to other participant populations, beyond clinical ones, is whether children and adolescents would show the same behavioral pattern because previous studies indicated developmental changes in reinforcement learning (e.g., Cohen et al., 2020; Shephard et al., 2014). Future work may also examine whether learning and transfer of cognitive flexibility show context specificity (e.g., Braem et al., 2020) and whether it is influenced by feedback type and feedback intensity (e.g., Yee et al., 2016, 2022) and by whether rules switch more stochastically rather than every fixed number of trials (e.g., Behrens et al., 2007).

In conclusion, we present a novel paradigm showing that participants transfer volatility-conditioned rule-learning rates to new temporal, task, and stimulus contexts. This transfer of a task- and stimulus-independent rule-learning parameter represents the formation and generalization of structural task knowledge for guiding cognitive control strategies. Given that impairments in the ability to adopt a contextually appropriate level of cognitive flexibility are thought to be central to various clinical conditions (e.g., Browning et al., 2015; Manly et al., 2002; Nassar & Troiani, 2020; Van Eylen et al., 2011), this new task protocol holds promise for developing a model-based assessment of individual differences in this ability in future studies. Furthermore, learning and transfer of cognitive strategies have been a central target in applied psychology, where "brain-training" interventions have been a popular idea to help improve cognitive functioning but have had little success at far transfer (Simons et al., 2016). Our demonstration of far transfer of cognitive flexibility settings acquired through trial-and-error learning may open the door to new, more successful approaches in this domain.

Transparency

Action Editor: M. Natasha Rajah
Editor: Patricia J. Bauer
Author Contribution(s)

Tanya Wen: Conceptualization; Data curation; Formal analysis; Methodology; Software; Visualization; Writing – original draft; Writing – review & editing.

Raphael M. Geddert: Formal analysis; Methodology; Software; Writing – review & editing.

Seth Madlon-Kay: Formal analysis; Methodology; Software; Supervision; Visualization; Writing – review & editing.

Tobias Egner: Conceptualization; Funding acquisition; Supervision; Writing – review & editing.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

This research was supported through NIH grant R01MH1169967 (T.E.).

Open Practices

All data and code have been made publicly available via the Open Science Framework and can be accessed at osf.io/wumry as well as https://github.com/tanya-wen/Meta-flexibility. This study was not preregistered.

ORCID iD

Tanya Wen https://orcid.org/0000-0002-4580-7831

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/09567976221141854

References

Baram, A. B., Muller, T. H., Nili, H., Garvert, M. M., & Behrens, T. E. J. (2021). Entorhinal and ventromedial prefrontal cortices abstract and generalize the structure of reinforcement learning problems. *Neuron*, *109*(4), 713–723.e7. https://doi.org/10.1016/j.neuron.2020.11.024

Barraclough, D. J., Conroy, M. L., & Lee, D. (2004). Prefrontal cortex and decision making in a mixed-strategy game. *Nature Neuroscience*, *7*(4), 404–410. https://doi.org/10.1038/nn1209

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. https://doi.org/10.1016/j.neuron.2018.10.002

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221. https://doi.org/10.1038/nn1954

Berg, E. A. (1948). A simple objective technique for measuring flexibility in thinking. *Journal of General Psychology*, *39*(1), 15–22. https://doi.org/10.1080/00221309.1948.9918159

Braem, S. (2017). Conditioning task switching behavior. *Cognition*, *166*, 272–276. https://doi.org/10.1016/j.cognition.2017.05.037

Braem, S., & Egner, T. (2018). Getting a grip on cognitive flexibility. *Current Directions in Psychological Science*, *27*(6), 470–476. https://doi.org/10.1177/0963721418787475

Braem, S., Liefooghe, B., & Abrahamse, E. (2020). Learning when to learn: Context-specific instruction encoding. *PsyArXiv*. https://doi.org/10.31234/OSF.IO/7TVX3

Browning, M., Behrens, T. E., Jocham, G., O'Reilly, J. X., & Bishop, S. J. (2015). Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nature Neuroscience*, *18*(4), 590–596. https://doi.org/10.1038/nn.3961

Buhrmester, M. D., Talaifar, S., & Gosling, S. D. (2018). An evaluation of Amazon's Mechanical Turk, its rapid rise, and its effective use. *Perspectives on Psychological Science*, *13*(2), 149–154. https://doi.org/10.1177/1745691617706516

Chiu, Y. C., & Egner, T. (2017). Cueing cognitive flexibility: Item-specific learning of switch readiness. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(12), 1950–1960. https://doi.org/10.1037/xhp0000420

Chiu, Y. C., Fröber, K., & Egner, T. (2020). Item-specific priming of voluntary task switches. *Journal of Experimental Psychology: Human Perception and Performance*, *46*(4), 434–441. https://doi.org/10.1037/xhp0000725

Cohen, A. O., Nussenbaum, K., Dorfman, H. M., Gershman, S. J., & Hartley, C. A. (2020). The rational use of causal inference to guide reinforcement learning strengthens with age. *npj Science of Learning*, *5*(1), Article 16. https://doi.org/10.1038/s41539-020-00075-3

Costa, V. D., Dal Monte, O., Lucas, D. R., Murray, E. A., & Averbeck, B. B. (2016). Amygdala and ventral striatum make distinct contributions to reinforcement learning. *Neuron*, *92*(2), 505–517. https://doi.org/10.1016/j.neuron.2016.09.025

Costa, V. D., Tran, V. L., Turchi, J., & Averbeck, B. B. (2015). Reversal learning and dopamine: A Bayesian perspective. *Journal of Neuroscience*, *35*(6), 2407–2416. https://doi.org/10.1523/JNEUROSCI.1989-14.2015

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLOS ONE*, *8*(3), Article e57410. https://doi.org/10.1371/journal.pone.0057410

Daw, N. D. (2011). Trial-by-trial data analysis using computational models. In M. R. Delgado, E. A. Phelps, & T. W. Robbins (Eds.), *Decision making, affect, and learning: Attention and Performance XXIII*. Oxford University Press.

Donahue, C. H., & Lee, D. (2015). Dynamic routing of task-relevant signals for decision making in dorsolateral prefrontal cortex. *Nature Neuroscience*, *18*(2), 295–301.

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. https://doi.org/10.1177/0963721418800030

Dreisbach, G., & Haider, H. (2006). Preparatory adjustment of cognitive control in the task switching paradigm. *Psychonomic Bulletin and Review*, *13*(2), 334–338. https://doi.org/10.3758/BF03193853

Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, *12*(7), 2121–2159.

Fröber, K., & Dreisbach, G. (2017). Keep flexible—keep switching! The influence of forced task switching on voluntary task switching. *Cognition*, *162*, 48–53. https://doi.org/10.1016/j.cognition.2017.01.024

Fröber, K., Jurczyk, V., & Dreisbach, G. (2021). Keep flexible—keep switching? Boundary conditions of the influence of forced task switching on voluntary task switching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(9), 1249–1262. https://doi.org/10.1037/xlm0001104

Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness*, *5*(2), 189–211.

Goschke, T. (2003). Voluntary action and cognitive control from a cognitive neuroscience perspective. In S. Maasen, W. Prinz, & G. Roth (Eds.), *Voluntary action: Brains, minds, and sociality* (pp. 49–85). Oxford University Press.

Goschke, T. (2013). Volition in action: Intentions, control dilemmas and the dynamic regulation of intentional control. In W. Prinz, A. Beisert, & A. Herwig (Eds.), *Action science: Foundations of an emerging discipline* (pp. 408–434). MIT Press.

Halford, G. S., Bain, J. D., Maybery, M. T., & Andrews, G. (1998). Induction of relational schemas: Common processes in reasoning and complex learning. *Cognitive Psychology*, *35*(3), 201–245. https://doi.org/10.1006/cogp.1998.0679

Jacobs, R. A. (1988). Increased rates of convergence through learning rate adaptation. *Neural Networks*, *1*(4), 295–307. https://doi.org/10.1016/0893-6080(88)90003-2

Jiang, J., Beck, J., Heller, K., & Egner, T. (2015). An insula-frontostriatal network mediates flexible cognitive control by adaptively predicting changing control demands. *Nature Communications*, *6*(1), 1–11. https://doi.org/10.1038/ncomms9165

Jiang, J., Heller, K., & Egner, T. (2014). Bayesian modeling of flexible cognitive control. *Neuroscience & Biobehavioral Reviews*, *46*, 30–43. https://doi.org/10.1016/j.neubiorev.2014.06.001

Katahira, K. (2015). The relation between reinforcement learning parameters and the influence of reinforcement history on choice behavior. *Journal of Mathematical Psychology*, *66*, 59–69. https://doi.org/10.1016/j.jmp.2015.03.006

Kemp, C., Goodman, N. D., & Tenenbaum, J. B. (2010). Learning to learn causal models. *Cognitive Science*, *34*(7), 1185–1243. https://doi.org/10.1111/j.1551-6709.2010.01128.x

Leboe, J. P., Wong, J., Crump, M., & Stobbe, K. (2008). Probe-specific proportion task repetition effects on switching costs. *Perception and Psychophysics*, *70*(6), 935–945. https://doi.org/10.3758/PP.70.6.935

Lee, D., Seo, H., & Jung, M. W. (2012). Neural basis of reinforcement learning and decision making. *Annual Review of Neuroscience*, *35*, 287–308. https://doi.org/10.1146/annurev-neuro-062111-150512

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135.

Manly, T., Hawkins, K., Evans, J., Woldt, K., & Robertson, I. H. (2002). Rehabilitation of executive function: Facilitation of effective goal management on complex tasks using periodic auditory alerts. *Neuropsychologia*, *40*(3), 271–281. https://doi.org/10.1016/S0028-3932(01)00094-X

Mark, S., Moran, R., Parr, T., Kennerley, S. W., & Behrens, T. E. J. (2020). Transferring structural knowledge across cognitive maps in humans and models. *Nature Communications*, *11*(1), Article 4783. https://doi.org/10.1038/s41467-020-18254-6

Massi, B., Donahue, C. H., & Lee, D. (2018). Volatility facilitates value updating in the prefrontal cortex. *Neuron*, *99*(3), 598–608. https://doi.org/10.1016/j.neuron.2018.06.033

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*(3), 134–140. https://doi.org/10.1016/S1364-6613(03)00028-7

Monsell, S., & Mizon, G. A. (2006). Can the task-cuing paradigm measure an endogenous task-set reconfiguration process? *Journal of Experimental Psychology: Human Perception and Performance*, *32*(3), 493–516. https://doi.org/10.1037/0096-1523.32.3.493

Nassar, M. R., & Troiani, V. (2020). The stability flexibility tradeoff and the dark side of detail. *Cognitive, Affective, & Behavioral Neuroscience*, *21*(3), 607–623. https://doi.org/10.3758/s13415-020-00848-8

Niv, Y. (2019). Learning task-state representations. *Nature Neuroscience*, *22*(10), 1544–1553. https://doi.org/10.1038/s41593-019-0470-8

Niv, Y., Daniel, R., Geana, A., Gershman, S. J., Leong, Y. C., Radulescu, A., & Wilson, R. C. (2015). Reinforcement learning in multidimensional environments relies on attention mechanisms. *Journal of Neuroscience*, *35*(21), 8145–8157. https://doi.org/10.1523/JNEUROSCI.2978-14.2015

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, *87*(6), 532–552.

R Core Team. (2022). *R: A language and environment for statistical ## computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rescorla, R., & Wagner, A. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). Appleton-Century-Crofts.

Sabah, K., Dolk, T., Meiran, N., & Dreisbach, G. (2019). When less is more: Costs and benefits of varied vs. fixed content and structure in short-term task switching training. *Psychological Research*, *83*(7), 1531–1542. https://doi.org/10.1007/s00426-018-1006-7

Sabah, K., Dolk, T., Meiran, N., & Dreisbach, G. (2021). Enhancing task-demands disrupts learning but enhances transfer gains in short-term task-switching training. *Psychological Research*, *85*(4), 1473–1487. https://doi.org/10.1007/S00426-020-01335-Y/FIGURES/5

Schulz, E., Franklin, N. T., & Gershman, S. J. (2020). Finding structure in multi-armed bandits. *Cognitive Psychology*, *119*, Article 101261. https://doi.org/10.1016/j.cogpsych.2019.101261

Shephard, E., Jackson, G. M., & Groom, M. J. (2014). Learning and altering behaviours by reinforcement: Neurocognitive differences between children and adults. *Developmental Cognitive Neuroscience*, *7*, 94–105. https://doi.org/10.1016/J.DCN.2013.12.001

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. L. (2016). Do "brain-training" programs work? *Psychological Science in the Public Interest*, *17*(3), 103–186. https://doi.org/10.1177/1529100616661983

Siqi-Liu, A., & Egner, T. (2020). Contextual adaptation of cognitive flexibility is driven by task- and item-level learning. *Cognitive, Affective, & Behavioral Neuroscience*, *20*(4), 757–782. https://doi.org/10.3758/s13415-020-00801-9

Stan Development Team. (2020). *RStan: The R interface to Stan*. R package version 2.21.2. http://mc-stan.org/

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. MIT Press.

Van Eylen, L., Boets, B., Steyaert, J., Evers, K., Wagemans, J., & Noens, I. (2011). Cognitive flexibility in autism spectrum disorder: Explaining the inconsistencies? *Research in Autism Spectrum Disorders*, *5*(4), 1390–1401. https://doi.org/10.1016/j.rasd.2011.01.025

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/S11222-016-9696-4

Watkins, C. J. C. H., & Dayan, P. (1992). *Q*-learning. *Machine Learning*, *8*, 279–292. https://doi.org/10.1007/BF00992698

Yee, D. M., Krug, M. K., Allen, A. Z., & Braver, T. S. (2016). Humans integrate monetary and liquid incentives to motivate cognitive task performance. *Frontiers in Psychology*, *6*, Article 2037. https://doi.org/10.3389/FPSYG.2015.02037/BIBTEX

Yee, D. M., Leng, X., Shenhav, A., & Braver, T. S. (2022). Aversive motivation and cognitive control. *Neuroscience & Biobehavioral Reviews*, *133*, Article 104493. https://doi.org/10.1016/j.neubiorev.2021.12.016

Yu, L. Q., Wilson, R. C., & Nassar, M. R. (2020). Adaptive learning is structure learning in time. *PsyArXiv*. https://doi.org/10.31234/OSF.IO/R637C