Electrophysiological Responses to Feedback during the Application of Abstract Rules

Matthew M. Walsh[1] and John R. Anderson[2]

Abstract

■ Much research focuses on how people acquire concrete stimulus–response associations from experience; however, few neuroscientific studies have examined how people learn about and select among abstract rules. To address this issue, we recorded ERPs as participants performed an abstract rule-learning task. In each trial, they viewed a sample number and two test numbers. Participants then chose a test number using one of three abstract mathematical rules they freely selected from: greater than the sample number, less than the sample number, or equal to the sample number. No one rule was always rewarded, but some rules were rewarded more frequently than others. To maximize their earnings, participants needed to learn which rules were rewarded most frequently. All participants learned to select the best rules for repeating and novel stimulus sets that obeyed the overall reward probabilities. Participants differed, however, in the extent to which they overgeneralized those rules to repeating stimulus sets that deviated from the overall reward probabilities. The feedback-related negativity (FRN), an ERP component thought to reflect reward prediction error, paralleled behavior. The FRN was sensitive to item-specific reward probabilities in participants who detected the deviant stimulus set, and the FRN was sensitive to overall reward probabilities in participants who did not. These results show that the FRN is sensitive to the utility of abstract rules and that the individual's representation of a task's states and actions shapes behavior as well as the FRN. ■

INTRODUCTION

Psychologists have long recognized the role of experience in learning. Thorndike (1911) famously articulated this idea in the law of effect: Humans and animals tend to repeat actions that have produced positive outcomes, and they tend to avoid actions that have produced negative outcomes. Thorndike first demonstrated the law of effect by placing a cat in a puzzle box that contained a lever. The lever opened a door, which released the cat from the box. Although the feline initially performed many ineffective actions, it quickly learned to press the lever to escape. The simplicity of Thorndike's law belies the complexity of behaviors that individuals learn from experience. In addition to forming associations that map specific stimuli to responses, we acquire abstract principles from experience. These principles allow us to respond to familiar and novel stimuli alike. For example, mathematical operations such as addition and multiplication are applicable to studied and arbitrary number pairs as well as to alphabetic variables. Although the ability to generalize rules from experience is central to intelligent behavior, few neuroscientific studies have examined how humans learn about and choose among such abstract rules.

The psychological literature contains several definitions of abstraction (for a review, see Bunge & Wallis, 2008). Common to these definitions are the ideas that abstract rules map multiple stimuli to responses, that abstract rules are not bound to exemplar knowledge, and that abstract rules are applicable to novel stimuli and in novel contexts. Here, we define abstraction in terms of relational integration (Badre & D'Esposito, 2009; Ramnani & Owen, 2004; Kroger et al., 2002). Concrete rules depend on first-order relational integration, which involves assigning a property to an item. For example, is the sample stimulus red, or is the sample number five? Abstract rules depend on second- or higher-order relational integration, which involves assigning relations between properties of items. For example, do the colors of the test and sample stimuli match, or does the value of the test number exceed the value of the sample number?

Abstract Rule Use

The pFC is thought to represent rules and their associated contexts (Passingham, 1993). In support of this idea, neurophysiological studies have found that pFC neurons encode the sensory dimension or task set that actively governs mappings between stimuli and responses (Yamada, Pita, Iijima, & Tsutsui, 2010; Mansouri, Matsumoto, & Tanaka, 2006; Asaad, Rainer, & Miller, 2000; White & Wise, 1999). When the relevant sensory dimension or task set changes, so too does the subset of active pFC neurons. pFC neurons also encode abstract rules and strategies

[1]Air Force Research Laboratory, Wright-Patterson Air Force Base, OH, [2]Carnegie Mellon University, Pittsburgh, PA

© 2013 Massachusetts Institute of Technology

Journal of Cognitive Neuroscience 25:11, pp. 1986–2002
doi:10.1162/jocn_a_00454

(Bongard & Nieder, 2010; Muhammad, Wallis, & Miller, 2006; Genovesio, Brasted, Mitz, & Wise, 2005; Wallis & Miller, 2003; Wallis, Anderson, & Miller, 2001). For example, in the match-to-sample task, an individual must select test stimuli that are identical to or different from the sample stimulus. Heterogeneous populations of pFC neurons respond depending on whether the match or nonmatch rule is in effect. Complementing these results, lesion studies have demonstrated that ablation of pFC regions in animals eliminates rule-guided behavior while leaving other types of responses intact (Baxter, Gaffan, Kyriazis, & Mitchell, 2009; Gaffan, Easton, & Parker, 2002; Bussey, Wise, & Murray, 2001; Dias, Robbins, & Roberts, 1996). Likewise, pFC damage selectively impairs rule-guided behavior in humans (Shallice & Burgess, 1991; Milner, 1963).

Neuroimaging studies have further contributed to our understanding of the neural basis of rule use (for reviews, see Schneider & Logan, 2009; Bunge, 2004). These studies detail a gradient of abstraction along the rostro-caudal axis of pFC, with progressively anterior regions representing increasingly abstract responses (Badre & D'Esposito, 2009; Koechlin & Summerfield, 2007; Christoff & Gabrieli, 2000). At a concrete level, simple conditional motor and response selection tasks activate the dorsal motor and premotor cortex (Badre & D'Esposito, 2007; Picard & Strick, 2001). At a more abstract level, tasks that require identifying relationships among stimuli engage more anterior lateral pFC regions (Badre & D'Esposito, 2007; Bunge, Kahn, Wallis, Miller, & Wagner, 2003). At a still more abstract level, tasks that necessitate simultaneously integrating multiple relationships among stimuli activate the frontopolar cortex (Bunge, Wendelken, Badre, & Wagner, 2005; Christoff, Ream, Geddes, & Gabrieli, 2003). Thus, tasks that engender the greatest relational complexity recruit the anterior extent of pFC.

Concrete Stimulus–Response Associations

Research on the use of abstract rules has proceeded largely independently of work on the incremental acquisition of stimulus–response associations. Whereas the former emphasizes the involvement of pFC in behavioral control, the latter focuses on the role of dopamine in reward learning (Berridge, 2007). In a series of studies, Schultz and colleagues demonstrated that the phasic response of dopamine neurons mirrored a reward prediction error signal (Schultz, 1998). When a reward was unexpectedly delivered, neurons showed enhanced activity. When a conditioned stimulus preceded reward, however, neurons no longer responded to reward delivery. Rather, they responded to the earlier conditioned stimulus. Finally, when a reward was unexpectedly omitted, dopamine neurons showed depressed activity at the expected time of reward delivery. These results demonstrate that the responses of dopamine neurons depend on differences between actual and expected outcomes, or reward prediction errors.

Neuroimaging experiments have since extended these results to humans. BOLD responses in the BG, a collection of structures innervated by dopamine neurons, mirror reward prediction errors (McClure, York, & Montague, 2004; O'Doherty, 2004). These findings have given rise to the idea that dopamine neurons compute prediction errors. The BG use these prediction errors to modify the strength of stimulus–response associations (Packard & Knowlton, 2002). Positive prediction errors strengthen associations, ensuring that the individual repeats those responses in the future, and negative prediction errors weaken associations, ensuring that the individual avoids those responses in the future.

The Feedback-related Negativity

Studies using scalp-recorded ERPs have also been informative with respect to this issue. ERP studies of reward learning have revealed a frontocentral negativity that emerges from 200 to 400 msec after negative feedback (Miltner, Braun, & Coles, 1997). This feedback-related negativity (FRN) sometimes appears as a negative deflection following losses and sometimes as a positive deflection following wins (Holroyd, Pakzad-Vaezi, & Krigolson, 2008). Several features of the FRN indicate that it reflects a reward prediction error signal (Walsh & Anderson, 2012). First, the FRN is sensitive to violations of reward probability and magnitude (Walsh & Anderson, 2011a, 2011b; Holroyd, Nieuwenhuis, Yeung, & Cohen, 2003). Second, FRN amplitude correlates with posterror behavioral adjustment (Cohen & Ranganath, 2007). Third, source localization studies (Gehring & Willoughby, 2002; Miltner et al., 1997), single cell recordings (Ito, Stuphorn, Brown, & Schall, 2003; Niki & Watanabe, 1979), and neuroimaging experiments (Holroyd et al., 2004) indicate that the FRN originates in the ACC, a region involved in integrating reward history to guide action selection.

These ideas underlie the reinforcement learning theory of the error-related negativity (Holroyd & Coles, 2002). According to this theory, the dopamine system monitors outcomes to determine whether things have gone better or worse than expected. Positive prediction errors induce phasic increases in dopamine firing rates, and negative prediction errors induce phasic decreases in dopamine firing rates. Dopamine neurons convey error signals to the BG where they are used to revise expectations. Dopamine neurons also convey error signals to cortical structures such as the anterior cingulate where they are used to integrate reward information with action selection. The scalp-recorded FRN reflects the impact of dopamine signals on neurons in the anterior cingulate. Phasic decreases in dopamine activity yield a more negative FRN, and phasic increases in dopamine activity yield a less negative FRN.

Although the FRN has been the focus of many studies (for a review, see Walsh & Anderson, 2012), nearly all involve the acquisition of concrete stimulus–response

associations. For example, in one of the earliest reports of the FRN, participants responded to a visual stimulus with a left or a right button press (Holroyd & Coles, 2002). In this and subsequent experiments, the mapping between stimuli and responses was arbitrary. Consequently, participants learned each stimulus–response association separately and could not generalize responses to novel stimuli. In an experiment that did permit generalization, participants categorized visual stimuli that varied along multiple dimensions (Krigolson, Pierce, Holroyd, & Tanaka, 2009). The probabilistic relationship between dimensional values and category membership determined the mapping between stimuli and responses. Consequently, participants could respond correctly to novel stimuli. Although this experiment permitted generalization, participants' responses may or may not have depended on relational integration, which is the concern of this article. When interactions among multiple stimulus dimensions define category membership, as in Krigolson et al.'s (2009) experiment, people do not use explicit rules to compare stimulus dimensions against critical values (Ashby & O'Brien, 2005); rather, they form graded representations that map the perceptual space to responses.

Overview of Present Research

Abstract rules and concrete stimulus–response associations likely draw on partially overlapping neural structures. However, three differences in the paradigms used to study these types of responses obscure their similarities. First, most studies of rule use focus on asymptotic performance, whereas most studies of stimulus–response associations focus on learning. Second, most rule tasks employ deterministic rewards, whereas most stimulus–response association tasks employ probabilistic rewards. Third, most studies of rule use report tonic neural signals that appear before or in concert with rule application, whereas most studies of stimulus–response associations report phasic neural signals evoked by outcomes or feedback-related events.

The present research borrows elements from both types of paradigm. We asked participants to select among three abstract mathematical rules that were probabilistically rewarded. Mathematical principles operate on variable numerical quantities rather than fixed sensory stimuli. As such, mathematical principles can be used to study abstract rule use. In each trial, we presented participants with one sample number and two test numbers (Figure 1). Participants chose a test number using one of three abstract mathematical rules they freely selected from: greater than the sample number, less than the sample number, or equal to the sample number. No rule was always rewarded, but some rules were rewarded more frequently than others were. To maximize their earnings, participants needed to learn which rules were rewarded most frequently.

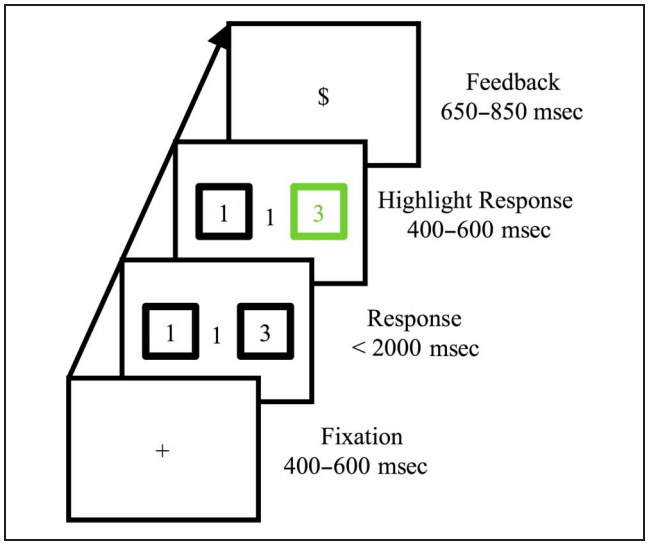

Figure 1. Trial procedure. Participants viewed a sample number at the center of the screen. Two test numbers appeared beside the sample number. After participants selected a test number, feedback appeared.

Some configurations of sample and test numbers repeated in 225 trials. In these "standard" trials, the globally optimal rule was rewarded most frequently. Participants could respond correctly using concrete stimulus–response associations (i.e., "Choose three because three has been rewarded") or abstract rules (i.e., "Choose the test number that is greater than the sample").[1] Other configurations of sample and test numbers appeared only once. In these "novel" trials, the globally optimal rule was rewarded most frequently. Participants could respond correctly only by using abstract rules. Still other configurations of sample and test numbers repeated in 225 trials but deviated from the overall reward probabilities. In these "deviant" trials, the globally optimal rule was rewarded less frequently than the globally suboptimal rule. Participants could respond correctly using concrete stimulus–response associations. Participants could also respond correctly using abstract rules, although doing so required that they not overgeneralize knowledge from standard and novel trials.[2] We analyzed choice data to determine whether participants used abstract rules, concrete stimulus–response associations, or both. We also modeled their behavior to address the question of whether individuals learn about the utility of abstract rules in the same way that they acquire concrete stimulus–response associations.

Besides studying overt behavior, we used ERPs to explore how people learn about and choose among abstract rules. The reinforcement learning theory of the error-related negativity grants the FRN (and by extension, the anterior cingulate) a role in the acquisition of concrete stimulus–response associations and abstract rules (Holroyd & Coles, 2002). Although studies have demonstrated the role of the FRN in the formation of concrete

stimulus–response associations, no study has yet examined the involvement of the FRN in the use of abstract rules. To that end, we asked whether negative feedback would evoke an FRN as participants chose among abstract response rules. We expected that the FRN would reflect reward probabilities in standard trials: Abstract rules and concrete associations were both applicable in such trials. If the FRN were sensitive to the utility of abstract rules, the FRN would also reflect reward probabilities in novel trials where abstract rules alone were applicable. Furthermore, if participants overgeneralize abstract rules to deviant items, the FRN would reflect overall reward probabilities rather than item-specific reward probabilities in deviant trials. Alternatively, if the FRN were sensitive only to the utility of concrete stimulus–response associations, the FRN would not reflect reward probabilities in novel trials where there was no opportunity to form concrete associations. Additionally, the FRN would reflect item-specific reward probabilities rather than overall reward probabilities in deviant trials.

EXPERIMENT 1

Methods

Participants

Fourteen graduate and undergraduate students participated on a paid volunteer basis (seven men and seven women, ages ranging from 19 to 31 years with a mean age of 23 years). All were right-handed, and none reported a history of neurological impairment.

Procedure

At the start of each trial, a sample number appeared at the center of the screen. Two test numbers appeared beside the sample number (Figure 1). Participants pressed a key bearing a left arrow to select the test number on the left, and they pressed a key bearing a right arrow to select the test number on the right. They used their left and right index fingers to make selections. The selected number turned green. Participants had 2,000 msec to respond. If they failed to respond within 2,000 msec, both numbers turned red.

Participants selected among three rules that were based on the relationship between the values of the test numbers and the sample number: greater than—the test number was greater than the sample number; less than—the test number was less than the sample number; and equal to—the test number was equal to the sample number. Participants were informed about the significance of the relationship between the values of the test numbers and the sample number. They were told, "Use the feedback to learn whether it is best to select test numbers that are greater than the sample number, less than the sample number, or equal to the sample number." Positive feedback signifying reward was denoted by $, negative feedback signifying no reward was denoted by X, and missed responses were denoted by !.

The sample and test numbers defined the stimulus sets (Figure 2). The experiment contained two standard sets that repeated in 225 trials each, 450 novel sets that appeared in one trial each, and one deviant set that

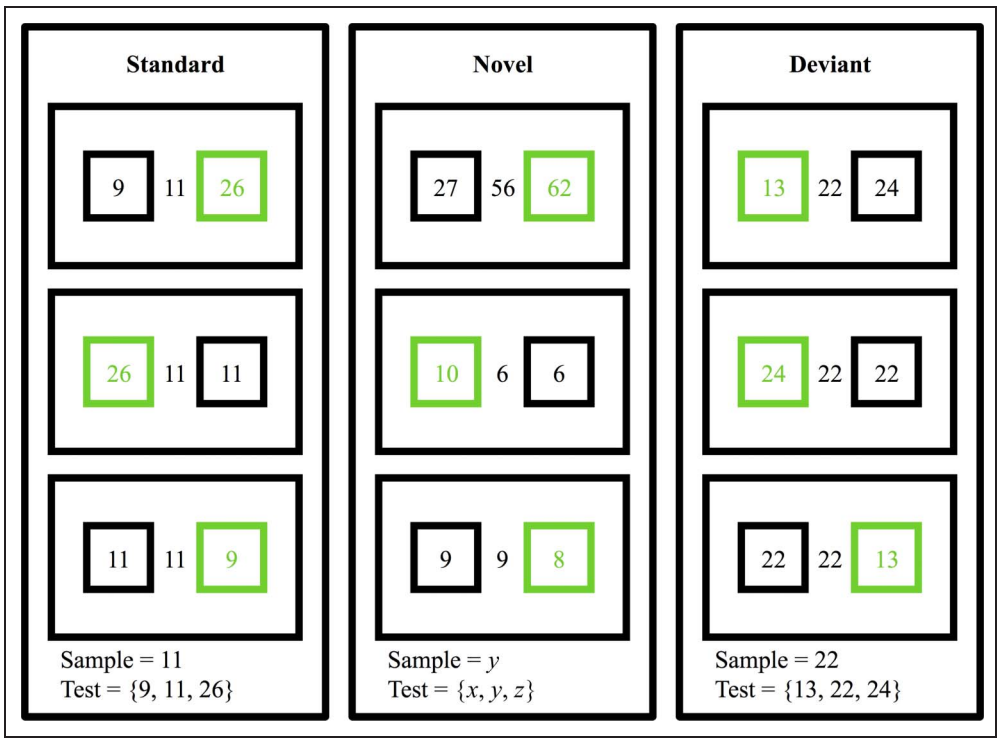

Figure 2. Example of a standard stimulus set appearing in three trials (left). In each trial, the sample number is fixed and two test numbers are drawn from a pool of three test numbers. Example of three novel stimulus sets appearing in three trials (center). Sample and test numbers are unique in each trial. Example of the deviant stimulus set appearing in three trials (right). In each trial, the sample number is fixed and two test numbers are drawn from a pool of three test numbers.

Table 1. Stimulus Sets and Reward Probabilities

			Reward Probability (%)		
	Stimulus Sets	*Trials per Set*	*Optimal*	*Suboptimal*	*Equal to*
Standard	2	225	75	25	0
Novel	450	1	75	25	0
Deviant	1	225	25	75	0

repeated in 225 trials (Table 1). For standard and novel sets, the globally optimal rule was rewarded with 75% probability and the globally suboptimal rule was rewarded with 25% probability. For the deviant set, these probabilities reversed (i.e., the globally optimal rule was rewarded with 25% probability and the globally suboptimal rule was rewarded with 75% probability). Participants learned these probabilities from experience. Greater than and less than were randomly assigned as globally optimal and suboptimal rules across participants and in a counterbalanced manner, and equal to was never rewarded.[3]

The experiment contained 1125 trials. In each trial, the inclusion of two test numbers allowed participants to select from two rules. This yielded three choice pairs of reward probabilities (75/25, 75/0, 25/0) that occurred with equal frequency. For example, the test numbers in the top row of Figure 2 permit choice between less than and greater than, the test numbers in the middle row permit choice between greater than and equal to, and the test numbers in the bottom row permit choice between equal to and less than. Although no rule was rewarded with 100%, participants could maximize pay by selecting the rule that was rewarded most often within each pair. For standard and novel sets, this meant selecting the globally optimal rule in two cases (75/25 and 75/0) and the globally suboptimal rule in one case (25/0). For deviant sets, this meant selecting the globally suboptimal rule in two cases (75/25 and 25/0) and the globally optimal rule in one case (75/0).

Stimuli were drawn from the numbers 1 to 99. Numbers assigned to the standard and deviant sets appeared only within those sets (Figure 2). The remaining numbers were assigned to the novel sets. Although numbers necessarily repeated within novel sets, no combination of test numbers or of test and sample numbers repeated.

Participants completed 30 practice trials before beginning the main experiment. To prevent participants from learning about the utility of rules used in the experiment prematurely, we instructed them to select between two different rules during practice: odd (the value of the test number is odd) and even (the value of the test number is even). The rules were rewarded with 75% probability and 25% probability and were counterbalanced across participants. Participants were told that these rules no longer applied once the experiment began.

EEG Recording and Analysis

Participants sat in an electromagnetically shielded booth. Stimuli appeared on a CRT monitor placed behind radio frequency shielded glass and set 60 cm from participants. The EEG was recorded from 32 Ag-AgCl sintered electrodes (10–20 system). Electrodes were also placed on the right and left mastoids. The right mastoid served as the reference electrode, and scalp recordings were algebraically re-referenced off-line to the average of the right and left mastoids. The vertical EOG was recorded as the potential between electrodes placed above and below the left eye, and the horizontal EOG was recorded as the potential between electrodes placed at the external canthi. The EEG and EOG signals were amplified by a Neuroscan bioamplification system with a bandpass of 0.1–70.0 Hz and were digitized at 250 Hz. Electrode impedances were kept below 5 kΩ.

The EEG recording was decomposed into independent components using the EEGLAB infomax algorithm (Delorme & Makeig, 2004). Components associated with eye blinks were visually identified and projected out of the EEG recording. Epochs of 850 msec (including a 200-msec baseline) were then extracted from the continuous recording and corrected over the prestimulus interval. Epochs containing voltages above +75 μV or below −75 μV were excluded from further analysis (<9% epochs).

We created feedback-locked ERPs for trials where participants selected the 75 rule or the 25 rule. A series of paired t tests revealed that neural responses depended only on the selected rule and not the choice pair in which it appeared (i.e., neural responses after the 75 rule did not depend on whether it appeared with the 25 rule or the 0 rule and neural responses after the 25 rule did not depend on whether it appeared with the 75 rule or the 0 rule). Consequently, we excluded the factor of choice pair from further ERP analyses. The P300, an endogenous ERP component evoked by stimulus presentation, is sensitive to event probabilities (Duncan-Johnson & Donchin, 1977). To isolate the effects of outcome likelihood on the FRN from the effects of outcome likelihood on the P300, we adopted the

difference wave approach advocated by Holroyd, Krigolson, Baker, Lee, and Gibson (2009), and we compared losses and wins that were equally likely (see also Luck, 2005). We created a probable outcome difference wave (losses after 25 rule − wins after 75 rule) and an improbable outcome difference wave (losses after 75 rule − wins after 25 rule). We measured the FRN as the mean voltage of the difference waves from 240 to 400 msec after feedback onset. We analyzed data from three midline sites (FCz, Cz, and CPz), and we applied the Greenhouse–Geisser correction when factors had more than two levels.

Results

Behavioral Results

Participants favored the 75 rule from the 75/25 and 75/0 pairs, and they favored the 25 rule from the 25/0 pair (Figure 3). A 3 (Trial Type: standard, novel, deviant) × 3 (Rule Pair: 75/25, 75/0, 25/0) repeated-measures ANOVA revealed main effects of Trial Type, $F(2, 26) = 7.690$, $p < .01$, and Rule Pair, $F(2, 26) = 11.454$, $p < .001$. Participants were somewhat more likely to select the globally best rules in standard and novel trials than in deviant trials, and they were far more likely to select the globally best rules in trials with the 75/25 and 75/0 pairs than in trials with the 25/0 pair. The interaction between Trial Type and Rule Pair was also significant, $F(4, 52) = 3.253$, $p < .05$. Selection of the 75 rule from the 75/25 pair maximized the probability of reward in standard and novel trials but minimized the probability of reward in deviant trials. Accordingly, participants were somewhat less likely to choose the 75 rule from the 75/25 pair in deviant trials as compared with standard trials, $t(13) = 2.260$, $p < .05$, and novel trials, $t(13) = 2.066$, $p < .1$.

We derived two indices to determine whether individuals responded differently to the deviant stimulus set. First, we isolated trials that featured the 75/25 pair, and we calculated the difference in preference for the 75 rule during standard and deviant trials (i.e., Standard(75/25) − Deviant(75/25)). If participants detected the probability reversal, they would favor the 75 rule in standard trials and the 25 rule in deviant trials. Second, we isolated trials that featured the 75/0 and 25/0 pairs. We calculated the difference in accuracy between 75/0 and 25/0 pairs and for standard and deviant trials, (i.e., [Standard(75/0) − Standard(25/0)] − [Deviant(75/0) − Deviant(25/0)]). We expected that participants would respond more accurately to the 75/0 pair than to the 25/0 pair during standard trials: The 75 rule had far greater utility than the 0 rule, whereas the 25 rule had only slightly greater utility than the 0 rule. Furthermore, if participants detected the probability reversal, they would respond more accurately to the 25/0 pair than to the 75/0 pair during deviant trials; because probabilities reversed in deviant trials, the 25/0 pair now contained a clear winner and the 75/0 pair did not. We calculated these two indices for each participant. As seen in Figure 4, they were strongly correlated, $r = .74$, $p < .01$. Thus, the weak interaction present in the selection data (Figure 3) arose from a subset of participants that exhibited moderate sensitivity to the deviant stimulus set.

ERP Results

We quantified the FRN as the difference between waveforms following losses and wins that were equally likely. Participants displayed an FRN for improbable outcomes (losses after 75 rule − wins after 25 rule) and probable outcomes (losses after 25 rule − wins after 75 rule) and for all trial types (Figure 5). A 3 (Trial Type: standard, novel, deviant) × 2 (Outcome Likelihood: probable, improbable) × 3 (Site: FCz, Cz, CPz) ANOVA of FRN amplitude revealed a significant effect of Outcome Likelihood, $F(1, 13) = 20.545$, $p < .001$, but not of Site, $F(2, 26) = .541$, $p > .1$, or Trial Type, $F(2, 26) = .847$, $p > .1$. The FRN was greater for improbable than for probable

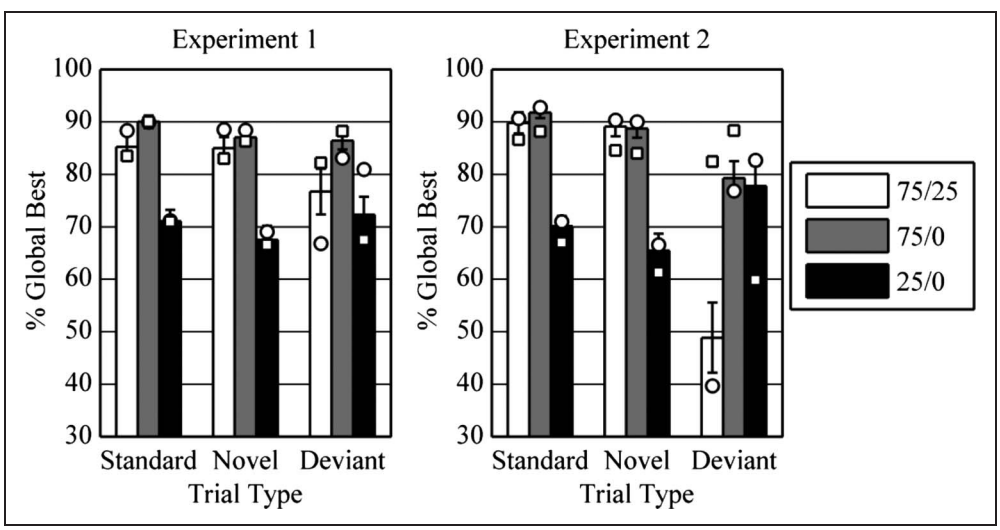

Figure 3. Percentage of trials where participants selected the globally best rule (i.e., 75/25, 75/0, 25/0) by trial type and choice pair (±1 within-subject *SE*). Circles show performance of aware participants, and squares show performance of unaware participants (for explanation of groups, see Model-based analysis).

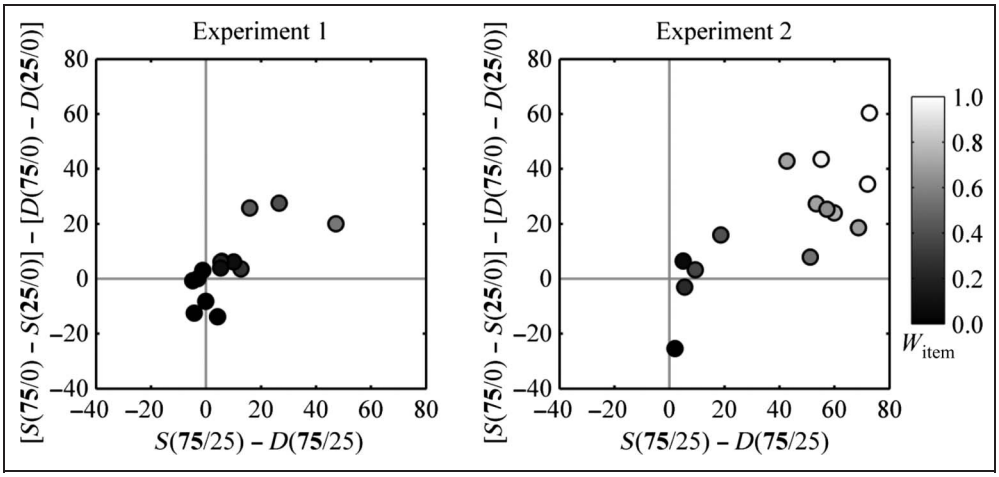

Figure 4. Relationship between behavioral indices of sensitivity to deviant set. Positive values denote greater sensitivity. Grayscale depicts individual weight parameter (W_{item}) estimates for Combination model (for explanation of W_{item}, see Model-based analysis).

outcomes at all sites. No interactions involving the factor of Trial Type approached significance (all $p > .1$).

We then focused on site FCz where the FRN was maximal (Figure 6). A 3 (Trial Type) × 2 (Outcome Likelihood) ANOVA revealed a main effect of Outcome Likelihood, $F(1, 13) = 19.443$, $p < .001$, but not of Trial Type, $F(2, 26) = 1.109$, $p > .1$. The interaction was also not significant, $F(2, 26) = 1.540$, $p > .1$. Improbable outcomes uniformly yielded the largest FRN across all trial types (Figure 7). Thus, at the level of the group, the FRN was predominantly sensitive to the utility of abstract rules for standard, novel, and deviant items alike.

Model-based Analysis

All participants appeared to use abstract rules, but they differed in the extent to which they overgeneralized those rules to deviant items. To quantify the contributions of general and item-specific knowledge to choice behavior, we modeled the learning and decision-making process. To do so, we used temporal difference learning (Sutton & Barto, 1998), an influential technique in the field of artificial intelligence with strong ties to psychological theories of human and animal conditioning (Walsh & Anderson, in press) and physiological models of phasic dopamine responses (Schultz, 1998). Central to this technique is the idea that differences between actual and expected outcomes, or reward prediction errors, provide teaching signals. After the individual experiences an outcome, a prediction error is calculated and used to update the estimated value of the previous action. In this way, the individual can learn to associate states and actions with rewards.

The temporal difference learning framework does not directly specify what constitutes an action. For example, in our experiment, individuals could represent actions in terms of concrete stimulus–response associations (i.e., "Choose three if the sample number is one") or abstract rules (i.e., "Choose the test number that is greater than the sample number"). The temporal difference learning framework also does not specify what constitutes a state. For example, individuals could learn about the utility of general rules applicable to all stimulus sets or specific rules applicable to one stimulus sets. To understand how these different representations contributed to participants' behavior, we compared three models: One model

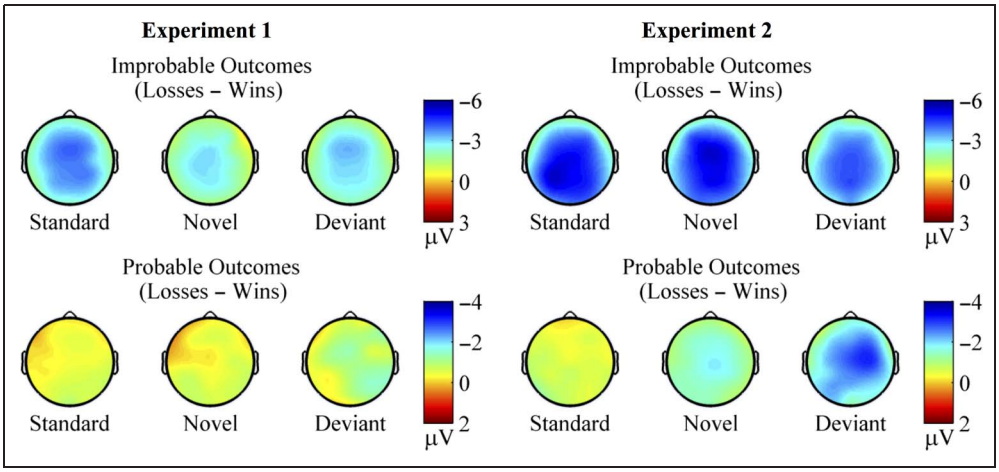

Figure 5. Topography of the FRN by condition and outcome likelihood. Time is from 240 to 400 msec with respect to feedback onset.

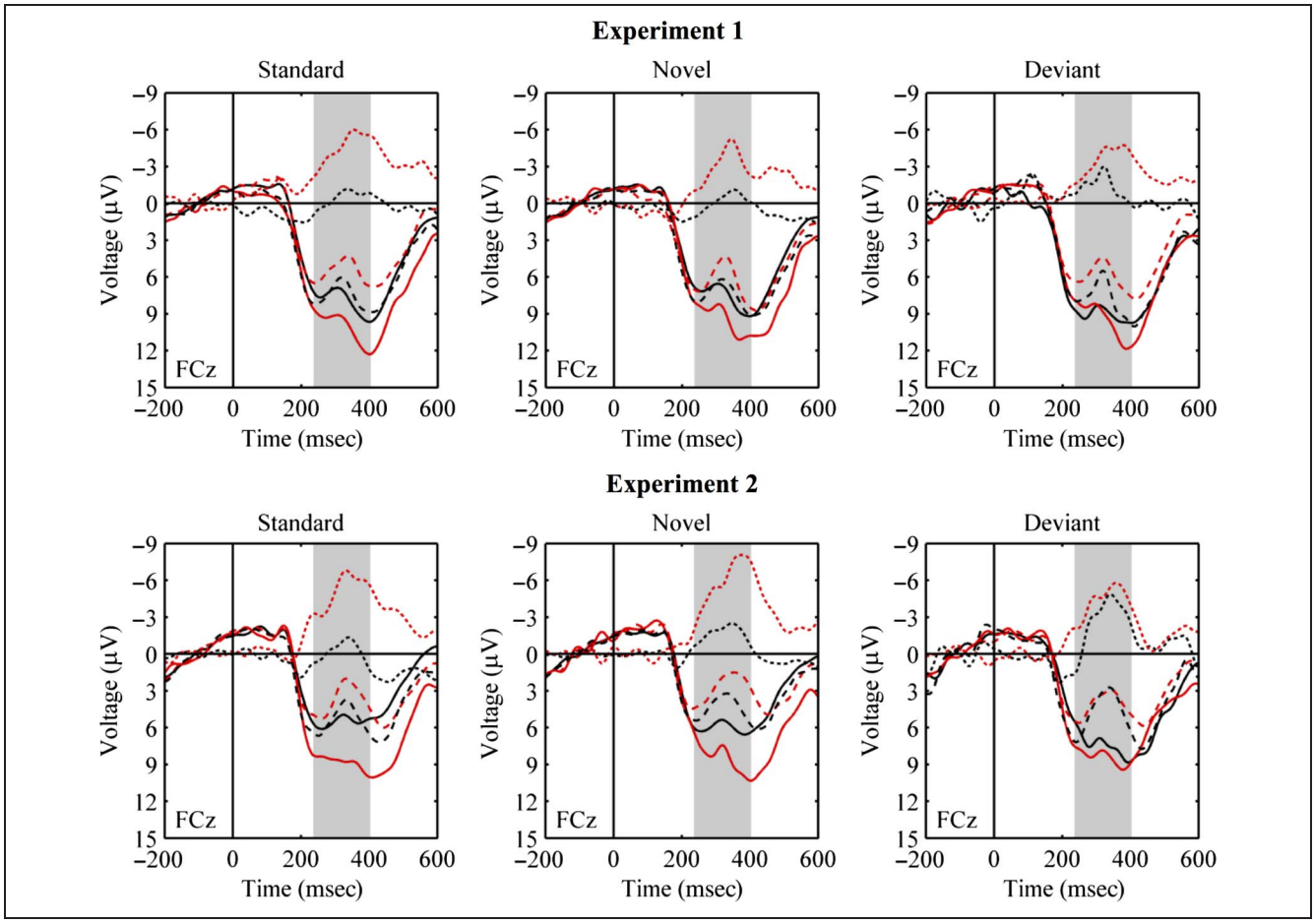

Figure 6. ERPs for improbable losses (dashed red), probable losses (dashed black), probable wins (solid black), and improbable wins (solid red) by trial type at site FCz. FRN (calculated as the difference between loss and win waveforms) for improbable outcomes (dotted red) and probable outcomes (dotted black).

learned the utility of general rules applicable to all stimulus sets, another model learned item-specific rules applicable to one stimulus sets, and the final model learned both.

The Combination model represented actions in terms of general rules (i.e., greater than the sample, less than the sample, and equal to the sample) that were applicable to all stimulus sets ($Q_{general}$). The Combination model also represented actions in terms of item-specific rules that were applicable to one stimulus set each (Q_{item}). Upon encountering standard and deviant stimulus sets, which repeated throughout the experiment, the Combination model calculated a weighted average of the utility estimates

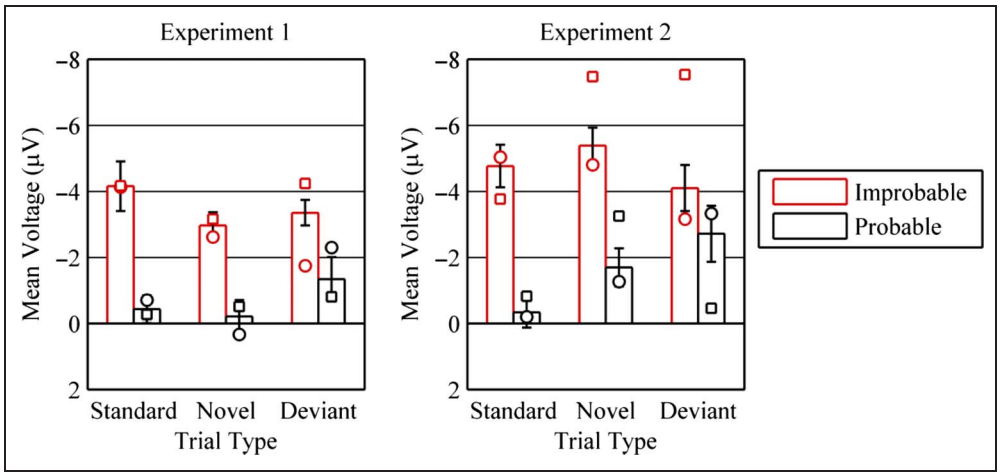

Figure 7. Mean FRN amplitude (losses − wins) at site FCz from 240 to 400 msec (±1 within-subject *SE*). Circles show FRN for aware participants, and squares show FRN for unaware participants (for explanation of groups, see Model-based Analysis).

provided by the general and item-specific components, $Q_{\text{combine}} = (1 - W_{\text{item}}) \times Q_{\text{general}} + W_{\text{item}} \times Q_{\text{item}}$. Upon experiencing novel trials, which did not repeat throughout the experiment, the Combination model calculated utility as the estimate provided by the general component only, $Q_{\text{combine}} = Q_{\text{general}}$.

In each trial, the model chose from two test numbers. The probability of selecting a number (π_a) was determined by a softmax decision rule (Sutton & Barto, 1998),

$$\pi_a = \frac{\exp(Q_{\text{combine}}(a)/\tau)}{\exp(Q_{\text{combine}}(a)/\tau) + \exp(Q_{\text{combine}}(b)/\tau)}$$

Selection noise (τ) controlled the degree of randomness in choices. The softmax selection rule resembles Luce's (1977) choice axiom. The softmax selection rule also approximates greedy selection among actions whose utility estimates are subject to continuously varying noise (Fu & Anderson, 2006).

After each outcome (r), the model computed a reward prediction error for the general rule that contributed to the combined utility estimate, $\delta_{\text{general}} = r - Q_{\text{general}}(a)$. The model also computed a reward prediction error for the item-specific rule that contributed to the combined utility estimate, $\delta_{\text{item}} = r - Q_{\text{item}}(a)$. The model used these prediction errors to update the utility of the general rule, $Q_{\text{general}}(a) \leftarrow Q_{\text{general}}(a) + \alpha \times \delta_{general}$, and the item-specific rule, $Q_{\text{item}}(a) \leftarrow Q_{\text{item}}(a) + \alpha \times \delta_{item}$, that contributed to the combined utility estimate. Learning rate (α) scaled the size of utility updates. The model received rewards of +1 and 0 for positive and negative feedback, respectively.

We used the behavioral data to estimate parameter values for each participant (α, τ, and W_{item}). We presented the model with the history of choices and rewards that the participant experienced. For each trial, t, we calculated the probability that the model would make the same choice as the participant, $p_k(t)$. We used the simplex optimization algorithm (Nelder & Mead, 1965) with multiple start points to identify parameter values that maximized the log likelihood of the observed choices, LLE = $\Sigma_t \ln(p_k(t))$.

We also evaluated the performance of two nested variants of the Combination model. In the General model, we set $W_{\text{item}} = 0$ to simulate exclusive use of general rules. In the Item model, we set $W_{\text{item}} = 1$ to simulate exclusive use of item-specific rules. For each participant, we determined whether the General or Item models outperformed the Combination model using the likelihood ratio test (Lewandowsky & Farrell, 2011).

Behavioral Results

Averaged across trials, the likelihood of the observed choices produced by the best fitting parameterization of the Combination model for each participant equaled .74 ± .02. Estimates for the weight parameter approached zero (indicative of exclusive use of general rules) for all but five participants (Figure 4). These same five participants exhibited the greatest scores for the two indices used to assess behavioral sensitivity to deviant items.

The Item model did not outperform the Combination model for any participant (all $p > .1$ by the likelihood ratio test). The General model, in contrast, outperformed the Combination model for all but five participants ($p < .05$). On the basis of these contrasts, we identified the five participants best described by the Combination model, and the nine best described by the General model. Because item-specific utility values influenced behavior in the former group but not in the latter, we simply refer to these groups as "aware" and "unaware." We compared parameter estimates for the Combination model between aware and unaware participants. Expectedly, the value of the weight parameter (W_{item}) was greater for aware participants, $t(12) = 7.524$, $p < .0001$ (Table 2). No other parameters differed between groups.

We then separated the behavioral data according to participants' model-based classifications (Figure 3). The interaction between trial type and rule pair was significant in aware participants, $F(4, 16) = 6.343$, $p < .01$, but not in unaware participants, $F(4, 32) = 1.066$, $p > .1$. Although aware participants exhibited sensitivity to deviant items, they did not completely reverse their decisions for the 75/25 pair in deviant trials. Response accuracy for the 75/25 pair, defined as the percentage of trials where participants selected the locally optimal rule, remained lower for deviant trials than for standard trials, $t(4) = 3.595$, $p < .05$, or novel trials, $t(4) = 3.799$, $p < .05$. These results describe aggregate responding over the course of the experiment. Aware participants exhibited increasing sensitivity to deviant items, although they never fully reversed their decisions for the 75/25 pair in deviant trials

Table 2. Model Parameter Estimates by Participant Subgroup for Experiments 1 and 2

Experiment	*Subgroup*	α	W_{item}	τ
1	*Aware* ($n = 5$)	.02 ± .01	.37 ± .06	.14 ± .02
	Unaware ($n = 9$)	.07 ± .05	.02 ± .01	.17 ± .03
2	*Aware* ($n = 11$)	.06 ± .03	.71 ± .07	.19 ± .02
	Unaware ($n = 3$)	.16 ± .16	.09 ± .07	.29 ± .18

(Appendix). Unaware participants never exhibited any sensitivity to deviant items.

Ancillary ERP Analysis

We divided the ERP data according to participants' model-based classifications and computed FRN amplitude at site FCz by trial type and outcome likelihood (Figure 7). Aware participants showed an interaction between trial type and outcome likelihood, $F(2, 8) = 5.753, p < .05$, but unaware participants did not, $F(2, 18) = .524, p > .1$. These results show that participants who exhibited behavioral awareness of deviant items also displayed neural sensitivity in the form of a differential FRN to deviant items.

Discussion

Participants responded correctly to standard stimulus sets. Because these sets repeated throughout the experiment, participants could base selections on concrete stimulus–response associations or abstract rules. Participants also responded correctly to novel stimulus sets. Because these sets never repeated, participants could base selections only on abstract rules. Most participants' responses to the deviant stimulus set were indistinguishable from their responses to standard and novel sets. In principle, participants could have responded correctly to the deviant set by using concrete stimulus–response associations or by learning about the utility of abstract rules as applied to those items specifically. In actuality, the majority overgeneralized the abstract rules that were most suitable for standard and novel sets.

Our model-based analysis supports the idea that most participants applied a common set of abstract rules to standard, novel, and deviant stimulus sets alike. Because the Item model cannot generalize and because sample and test numbers never repeat in novel trials, the Item model performs at chance in novel trials. Participants performed equally well in standard and novel trials, however. Because the General model ignores information about specific number values, the General model treats standard, novel, and deviant trials identically. So too did most participants. Some exhibited sensitivity to the probability reversal in deviant trials, however. For these participants, the Combination model permitted generalization in novel trials and item-specific learning in deviant trials.

Negative feedback produced an FRN. The amplitude of the FRN was greater for improbable outcomes than for probable outcomes and was comparable during standard, novel, and deviant trials. The finding that the FRN was equivalent during standard and novel trials demonstrates that the FRN is sensitive to the utility of abstract response rules. That the FRN was also equivalent during deviant trials, where the utility of concrete stimulus–response associations directly opposed the utility of abstract rules, underscores this point.

The modeling results revealed a subset of individuals who displayed behavioral sensitivity to deviant items. Interestingly, these participants' neural responses reflected some sensitivity to the item-specific reward probabilities as well. This suggests that behavioral and neural adaptation take place over a shared representation of the task's states and actions, a point that we explore further in Experiment 2.

EXPERIMENT 2

In Experiment 1, the FRN tracked the item-specific probability of reward in participants who detected deviant trials, whereas the FRN tracked the overall probability of reward in participants who did not. These results indicate that the representation of states and actions that guides behavior also influences the FRN. Because so few participants detected the deviant stimulus set, however, this conclusion remains speculative. Our goal in Experiment 2 was to increase the number of participants who distinguished between deviant items as compared with standard and novel items.

To do so, we adopted two theoretically motivated manipulations (Lovett & Schunn, 1999). First, we presented standard, novel, and deviant stimulus sets in distinctive font colors. Second, we instructed participants to pay attention to the font color, and we informed them that the reward probabilities differed for one color. We predicted that these manipulations would produce a greater number of aware participants in Experiment 2. If behavioral adaptation and neural adaptation take place over a shared representation of a task's states and actions, the FRN will be sensitive to item-specific reward probabilities in aware participants, and the FRN will be insensitive to item-specific reward probabilities in unaware participants.

Methods

Participants

Fourteen graduate and undergraduate students participated on a paid volunteer basis (six men and eight women, ages ranging from 19 to 38 years with a mean age of 25 years). All were right-handed, and none reported a history of neurological impairment.

Procedure

The procedure was identical to Experiment 1 with one exception: Font color varied by trial type. Standard, novel, and deviant stimulus sets appeared in different colors that were randomized across participants. For example, one standard stimulus set always appeared in red and the other in yellow, half of the novel stimulus sets appeared in blue and the others in green, and the deviant stimulus set always appeared in cyan.[4] As in Experiment 1,

participants were informed about the significance of the relationship between the values of the test numbers and the sample number. They were further informed about the significance of font colors. They were told, "In different trials, the numbers will appear in different colors. Pay attention to the color of the numbers. The reward probabilities are the same for most colors, but they are different for one color."

Results

Behavioral Results

Participants favored the 75 rule from the 75/25 and 75/0 pairs, and they favored the 25 rule from the 25/0 pair (Figure 3). A 3 (Trial Type: standard, novel, deviant) × 3 (Rule Pair: 75/25, 75/0, 25/0) ANOVA revealed main effects of Trial Type, $F(2, 26) = 23.216$, $p < .0001$, and Rule Pair, $F(2, 26) = 9.669$, $p < .001$. Participants were more likely to select the globally best rules in standard and novel trials than in deviant trials, and they were far more likely to select the globally best rules in trials with the 75/25 and 75/0 pairs than in trials with the 25/0 pair. The interaction between trial type and rule pair was also significant, $F(4, 52) = 21.652$, $p < .0001$. Participants were less likely to choose the 75 rule from the 75/25 pair for deviant trials as compared with standard trials, $t(13) = 5.702$, $p < .0001$, and novel trials, $t(13) = 5.590$, $p < .0001$, indicating that they detected that the 75 rule was less likely to be rewarded in deviant trials.

As in Experiment 1, we derived two indices to determine whether individuals responded differently to the deviant stimulus set. For trials with the 75/25 pair, we calculated the difference in preference for the 75 rule during standard and deviant trials. For trials with the 75/0 and 25/0 pairs, we calculated the difference in accuracy between 75/0 and 25/0 pairs and for standard and deviant trials. As seen in Figure 4, the measures were highly correlated, $r = .78$, $p < .01$. Thus, the strong interaction present in the choice data (Figure 3) arose from the subset of participants sensitive to deviant items.

ERP Results

We quantified the FRN as the difference between waveforms following losses and wins that were equally likely. Participants displayed an FRN for improbable outcomes (losses after 75 rule − wins after 25 rule) and probable outcomes (losses after 25 rule − wins after 75 rule) and for all trial types (Figure 5). A 3 (Trial Type: standard, novel, deviant) × 2 (Outcome Likelihood: probable, improbable) × 3 (Site: FCz, Cz, CPz) ANOVA of FRN amplitude revealed a significant effect of Outcome Likelihood, $F(1, 13) = 15.912$, $p < .01$, but not of Site, $F(2, 26) = .542$, $p > .1$, or Trial Type, $F(2, 26) = 1.642$, $p > .1$. The FRN was greater for improbable than for probable outcomes at all sites. The interaction between Trial Type and Outcome Likelihood approached significance, $F(2, 26) = 3.071$, $p < .1$.

We then focused on site FCz, where the FRN was maximal (Figure 6). A 3 (Trial Type) × 2 (Outcome Likelihood) ANOVA revealed a main effect of Outcome Likelihood, $F(1, 13) = 14.259$, $p < .01$, but not of Trial Type, $F(2, 26) = 1.781$, $p > .1$. Again, the interaction was marginally significant, $F(2, 26) = 3.301$, $p < .1$. Improbable outcomes yielded larger FRNs than probable outcomes for standard and novel trials, but not for deviant trials (Figure 7). It is not surprising that the interaction between trial type and outcome likelihood was weak because the aggregate data contained a mixture of unaware and aware participants, as we address next.

Model-based Analysis

To quantify the contributions of general and item-specific knowledge to choices, we fit the Combination model to each participant. The Combination model was implemented as in Experiment 1, with the following exception. Because sample and test numbers defined states in Experiment 1 and because these numbers did not repeat during novel trials, the Combination model could not apply item-specific knowledge to novel stimulus sets in Experiment 1. Because stimulus color defined states in Experiment 2 and because these colors did repeat during novel trials, the Combination model could apply item-specific knowledge to novel stimulus sets in Experiment 2. This knowledge was specific to an item's color but characterized abstract mathematical principles that were applicable to the general class of numerical stimuli. We presented the model with the history of choices and rewards that the participant experienced, and we identified the parameter values that maximized the log likelihood of the observed sequence of choices (α, τ, and W_{item}). Additionally, we tested whether the General ($W_{\text{item}} = 0$) or Item ($W_{\text{item}} = 1$) models outperformed the Combination model for each participant using the likelihood ratio test.

Model Results

Averaged over trials, the likelihood of the observed choices produced by the best fitting parameterization of the Combination model for each participant equaled $.74 \pm .02$ (coincidently, the same as in Experiment 1). Estimates for the weight parameter greatly exceeded zero for all but three participants (Figure 4). These same three participants exhibited the lowest scores for the two indices used to assess behavioral sensitivity to deviant items.

The Combination model outperformed the Item model for all but three participants ($p < .05$), and the Combination model outperformed the General model for all but three participants ($p < .05$). Thus, these contrasts revealed eleven participants best described by the Item or

Combination model (aware participants), and three best described by the General model (unaware participants). We compared parameter estimates for the Combination model between aware and unaware participants. The value of the weight parameter (W_{item}) was greater for aware participants, $t(12) = 4.529$, $p < .001$ (Table 2). No other parameters differed between groups.

We then separated the behavioral data according to participants' model-based classifications (Figure 3). The interaction between Trial Type and Rule Pair was significant in aware participants, $F(4, 40) = 42.502$, $p < .0001$, but not in unaware participants, $F(4, 8) = .520$, $p > .1$. Although aware participants exhibited sensitivity to deviant items, they did not completely reverse their decisions for the 75/25 pair in deviant trials. Response accuracy for the 75/25 pair, defined as the percentage of trials where participants selected the locally optimal rule, remained lower for deviant trials than for standard trials, $t(10) = 4.931$, $p < .001$, or novel trials, $t(10) = 4.798$, $p < .001$. These results describe aggregate responding over the course of the experiment. Sensitivity to deviant items gradually increased over the course of the experiment in aware participants, but unaware participants never exhibited any sensitivity to deviant items (Appendix).

Ancillary ERP Analysis

We divided the ERP data according to participants' model-based classifications and computed FRN amplitude at site FCz by trial type and outcome likelihood (Figure 7). Aware participants showed an interaction between Trial Type and Outcome Likelihood, $F(2, 20) = 9.995$, $p < .01$, but unaware participants did not, $F(2, 4) = 2.152$, $p > .1$.

Although the relationship between Awareness, Trial Type, and Outcome Likelihood was evident in both experiments, few participants fell into the aware subgroup in Experiment 1, and few fell into the unaware subgroup in Experiment 2. To overcome this limitation, we combined data from Experiments 1 and 2 and replicated the preceding analysis. The critical three-way interaction between Trial Type, Outcome Likelihood, and Awareness was significant, $F(2, 48) = 9.711$, $p < .001$, and was not influenced by experiment, $F(2, 48) = 1.661$, $p > .1$. The effect of Outcome Likelihood was modulated by Trial Type in aware participants, $F(2, 28) = 12.002$, $p < .001$, but not in unaware participants, $F(2, 24) = 1.504$, $p > .1$.

Even within subgroups, participants displayed a range of values for the weight parameter. As such, we examined whether individual values of the weight parameter related to neural sensitivity toward deviant items. To do so, we calculated the effect of outcome likelihood during standard and novel trials (improbable FRN − probable FRN) and subtracted from it the effect of outcome likelihood during deviant trials (improbable FRN − probable FRN).[5] Negative values of this index reflect neural sensitivity toward deviant items. The correlation between the weight parameter and neural sensitivity was significant within each experiment (Experiment 1, $r = -.50$, $p < .05$; Experiment 2, $r = -.70$, $p < .01$) and across the two experiments, $r = -.60$, $p < .001$. Neural sensitivity to deviant items increased with the value of the behavioral weight parameter (Figure 8).

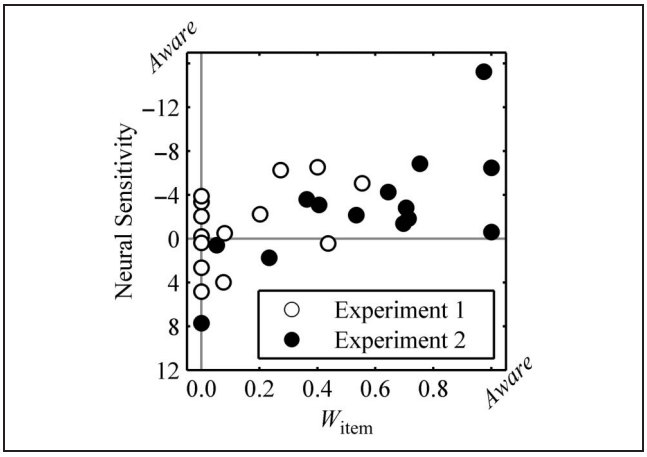

Figure 8. Scatter plot of individuals' neural sensitivity to deviant items (see text for details) and estimates for weight parameter in Combination model.

Discussion

The color grouping and instruction manipulations in Experiment 2 increased participants' sensitivity to the deviant stimulus set. All participants responded correctly to standard stimulus sets, all correctly generalized abstract rules to novel stimulus sets, and most avoided overgeneralizing abstract rules to the deviant stimulus set. The larger estimates for the weight parameter in the Combination model reflected participants' enhanced sensitivity to deviant items, as did the fact that the Item and Combination models matched participants much more closely than the General model did.

What effect, if any, did participants' heightened sensitivity to deviant items have on the FRN? As in Experiment 1, the amplitude of the FRN was greater for improbable outcomes than for probable outcomes and was comparable during standard and novel trials. The FRN nearly reversed for deviant items, however. This effect was especially apparent for aware participants in Experiments 1 and 2 and was completely absent for unaware participants. Moreover, the degree to which neural responses reversed for deviant items, as compared with standard and novel items, strongly related to individual differences in the behavioral weight parameter from the Combination model.

GENERAL DISCUSSION

The two main results of these experiments can be summarized quite simply. First, participants fluently applied

abstract rules to repeating and novel stimuli. Participants differed, however, in their ability to detect deviant items and to adjust their responses accordingly. Second, the FRN tracked behavioral performance. Although previous studies have demonstrated that the FRN is sensitive to the utility of concrete stimulus–response associations (for a review, see Walsh & Anderson, 2012), the current results show for the first time that the FRN is also sensitive to the utility of abstract response rules applied to novel stimuli. More strikingly, the FRN reflected item-specific reward probabilities in participants who detected the deviant set, and the FRN reflected overall reward probabilities in participants who did not. Thus, these results show that the representation of states and actions that guides behavior shapes the FRN.

Before conducting these experiments, it was unclear whether the FRN would track behavior. Some studies have found that only the FRN exhibits sensitivity to outcome likelihood in participants who display concurrent behavioral adaptation (Krigolson et al., 2009; Bellebaum & Daum, 2008). Yet other studies have reported a dissociation between the FRN and behavior. For example, Walsh and Anderson (2011b) recorded ERPs as participants performed a probabilistic choice task. Before the task, participants received a description of the reward probabilities associated with each response. Instruction eliminated participants' reliance on feedback as evidenced by their immediate asymptotic performance. In striking contrast, the FRN continued to adapt with experience. The FRN distinguished between probable and improbable outcomes only after participants had amassed significant practice. Taken together, the results from that study and the current experiments indicate that the representations of states and actions that guide behavior serve as a lens through which the FRN acquires information about reward probabilities from experience. These representations can be concrete, as in the case of earlier studies, or abstract, as in the case of the current experiments.

These results raise several additional questions. First, why did the FRN fail to reverse for deviant items in aware participants? Likewise, although aware participants came to favor the alternative rule for the deviant stimulus set, why did their item-specific accuracy remain lower in deviant trials than in standard and novel trials? Our model-based analysis revealed that the weight participants assigned to item-specific utilities (W_{item}) typically fell below one. In other words, even aware participants overgeneralized abstract rules to some extent. Our correlational analysis expanded upon this result by showing that overgeneralization of behavioral responses, as assessed by the value of the weight parameter, strongly related to overgeneralization in neural responses. Additionally, participants did not begin the experiments knowing the identity of the deviant stimulus set or the utilities of the abstract rules. As such, behavioral and neural responses to deviant items could reverse only in later trials.

Second, did participants actually calculate the utility of item-specific and general rules, and did they combine these estimates before responding in each trial? Our computational approach resembles Jacobs, Jordan, Nowlan, and Hinton's (1991) mixture of experts framework. The underlying notion is that complex problems can be decomposed into subparts that are more easily solvable. The mixture of experts framework divides problems into subparts and assigns each subpart to an expert module. A separate gating module controls the output and learning of expert modules.

Within this framework, the gating module can simultaneously assign weight to the outputs of multiple experts. Researchers have used such an approach to combine rule and exemplar knowledge in category learning (Erickson & Kruschke, 1998) and to combine utility estimates produced by different algorithms in reward learning tasks (Frank & Badre, 2012; Gläscher, Daw, Dayan, & O'Doherty, 2010). Alternatively, the gating module can select a single expert in a winner-take-all fashion. For example, in Nosofsky and Palmeri's (1998) RULEX model, individuals categorize stimuli using a rule process or exemplar memory. Likewise, in cognitive architectures such as ACT-R (Anderson, 2007), production rules that express procedural and declarative knowledge compete for expression. In these examples, the varying but singular processes evoked in individual trials can produce the appearance of averaging when viewed in aggregate.

In the Combination model, we assigned item-specific and general knowledge to separate experts, and we combined their outputs using a weight parameter. The Combination model might also approximate an architecture in which item-specific responses compete with general responses on a trial-by-trial basis. By this view, individual differences in sensitivity to deviant items arise from the different utility values participants initially assign to item-specific and general responses. For example, an individual that begins with a strong preference for general rules will rarely select their item-specific counterparts. As such, the individual will respond similarly to all stimulus sets. Alternatively, an individual that begins with a weak preference for general rules will instead rely on item-specific responses. As such, the individual will respond differently to the deviant stimulus set.

Third, did participants label stimuli as "greater than," "less than," or "equal to," and did the FRN track concrete associations between these subvocalized labels and responses? By this view, the sensitivity of the FRN to the utility of abstract rules was a byproduct of the intermediate stimulus representations that participants formed. The same question applies to all studies of rule use: Do abstract rules give rise to neural responses, or do concrete associations built upon internal, intermediate stimulus representations give rise to neural responses? To the extent that the individual can apply a label to an abstract rule, this possibility can never be rejected. The more basic message of this article, however, is that behavioral and

neural responses are not bound to physical features of the stimulus and that abstract rules (or the intermediate representations they produce) influence behavioral and neural responses in a complementary manner.

Fourth and finally, how did participants acquire the correct representation of states and actions in the first place? People exhibit striking sensitivity to base rates (Lovett & Anderson, 1996; Maddox, 1995; Reder, 1987; Friedman et al., 1964). This was the case in both experiments, where all participants learned which abstract rules were rewarded most frequently. Acquisition of base rate information often occurs in parallel with the more difficult task of identifying the stimulus features and primitive actions that define a task (Wilson & Niv, 2012; Lovett & Schunn, 1999; Newell & Simon, 1972). In machine learning, this entails projecting a high-dimensional space onto a tractable set of states and actions over which learning can occur (i.e., the curse of dimensionality; Sutton & Barto, 1998).

Lovett and Schunn (1999) presented a theoretical framework for understanding how changes in mental representation influence base rate sensitivity. Most relevant to the current work is the first stage in their process model, representing the task. Lovett and Schunn proposed that people combine salient features of the task with prior knowledge to form an initial representation of states and actions. On the basis of these ideas, we presented standard, novel, and deviant stimulus sets in distinctive font colors during Experiment 2. We reasoned that font color, a salient feature, would allow participants to partition the different trial types into separate groups. Additionally, we instructed participants to pay attention to the font color, and we informed them that the reward probabilities differed for one color. We reasoned that instruction, a source of prior knowledge, would permit participants to use font color to form a multistate representation. These manipulations had the desired effect of engendering greater sensitivity to the base rates of success for deviant items.

Verbal instruction allowed participants to adopt the correct representation of the task immediately. However, verbal instruction is not necessary to discover abstract response rules. For example, monkeys can learn to select test pictures that match a sample picture (Wallis et al., 2001). Such learning takes time (i.e., tens of thousands of trials in the case of monkeys; Wallis & Miller, 2003). Although humans could presumably discover the same rules more quickly, the challenge is not trivial. In fact, we withheld instruction from participants in earlier versions of these experiments. Asymptotic behavior was identical to the behavior of participants reported here, but people varied greatly in how long they took to discover the abstract mathematical rules. The question of how people discover such rules is interesting in its own right and warrants further investigation (Wilson & Niv, 2012).

These questions aside, the current experiments clearly demonstrate that behavior and the FRN are not bound by concrete stimulus–response learning. In addition, these experiments demonstrate that the representation of states and actions that guides behavior shapes the FRN. Thus, these results advance our understanding of the FRN and strengthen the connection between the FRN and behavioral control.

APPENDIX

To what extent did participants' behavior change over the course of the experiment, and did the Combination model exhibit corresponding change? To address these questions, we divided each experiment into five blocks of 225 trials and computed choice data separately for each block. Because the best response differs between deviant stimulus sets and standard and novel stimulus sets in trials with the 75/25 pair, these trials provide the strongest test of whether participants were sensitive to deviant items. As such, we restricted our analyses to these trials. We performed 5 (Trial Block) × 3 (Trial Type) ANOVAs separately for aware and unaware participants in Experiments 1 and 2.

In Experiment 1, unaware participants exhibited a main effect of Block only, $F(4, 32) = 4.245$, $p < .01$. Aware participants exhibited a main effect of Trial Type only, $F(2, 8) = 6.994$, $p < .05$. Aware participants displayed moderate sensitivity to deviant items beginning around the second 225-trial block, yet they never favored the alternative response above chance in deviant trials. In Experiment 2, unaware participants did not exhibit any significant effects, likely owing to low statistical power. In contrast, aware participants exhibited a main effect of Trial Type, $F(2, 20) = 60.057$, $p < .001$, and a significant interaction between Block and Trial Type, $F(8, 80) = 20.028$, $p < .001$. Aware participants displayed moderate sensitivity to deviant items beginning around the second 225-trial block, and they increasingly favored the alternative response in deviant trials.

The Combination model captured these trends, including the transition from above- to below-chance selection of the globally optimal rule in deviant trials during Experiment 2. General rules in the Combination model are updated during every trial, whereas item-specific rules are updated only during trials when the corresponding items appear. Because of these disproportionate learning opportunities, general rules dominate performance initially, producing overgeneralization. As people amass experience with individual stimulus sets, however, item-specific rules increasingly contribute to performance.

These results indicate that the FRN for deviant items should reverse in aware participants following extended training. In aware participants of Experiment 1, FRN amplitude was equivalent for improbable and probable outcomes over the first half of trials (−2.79 μV vs. −2.78 μV) and reversed over the second half of trials (−0.87 μV vs. −2.28 μV). Likewise, in aware participants of Experiment 2, FRN amplitude was greater for

improbable than for probable outcomes over the first half of trials (−3.71 μV vs. −2.06 μV) and reversed over the second half of trials (−3.36 μV vs. −4.46 μV). Although suggestive, these results did not reach significance, likely because so few observations contributed to the averages of improbable events.

Acknowledgments

This work was supported by National Institute of Mental Health training grant T32MH019983 to Matthew Walsh and National Institute of Mental Health grant MH068243 to John Anderson.

Reprint requests should be sent to Matthew M. Walsh, Air Force Research Laboratory, 711 HPW/RHAC – Cognitive Models and Agents Branch, 2620 Q Street, Building 852, Wright-Patterson AFB, OH 45433, or via e-mail: mmwl88@gmail.com.

Notes

1. By correct, we mean the response that was more likely to be rewarded.
2. One could also present configurations of sample and test numbers that appeared only once and that deviated from the overall rule probabilities; however, learning to respond correctly to deviant items that appeared only once would be impossible.
3. Equal to was never the globally optimal or suboptimal rule. Such assignments would create scenarios where participants could ignore the sample number and maximize reward by always selecting the larger of the test numbers or by always selecting the smaller of the test numbers.
4. Color assignment was counterbalanced across participants. To equate color frequencies, half of the novel stimulus sets appeared in one color, and half appeared in another color.
5. This essentially measures the strength of the interaction between trial type and outcome likelihood.

REFERENCES

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York: Oxford University Press.

Asaad, W. F., Rainer, G., & Miller, E. K. (2000). Task-specific neural activity in the primate prefrontal cortex. *Journal of Neurophysiology, 84,* 451–459.

Ashby, F. G., & O'Brien, J. B. (2005). Category learning and multiple memory systems. *Trends in Cognitive Sciences, 9,* 83–89.

Badre, D., & D'Esposito, M. (2007). Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *Journal of Cognitive Neuroscience, 19,* 2082–2099.

Badre, D., & D'Esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical? *Nature Reviews Neuroscience, 10,* 659–669.

Baxter, M. G., Gaffan, D., Kyriazis, D. A., & Mitchell, A. S. (2009). Ventrolateral prefrontal cortex is required for performance of a strategy implementation task but not reinforcer devaluation effects in rhesus monkeys. *European Journal of Neuroscience, 29,* 2049–2059.

Bellebaum, C., & Daum, I. (2008). Learning-related changes in reward expectancy are reflected in the feedback-related negativity. *European Journal of Neuroscience, 27,* 1823–1835.

Berridge, K. C. (2007). The debate over dopamine's role in reward: The case for incentive salience. *Psychopharmacology, 191,* 391–431.

Bongard, S., & Nieder, A. (2010). Basic mathematical rules are encoded by primate prefrontal cortex neurons. *Proceedings of the National Academy of Sciences, U.S.A., 107,* 2277–2282.

Bunge, S. A. (2004). How we use rules to select actions: A review of evidence from cognitive neuroscience. *Cognitive, Affective, & Behavioral Neuroscience, 4,* 564–579.

Bunge, S. A., Kahn, I., Wallis, J. D., Miller, E. K., & Wagner, A. D. (2003). Neural circuits subserving the retrieval and maintenance of abstract rules. *Journal of Neurophysiology, 90,* 3419–3428.

Bunge, S. A., & Wallis, J. D. (Eds.) (2008). *Neuroscience of rule-guided behavior*. New York: Oxford University Press.

Bunge, S. A., Wendelken, C., Badre, D., & Wagner, A. D. (2005). Analogical reasoning and prefrontal cortex: Evidence for separable retrieval and integration mechanisms. *Cerebral Cortex, 15,* 239–249.

Bussey, T. J., Wise, S. P., & Murray, E. A. (2001). The role of ventral and orbital prefrontal cortex in conditional visuomotor learning and strategy use in rhesus monkeys (*Macaca mulatta*). *Behavioral Neuroscience, 115,* 971–982.

Christoff, K., & Gabrieli, J. D. E. (2000). The frontopolar cortex and human cognition: Evidence for a rostrocaudal hierarchical organization within the human prefrontal cortex. *Psychobiology, 28,* 168–186.

Christoff, K., Ream, J. M., Geddes, L. P. T., & Gabrieli, J. D. E. (2003). Evaluating self-generated information: Anterior prefrontal contributions to human cognition. *Behavioral Neuroscience, 117,* 1161–1168.

Cohen, M. X., & Ranganath, C. (2007). Reinforcement learning signals predict future decisions. *Journal of Neuroscience, 27,* 371–378.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134,* 9–21.

Dias, R., Robbins, T. W., & Roberts, A. C. (1996). Dissociation in prefrontal cortex of affective and attentional shifts. *Nature, 380,* 69–72.

Duncan-Johnson, C. C., & Donchin, E. (1977). On quantifying surprise: The variation of event-related potentials with subjective probability. *Psychophysiology, 14,* 456–467.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General, 127,* 107–140.

Frank, M. J., & Badre, D. (2012). Mechanisms of hierarchical reinforcement learning in corticostriatal circuits 1: Computational analysis. *Cerebral Cortex, 22,* 509–526.

Friedman, M. P., Burke, C. J., Cole, M., Keller, L., Millward, R. B., & Estes, W. K. (1964). Two-choice behavior under extended training with shifting probabilities of reinforcement. In R. C. Atkinson (Ed.), *Studies in mathematical psychology* (pp. 250–316). Stanford, CA: Stanford University Press.

Fu, W. T., & Anderson, J. R. (2006). From recurrent choice to skill learning: A reinforcement-learning model. *Journal of Experimental Psychology: General, 135,* 184–206.

Gaffan, D., Easton, A., & Parker, A. (2002). Interaction of inferior temporal cortex with frontal cortex and basal forebrain: Double dissociation in strategy implementation and associative learning. *Journal of Neuroscience, 22,* 7288–7296.

Gehring, W. J., & Willoughby, A. R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science, 295,* 2279–2282.

Genovesio, A., Brasted, P. J., Mitz, A. R., & Wise, S. P. (2005). Prefrontal cortex activity related to abstract response strategies. *Neuron, 47,* 307–320.

Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron, 66,* 585–595.

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review, 109,* 679–709.

Holroyd, C. B., Krigolson, O. E., Baker, R., Lee, S., & Gibson, J. (2009). When is an error not a prediction error? An electrophysiological investigation. *Cognitive, Affective, & Behavioral Neuroscience, 9,* 59–70.

Holroyd, C. B., Nieuwenhuis, S., Yeung, N., & Cohen, J. D. (2003). Errors in reward prediction are reflected in the event-related brain potential. *NeuroReport, 14,* 2481–2484.

Holroyd, C. B., Nieuwenhuis, S., Yeung, N., Nystrom, L., Mars, R. B., Coles, M. G. H., et al. (2004). Dorsal anterior cingulate cortex shows fMRI response to internal and external error signals. *Nature Neuroscience, 7,* 497–498.

Holroyd, C. B., Pakzad-Vaezi, K. L., & Krigolson, O. E. (2008). The feedback correct-related positivity: Sensitivity of the event-related brain potential to unexpected positive feedback. *Psychophysiology, 45,* 688–697.

Ito, S., Stuphorn, V., Brown, J. W., & Schall, J. D. (2003). Performance monitoring by the anterior cingulate cortex during saccade countermanding. *Science, 302,* 120–122.

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). Adaptive mixtures of local experts. *Neural Computation, 3,* 79–87.

Koechlin, E., & Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. *Trends in Cognitive Science, 11,* 229–235.

Krigolson, O. E., Pierce, L. J., Holroyd, C. B., & Tanaka, J. W. (2009). Learning to become an expert: Reinforcement learning and the acquisition of perceptual expertise. *Journal of Cognitive Neuroscience, 21,* 1833–1840.

Kroger, J. K., Sabb, F. W., Fales, C. L., Bookheimer, S. Y., Cohen, M. S., & Holyoak, K. J. (2002). Recruitment of anterior dorsolateral prefrontal cortex in human reasoning: A parametric study of relational complexity. *Cerebral Cortex, 5,* 477–485.

Lewandowsky, S., & Farrell, S. (2011). *Computational modeling in cognition: Principles and practice*. Thousand Oaks, CA: Sage.

Lovett, M. C., & Anderson, J. R. (1996). History of success and current context in problem solving. Combined influences on operator selection. *Cognitive Psychology, 31,* 168–217.

Lovett, M. C., & Schunn, C. D. (1999). Task representations, strategy variability, and base-rate neglect. *Journal of Experimental Psychology: General, 128,* 107–130.

Luce, R. D. (1977). The choice axiom after twenty years. *Journal of Mathematical Psychology, 15,* 215–233.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.

Maddox, W. T. (1995). Base-rate effects in multidimensional perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 288–301.

Mansouri, F. A., Matsumoto, K., & Tanaka, K. (2006). Prefrontal cell activities related to monkeys' success and failure in adapting to rule changes in a Wisconsin card sorting test analog. *Journal of Neuroscience, 26,* 2745–2756.

McClure, S. M., York, M. K., & Montague, P. R. (2004). The neural substrates of reward processing in humans: The modern role of fMRI. *Neuroscientist, 10,* 260–268.

Milner, B. (1963). Effects of different brain lesions on card sorting: The role of the frontal lobes. *Archives of Neurology, 9,* 90–100.

Miltner, W. H. R., Braun, C. H., & Coles, M. G. H. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a "generic" neural system for error detection. *Journal of Cognitive Neuroscience, 9,* 788–798.

Muhammad, R., Wallis, J. D., & Miller, E. K. (2006). A comparison of abstract rules in the prefrontal cortex, premotor cortex, inferior temporal cortex, and striatum. *Journal of Cognitive Neuroscience, 18,* 974–989.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal, 7,* 308–313.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.

Niki, H., & Watanabe, M. (1979). Prefrontal and cingulate unit activity during timing behavior in the monkey. *Brain Research, 171,* 213–224.

Nosofsky, R. M., & Palmeri, T. J. (1998). A rule-plus-exception model for classifying objects in continuous-dimension spaces. *Psychonomic Bulletin & Review, 5,* 345–369.

O'Doherty, J. P. (2004). Reward representations and reward-related learning in the human brain: Insights from neuroimaging. *Current Opinion in Neurobiology, 14,* 769–776.

Packard, M. G., & Knowlton, B. J. (2002). Learning and memory functions of the basal ganglia. *Annual Review of Neuroscience, 25,* 563–593.

Passingham, R. E. (1993). *The frontal lobes and voluntary action*. New York: Oxford University Press.

Picard, N., & Strick, P. L. (2001). Imaging the premotor areas. *Current Opinion in Neurobiology, 11,* 663–672.

Ramnani, N., & Owen, A. M. (2004). Anterior prefrontal cortex: Insights into function from anatomy and neuroimaging. *Nature Reviews Neuroscience, 5,* 184–194.

Reder, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology, 19,* 90–138.

Schneider, D. W., & Logan, G. D. (2009). Task switching. In L. R. Squire (Ed.), *Encyclopedia of neuroscience* (Vol. 9, pp. 869–874). Oxford, UK: Academic Press.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology, 80,* 1–27.

Shallice, T., & Burgess, P. W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain, 114,* 727–741.

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

Thorndike, E. L. (1911). *Animal intelligence: Experimental studies*. New York: Macmillan.

Wallis, J. D., Anderson, K. C., & Miller, E. K. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature, 411,* 953–956.

Wallis, J. D., & Miller, E. K. (2003). From rule to response: Neuronal processes in the premotor and prefrontal cortex. *Journal of Neurophysiology, 90,* 1790–1806.

Walsh, M. M., & Anderson, J. R. (2011a). Modulation of the feedback-related negativity by instruction and experience. *Proceedings of the National Academy of Sciences, U.S.A., 108,* 19048–19053.

Walsh, M. M., & Anderson, J. R. (2011b). Learning from delayed feedback: Neural responses in temporal credit assignment. *Cognitive, Affective, & Behavioral Neuroscience, 11,* 131–143.
Walsh, M. M., & Anderson, J. R. (2012). Learning from experience: Event-related potential correlates of reward processing, neural adaptation, and behavioral choice. *Neuroscience and Biobehavioral Reviews, 36,* 1870–1884.
Walsh, M. M., & Anderson, J. R. (in press). Navigating complex decision spaces: Problems and paradigms in sequential choice. *Psychological Bulletin*.
White, I. M., & Wise, S. P. (1999). Rule-dependent neuronal activity in the prefrontal cortex. *Experimental Brain Research, 126,* 315–335.
Wilson, R. C., & Niv, Y. (2012). Inferring relevance in a changing world. *Frontiers in Human Neuroscience, 5,* 1–14.
Yamada, M., Pita, M. C. R., Iijima, T., & Tsutsui, K. I. (2010). Rule-dependent anticipatory activity in prefrontal neurons. *Neuroscience Research, 67,* 162–171.