

# Foraging Value, Risk Avoidance, and Multiple Control Signals: How the Anterior Cingulate Cortex Controls Value-based Decision-making

Joshua W. Brown<sup>1</sup> and William H. Alexander<sup>2</sup>

## Abstract

Recent work on the role of the ACC in cognition has focused on choice difficulty, action value, risk avoidance, conflict resolution, and the value of exerting control among other factors. A main underlying question is what are the output signals of ACC, and relatedly, what is their effect on downstream cognitive processes? Here we propose a model of how ACC influences cognitive processing in other brain regions that choose actions. The model builds on the

earlier Predicted Response Outcome model and suggests that ACC learns to represent specifically the states in which the potential costs or risks of an action are high, on both short and long time-scales. It then uses those cost signals as a basis to bias decisions to minimize losses while maximizing gains. The model simulates both proactive and reactive control signals and accounts for a variety of empirical findings regarding value-based decision-making. ■

## INTRODUCTION

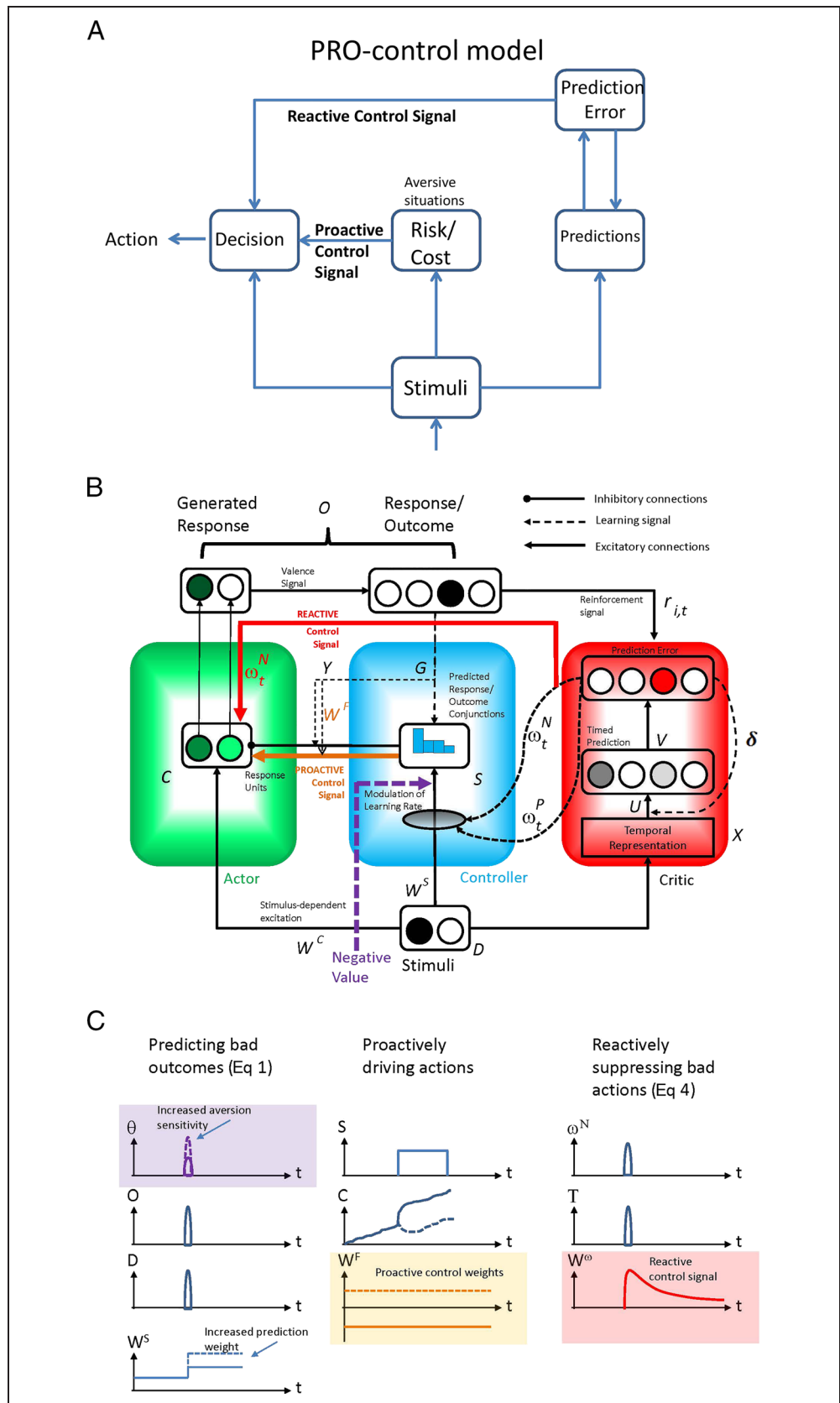
The ACC and surrounding medial pFC (mPFC) regions have been the subject of interest and debate, given that they are active in many tasks and also key to a number of clinical disorders (Alexander & Brown, 2011). ACC has been shown to be active in response to errors, conflict, and the likelihood of an error among other effects (Brown & Braver, 2005; Carter et al., 1998; Gehring, Goss, Coles, Meyer, & Donchin, 1993; Gemba, Sasaki, & Brooks, 1986). The role of ACC in decision-making has more recently been a focus in the literature. The earliest neuropsychological studies of ACC noted akinetic mutism as a result of ACC damage (reviewed in Devinsky, Morrell, & Vogt, 1995). Subjects typically lacked sensitivity to noxious stimuli and did not initiate movements or speech. Damage to ACC in animals leads to a lack of willingness to exert more effort to obtain a larger reward (Walton et al., 2009), and in humans, damage to areas including ACC can lead to failures of internally guided action (Amiez, Sophie, Charles, Procyk, & Petrides, 2015). Obsessive compulsive disorder prominently entails overactivity of ACC (Fitzgerald et al., 2005). Patients with obsessive compulsive disorder typically perceive risks and dangers as disproportionately large, so that they take unnecessarily effortful precautions to avoid perceived dangers. A recent neurosurgery study provides a fascinating insight to the role of ACC as providing the “will to persevere.” Patients undergoing electrical stimulation around dorsal ACC (Area 24) described a resulting sensation as

“a positive thing like...push harder, push harder, push harder to try to get through this” (Parvizi, Rangarajan, Shirer, Desai, & Greicius, 2013).

In humans, ACC activity has been associated with an inclination to search for better rewards (Kolling, Behrens, Mars, & Rushworth, 2012), to avoid risk (Fukunaga, Brown, & Bogg, 2012; Krawitz, Fukunaga, & Brown, 2010; Brown & Braver, 2007, 2008), and to make more normative decisions (Paulus & Frank, 2006). These findings collectively suggest that ACC provides a motivational signal, controlling decisions to pursue more rewarding options and avoid losses, even when more effort is required (Shenhav, Botvinick, & Cohen, 2013; Croxson, Walton, O’Reilly, Behrens, & Rushworth, 2009; Walton et al., 2009). This impetus provided to effortful behavior may act at the level of energizing particular overall strategies rather than individual actions (Holroyd & McClure, 2015). Still, debate has persisted over what information is actually represented in ACC, with various proposals of conflict (Botvinick, Braver, Barch, Carter, & Cohen, 2001), error likelihood (Brown & Braver, 2005; Holroyd, Yeung, Coles, & Cohen, 2005), surprise (Ferdinand, Mecklinger, Kray, & Gehring, 2012; Wessel, Danielmeier, Morton, & Ullsperger, 2012; Alexander & Brown, 2011), difficulty (Shenhav, Straccia, Cohen, & Botvinick, 2014; Shenhav et al., 2013), and the value of decision options (Shenhav et al., 2014; Kolling et al., 2012; Rushworth, Kolling, Sallet, & Mars, 2012; Hayden, Pearson, & Platt, 2011; Kennerley, Behrens, & Wallis, 2011; Walton et al., 2009; Rudebeck et al., 2008) and decision costs (Skvortsova, Palminteri, & Pessiglione, 2014). Despite a multiplication of empirical findings, the mechanisms by which ACC controls decision-making

<sup>1</sup>Indiana University, <sup>2</sup>Ghent University

**Figure 1.** The PRO-control model. (A) Overview of PRO-control model, showing proactive and reactive pathways. (B) Detailed view of new model. It consists of the PRO model (Alexander & Brown, 2011) but with three modifications highlighted as purple, orange, and red pathways. First, the Controller R-O conjunctions (blue box) are trained to respond to stimuli in proportion  $\theta$  (purple arrow) to how much the value of the current event falls below the expected value. In the original model, the R-O conjunctions were trained to reflect the probability of the corresponding outcome. Second, the proactive control signal from the Controller (blue box) to the response units (green box) now include learned but weak excitatory signals (orange arrows), in addition to the learned inhibitory control signals as in the original PRO model. The excitatory signals allow the controller to weakly activate specific responses. Third, the negative surprise signals  $\omega^N$  now provide a reactive control signal (red arrows) that rapidly and temporarily inhibits a response that leads to an undesirable outcome. These three modifications preserve the ability of the PRO model to simulate a wide range of ACC effects as shown previously (Alexander & Brown, 2011) and now add the ability to simulate the role of ACC in value-based decision making. (C) Graphical depiction of a representative time course of activities and synaptic strengths in the new mechanisms. For predicting bad outcomes (left column), greater  $\theta$  representing stronger avoidance learning leads to corresponding stronger increases in outcome predictions  $O$  by strengthening the weights  $W^S$ . For proactively driving actions (middle column), stimuli  $S$  excite or inhibit responses  $C$ . For reactively suppressing actions (right column), specific negative surprise representations  $\omega^N$  learn to generate a short acting specific activation (or suppression)  $W^{\omega}$  when they precede a rewarding (or aversive) outcome.



remain unclear, in part because there is presently no single computational model that can account for already observed effects. Without such a comprehensive model, additional empirical results may be less likely to resolve the issue efficiently (Brown, 2014).

In this article, we propose a computational neural model, the PRO-control model (Figure 1), aimed at unifying a body of empirical findings about how ACC influences decisions. Although there are many existing computational models (Alexander & Brown, 2010, 2011; Brown & Braver, 2005; Botvinick et al., 2001), we begin with the Predicted Response Outcome (PRO) model. The PRO model has been shown to account for a wide range of effects found within ACC, from fMRI, ERP, and monkey neurophysiology (Alexander & Brown, 2011). In particular, the PRO model can simulate effects of conflict (Botvinick et al., 2001), error likelihood (Brown & Braver, 2005; Holroyd et al., 2005), surprise (Garofalo, Maier, & di Pellegrino, 2014; Ferdinand et al., 2012; Wessel et al., 2012; Alexander & Brown, 2011; Jessup, Busemeyer, &

Brown, 2010; Oliveira, McDonald, & Goodman, 2007), ERP effects (Yeung & Nieuwenhuis, 2009; Holroyd et al., 2005; Holroyd & Coles, 2002), and individual differences in risk sensitivity (Brown & Braver, 2007). Recent extensions to the PRO model suggest a more general role for ACC in predicting salient events (Alexander & Brown, 2014) as well as how ACC may interact with regions outside the cingulate involved in cognitive control such as dorsolateral pFC (Alexander & Brown, 2015). Table 1 provides a nonexhaustive summary of effects simulated by the PRO model.

Nevertheless, the PRO model as published did not address in as much depth the question of how ACC activity influences decision-making, beyond providing learned inhibitory control over actions and modulating the learning rate of such control. In what follows, we present a modified version of the PRO model, which we call the PRO-control model. The new model retains the PRO model's ability to simulate a wide range of effects within ACC, and it now adds the ability to simulate a number of key effects

**Table 1.** Empirical Effects Simulated by the PRO Model

<i>Effect</i>	<i>Reference</i>	<i>Method</i>	<i>Mechanism</i>
Informative vs. uninformative cues	Aarts et al., 2008	fMRI	Negative surprise
Reward salience in substance use	Alexander et al., 2015	fMRI	Outcome prediction
Reward prediction	Amador et al., 2000	Single-unit monkey	Outcome prediction
Monitoring others' outcomes	Apps et al., 2012	fMRI	Negative surprise
Environmental volatility	Behrens et al., 2007	fMRI	Negative surprise
Global vs. local conflict	Blais & Bunge, 2010	fMRI	Outcome prediction
Conflict	Botvinick et al., 2001	fMRI	Outcome prediction
Error likelihood	Brown & Braver, 2005	fMRI	Outcome prediction
Unexpected error	Brown & Braver, 2005	fMRI	Negative surprise
Multiple responses	Brown, 2009	fMRI	Negative surprise
Attention signaling	Bryden et al., 2011	Single-unit rat	Negative surprise
Trial frequency	Carter et al., 2000	fMRI	Outcome prediction
Mismatch negativity	Crottaz-Herbette & Menon, 2006	EEG/fMRI	Negative surprise
Feedback delay	Forster & Brown, 2011	fMRI	Negative surprise
Surprising absence of pain	Garofalo et al., 2014	EEG	Negative surprise
Error effect	Gehring et al., 1993	EEG	Negative surprise
Time on task	Grinband et al., 2011	EEG	Negative surprise
Bayesian surprise	Ide et al., 2013	fMRI	Negative surprise
Multiple outcome predictions	Jahn et al., 2014	fMRI	Outcome prediction
Unexpected correct	Jessup et al., 2010	fMRI	Negative surprise
Stimulus prediction	Koyama et al., 2001	Single-unit monkey	Outcome prediction
Changes in predicted reward	Sallet et al., 2007	Single-unit monkey	Negative surprise
Speed-accuracy tradeoff	Yeung & Nieuwenhuis, 2009	EEG	Outcome prediction

by which ACC activity influences value-based decision-making. The ability to simulate a wider range of data derives from two key mechanisms of the new model, namely one for proactive control and one for reactive control (Figure 1A).

The first new mechanism of the PRO-control model affords a proactive control signal (Braver, Gray, & Burgess, 2007). The proactive component begins with ACC learning to signal situations in which the responses and outcomes are likely to be aversive, which entails lower overall value (Figure 1A). This occurs in various situations involving a probability of risk, costs, loss, or effort expended (Fukunaga et al., 2012; Brown & Braver, 2005, 2007). As these representations are learned, a subsequent pathway provides a way for these aversive predictions to influence decisions, guiding responses to minimize the possibility of an aversive outcome while maximizing rewards.

The second principle of the PRO-control model is that ACC provides reactive control. Specifically, prediction errors indicate a surprising outcome, and these in turn provide a basis for driving corrective actions by suppressing the action that led to the prediction error. This is not to say that prediction errors always mean that something bad has happened. Rather, only a subset of prediction errors indicate that something worse than expected has occurred or, more subtly, that some desirable outcome has failed to occur. The subsets of both “bad” and “good” surprise signals form the basis for reactively suppressing the actions that just led to the undesirable outcome while reactively increasing the probability of actions that recently led to rewarding outcomes. These signals act on a time-scale of a few trials, considerably shorter than some longer-lasting reinforcement learning effects (Schultz, Dayan, & Montague, 1997; Barto, Sutton, & Anderson, 1983). Collectively, these proactive and reactive control signals provide both rapid and lasting control over decisions, so that decisions maximize reward while minimizing unacceptable levels of aversive outcomes. In what follows, we describe how these mechanisms work in greater detail.

## METHODS

### The Original PRO Model

The PRO-control model builds on the earlier PRO model (Alexander & Brown, 2011), which consisted of several components—an Actor, a Critic, and a Controller (Figure 1B). The Actor implements stimulus response (S-R) mapping. The Controller learns to predict conjunctions of responses and outcomes (hence, the model name PRO for predictions of responses and outcomes) on the basis of incoming stimuli. These predicted R-O conjunctions form the basis for a control signal, which is trained to inhibit responses that lead to undesirable outcomes. The Critic resembles a temporal difference

model, but with some important differences as described previously (Alexander & Brown, 2011). The Critic’s Prediction layer forms timed predictions of R-O conjunctions, and the Critic’s Prediction Error layer compares the predicted outcomes against the actual outcomes. The prediction errors are then split into two components: The  $\omega^P$  component is the positive rectified prediction error: It is greater than zero when an outcome was not predicted to occur, but it occurred anyway. The  $\omega^P$  signal is referred to as positive surprise. Actual outcomes with a lower prior probability of occurrence yield greater  $\omega^P$ . The  $\omega^N$  component is the negative rectified prediction error: It is greater than zero when an outcome was predicted to occur, but it failed to occur. Outcomes with a larger prior probability of occurrence (and which nevertheless fail to occur) yield a larger  $\omega^N$  signal. It is important to note that positive surprise  $\omega^P$  refers to a surprising occurrence, whereas negative surprise  $\omega^N$  refers to surprise by omission. Thus positive and negative surprises are independent of valence—each can be either affectively good or bad. The  $\omega^N$  (or negative surprise) signal is crucial—in the PRO model simulations, it simulated the vast majority of empirical effects found in ACC (Alexander & Brown, 2011). In addition, negative surprise provides a useful signal—it is relatively easy to detect a surprising occurrence, but it is harder to detect a surprising non-occurrence, for example, an expected payment that fails to arrive.

### The New PRO-control Model

The new model consists of the PRO model but with three changes, highlighted as red pathways in Figure 1B. A description of the original PRO model is found in Alexander and Brown (2011). All simulation parameters used here are the same as in the original PRO model.

#### *Predicting Bad Outcomes*

First, the Controller R-O representations are now trained to represent the predicted R-O conjunction in proportion to how bad such an outcome would be, as in the error likelihood model (Brown & Braver, 2005). Here bad refers to a value below the expected value of outcomes over a longer term (Kolling et al., 2012; Charnov, 1976). In the original PRO model, the Controller R-O conjunction activities simply represented the probability of the R-O conjunction occurring. Now, the Controller R-O conjunction activity reflects approximately the product of how likely an R-O conjunction occurrence is and how bad such an occurrence would be (in terms of lost resources, cost, effort, etc.) relative to a baseline, for example, the average reward. This in turn means that the Controller R-O representations will be active in situations in which bad outcomes are likely to occur. With such a basis, it is possible to detect situations with potentially costly or

undesirable outcomes. The next change to the PRO model leverages this representation.

The modification is implemented (Figure 1C, left) by changing Equation 4 of Alexander and Brown (2011) to insert the value  $\theta$  as a modulator on the learning law:

$$W_{ij,t+1}^S = W_{ij,t}^S + A_{i,t}(\theta O_{i,t} - S_{i,t})G_t D_j \quad (1)$$

where  $\theta$  is a parameter that represents how subjectively bad a given outcome is, either because of individual differences in sensitivity to punishment (Brown & Braver, 2007, 2008; see Simulation 2), or because the outcome has a more negative value relative to the expected value (see Simulation 1), or some combination of those factors. The variable  $O$  represents the active R-O representation to be learned by the controller, and the weights  $W^S$  represent the weights from the Stimuli to the Controller R-O conjunctions. The remaining terms of Equation 1 are described in Equation 4 of the original PRO model (Alexander & Brown, 2011).

### Proactively Driving Good Actions

In the original PRO model, the Controller R-O conjunctions could be trained to inhibit responses that would lead to undesirable outcomes. This can simulate avoidance behavior and is consistent with the known anatomy of ACC to lateral pFC connections, which are mostly but not exclusively inhibitory (Medalla & Barbas, 2009). In the PRO-control model, the projections from the Controller R-O conjunction units to the responses can now include excitatory connections as well. This means that representations of costly or undesirable outcomes now can specifically activate more appropriate actions, not just suppress inappropriate ones. This captures the spirit of the patient report that ACC stimulation led to a “positive” feeling of “push[ing] harder to try to get through this” (Parvizi et al., 2013). In the PRO-control model, the change is implemented as follows. In the original PRO model, in Equation 16 of Alexander and Brown (2011), the inhibitory weights  $W^F$  were constrained to be not less than zero. In the new model, the inhibitory weights are allowed to be less than zero, which can lead to net excitation of the response units (Figure 1C, middle). Any net excitation is added to Equation 11 of Alexander and Brown (2011), where it is scaled along with excitatory S-R inputs, so that the original Equation 12 becomes:

$$E_{i,t} = \rho \left( \sum_j D_j W_{ij}^C + \sum_k [-S_k W_{ik}^F]^+ + \sum_j [-W_{ij,t}^{\omega}]^+ \right) \quad (2)$$

where the  $W^F$  term provides proactive control and the  $W^{\omega}$  term provides reactive control (described below). The  $[x]^+$  notation rectifies the argument to  $\max(0, x)$ . Equation 13 of Alexander and Brown (2011) is likewise modified to rectify the proactive inhibitory control input

and reactive control signals  $W^{\omega}$ , described in Equation 4 below:

$$I_{i,t} = \psi \left( \sum_j C_j W_{ij}^l \right) + \Phi \left( \sum_k [S_k W_{ik}^F]^+ + \sum_j [W_{ij,t}^{\omega}]^+ \right) \quad (3)$$

where all of the terms are described in Equation 13 of Alexander and Brown (2011). Also, the weights  $W^F$  are normalized such that the sum of the absolute values of the weights  $W^F$  does not exceed the number of R-O conjunctions represented. The control signals from the negative surprise units to the response units via weights  $W^{\omega}$  (described below) are assumed to be driven by an activation of unity, which persists briefly, that is, as long as the short-term weights  $W^{\omega}$  are greater than zero. This is a simplifying assumption for computational convenience, though it is realistic, provided that the negative surprise unit activities do not decay too quickly. There is evidence of sustained activity in monkey single units that may represent negative surprise (Shima & Tanji, 1998).

### Reactively Suppressing Bad Actions

In the original PRO model, the negative surprise  $\omega^N$  had only a small and indirect influence on behavior—it modulated the Controller R-O learning rate. Although this allowed the model to capture variable learning rate effects (Behrens, Woolrich, Walton, & Rushworth, 2007), it left unanswered the question of why such a rich representation of specific negative surprise exists if only to be condensed into a scalar that modulates the learning rate. In the PRO-control model, there is a new and richer projection from the negative surprise signals to the response units. When a bad outcome occurs, these projections are specifically and rapidly strengthened to inhibit the offending response (Figure 1C, right). The inhibitory weights (and the active  $\omega^N$  signal) both decay relatively quickly over the course of the next few trials, but in the meantime, they rapidly and powerfully suppress the offending action. Specifically, the new connection weights  $W^{\omega}$  from negative surprise signal  $j$  to action representation  $i$  are

$$W_{ij,t+1}^{\omega} = 0.25 W_{ij,t}^{\omega} + Y_i T_{i,t} \omega_{j,t}^N \quad (4)$$

where  $T$  is 1 if action  $i$  was executed and 0 otherwise, and  $Y$  is a valence signal corresponding to correct or error outcomes, as described in Equation 14 of Alexander and Brown (2011). Individual elements of the weights  $W^{\omega}$  are constrained to the range  $[-1, 1]$ .

The importance of these three changes is illustrated in the following simulations.

### Simulation 1. ACC and Foraging Value

A recent article proposed that ACC signals the value of abandoning a current known resource to forage for a potentially more valuable resource (Kolling et al., 2012). In that article, the authors showed that, as the value of foraging for new resources increased, ACC became more active, and it was more active when subjects chose to actually forage instead of engaging with and exploiting the current known resources, similar to Donoso, Collins, and Koehlin (2014).

A more recent article (Shenhav et al., 2014) showed that as the value of foraging increases beyond the indifference point, so that foraging is an obviously better decision, then the activity of ACC decreases. This finding suggests that ACC is not simply signaling the value of foraging, but rather ACC activity is greatest when the decision is most difficult, in the sense that the subject is closest to their indifference point. In support of this, the authors also found that the RTs were greatest near the indifference point. Thus, the authors argue that ACC signals the difficulty of the choice or perhaps the value of exerting cognitive control over the choice, although the authors do not exclude other models that may account for their effects (Shenhav et al., 2014).

To investigate this, we simulated the foraging task of Kolling et al. (2012) and Shenhav et al. (2014). Briefly, the model was presented with a choice between engaging with two possible tokens of known value (one of which will subsequently be won) versus foraging for a new set of two possible tokens from among a set of six existing tokens. The choice to engage will lead to an immediate payoff, whereas the choice to forage will lead to a potentially better set of options but no immediate payoff. Furthermore, there is a small value cost associated with foraging.

The PRO-control model carries out the task as follows. First, the average value of the engage options is calculated. Next, the average value of the forage options is calculated. Finally, the relative value of foraging is calculated as the degree by which the foraging value exceeds the engage value, less any associated cost of foraging. The foraging value may be negative if the average value of the engage options exceeds the average value of the forage options. There are 10 stimulus input representations to the PRO-control model for this simulation. The relative foraging value on a given trial is binned into 1 of 10 equally spaced bins, and a stimulus input corresponding to the currently binned trial is activated as input to the model. The S-R weights are fixed to favor the engage option regardless of the stimulus bin, so that, by default, the model will choose to engage rather than forage for most trials. The model is given a set of trials such that the distribution of relative foraging values is uniform across the ten bins.

As the model performs the task, a choice to engage is deemed correct if the relative value of foraging is less than zero, but a choice to engage is deemed an error

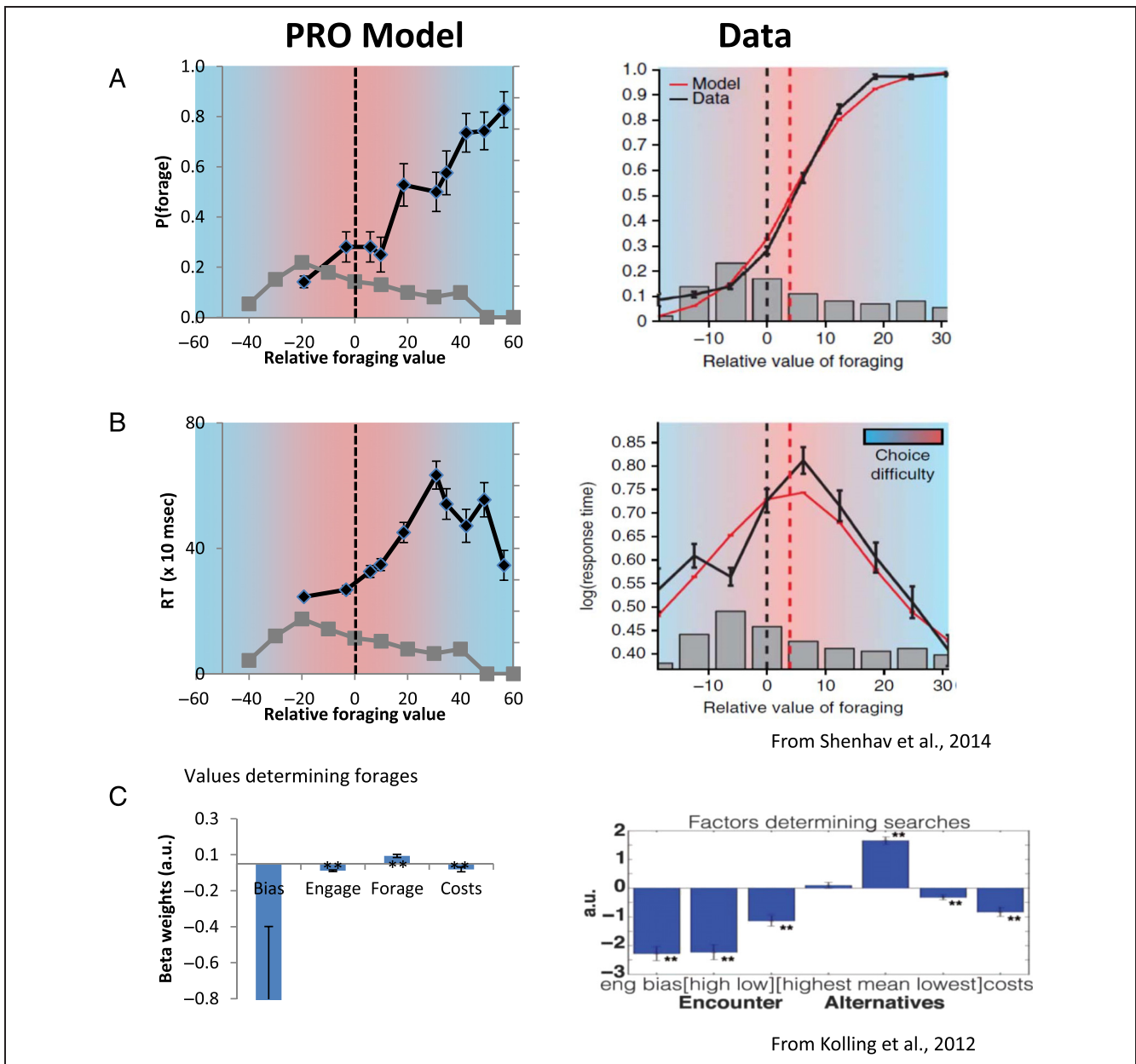
otherwise. Likewise, a choice to forage is deemed correct if the average of the newly acquired engage values is greater than the current engage values (less the foraging cost), and otherwise the choice to forage is deemed an error.

Errors have the main effect of strengthening the weights from stimuli to Controller R-O conjunctions (Figure 1), in proportion to how bad the choice was, that is, how far below (or above) zero was the relative value of foraging. This could be accounted for as a dopamine pause (Ljungberg, Apicella, & Schultz, 1992), which might lead to stronger weights by modulating D2 receptors. Indeed, both D2 receptor blockade and ACC lesions lead to reduced willingness to expend effort to achieve a greater reward (Walton et al., 2009). In the end, more obviously poor choices lead to stronger weights by which the current stimuli activate the Controller R-O conjunctions. In this way, the Controller learns to represent predicted R-O conjunctions in proportion to how likely they are to coincide with poor outcomes and also with how poor those outcomes would be (Brown & Braver, 2005, 2007).

The learned Controller R-O conjunctions in turn formed a basis for proactive control signals. When an error occurred, the activated Controller R-O conjunctions were trained to inhibit the just activated response unit, so that, in the future, the erroneous response would be less likely to be repeated under similar circumstances. Conversely, when a correct outcome occurred, the Controller R-O conjunctions were trained to weakly excite the just activated response unit. Thus, the controller exerts a “push-pull” effect on the response representations, inhibiting the wrong responses and weakly exciting the correct responses. The key is that the controller representations are only active in situations in which errors are likely to occur. (There is also a short-term, reactive control mechanism in this simulation, but its function is minimal here, and a fuller exposition of the reactive control mechanism is provided in Simulation 3.)

#### *Simulation 1 Results—Behavior*

The results of the simulation are shown in Figure 2, juxtaposed with earlier results (Kolling, Behrens, Wittmann, & Rushworth, 2016; Shenhav, Straccia, Botvinick, & Cohen, 2016; Shenhav et al., 2014; Kolling et al., 2012). Behaviorally, the model showed a preponderance of choices with negative relative foraging value, despite sampling the initial relative foraging values from a uniform distribution. This is consistent with the earlier empirical results and is due to the fact that high relative foraging value trials tended to lead to more foraging, which produced another round of relative foraging values; this continues until a trial occurs with a negative relative foraging value. Also, the model showed a bias toward engaging (Figure 2A), consistent with animal behavior following ACC lesions (Walton, Bannerman, &



**Figure 2.** PRO-control model simulations of foraging behavior. The PRO-control model was simulated on Experiment 2 of Shenhav et al. (2014). All PRO model parameters are taken from the flanker task simulation of Alexander and Brown (2011). Relative foraging values were discretized into 10 stimulus bins, one of which was active on any given trial. Foraging costs were included in the total relative foraging value. Stimulus–response weights were fixed at 0.6 for input to the foraging response and 1.0 for input to the engage response, which biased the model toward engage responses. The model had two response units, one indicating a forage response and the other indicating an engage response. The second stage of the decision, that is, between the two engage options, was not modeled. Feedback regarding the actual response was given to the PRO model immediately following a response. The results shown reflect 1000 simulated trials of the PRO model. (A) The PRO model simulates the choice probability with fits comparable to the drift-diffusion model. Of note, the model shows a bias toward engaging. (B) The PRO model shows RT as greatest near the indifference point, due to maximal competition between forage and engage responses which slows RT. (C) In agreement with human behavioral data, the model is more likely to choose foraging when the relative value of foraging is higher. It is less likely to choose to forage when the value of engaging is higher or the costs of foraging are higher.

Rushworth, 2002). This was implemented in the model by setting the S-R weights to be uniformly biased towards engaging, and so a certain number of incorrect “engage” choices had to be made to kick-start the learning of R-O conjunctions in the controller. The choice of a uniform bias toward engaging is a simplifying assumption here

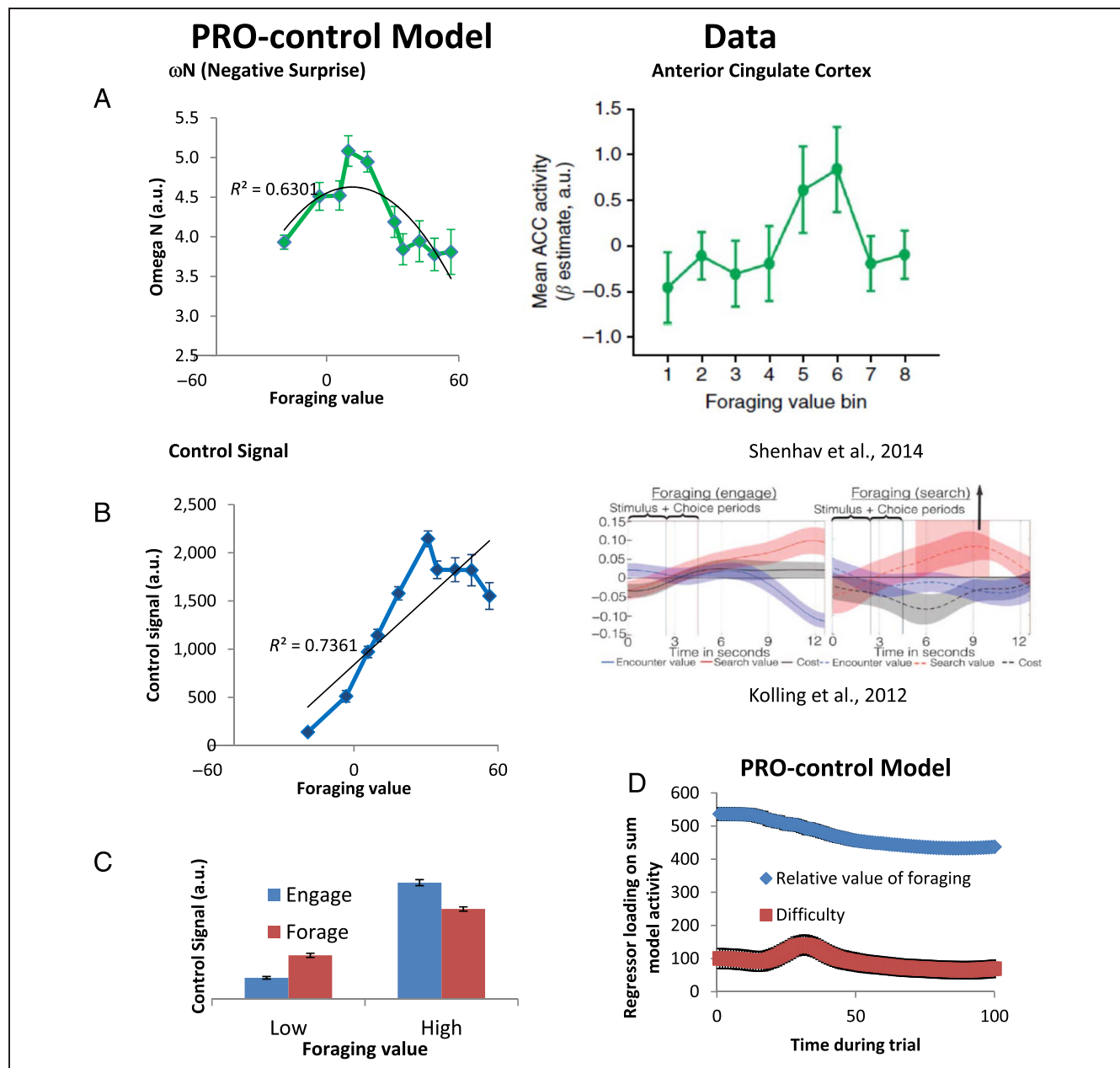
and does not address possible reinforcement learning mechanisms elsewhere that might counteract the engage bias (Walton et al., 2009). We then carried out a logistic regression of the forage versus engage choice probability as a function of average engage value, average forage value, and forage costs. As with the human data (Kolling

et al., 2012), the model was more likely to forage when the average forage value was higher, but less likely to forage when the average engage value was higher or when the cost of foraging was higher (Figure 2C).

*Simulation 1 Results—Neural Activity*

Figure 3 shows the simulated neural activity in the model as a function of relative foraging value. It is essential to note that there are two parts of the PRO-control model

shown here. Figure 3A shows the negative surprise (summed  $\omega^N$ ) generated by the model. This signal is maximal near the indifference point, that is, when foraging and engaging are equiprobable. The activity is greatest near the indifference point because whenever one option is chosen, the other option was expected with 50% probability, but it failed to occur, leading to a negative surprise signal. In contrast, when one option is more likely, it is in fact chosen more often. Furthermore, the choice is less surprising, all of which leads to a lower



**Figure 3.** PRO-control model simulations of foraging-related ACC activity. (A) Negative surprise (measured as the sum of  $\omega^N$  in the 2 sec following a response) is maximal near the indifference point, in agreement with fMRI results from Shenhav et al. (2014). (B) The model ACC Controller unit shows activity that correlates strongly and positively with the relative value of foraging. Of note, control activity decreases at the highest relative foraging values, there are fewer surprising engage choices to drive learning of the control signal. (C) This correlation occurs whether the ultimate choice is foraging or engaging, in agreement with Kolling et al. (2012). (D) The model activation (summed across proactive controller and negative surprise activity) correlates with relative foraging value early in the trial, but it correlates with difficulty later in the trial.



average negative surprise signal. The net result is an inverted-U function of activity versus relative foraging value, as observed by Shenhav et al. (2014). This simulation result suggests that ACC activity observed by Shenhav et al. (2014) may reflect negative surprise regarding the choice as a function of relative foraging value.

Kolling et al. (2012) suggest that ACC activity signals the relative value of foraging, such that greater relative foraging value will correlate with greater ACC activity. Shenhav et al. (2014, 2016) find only the inverted-U pattern of ACC activity (Figure 3A) and argue against an interpretation in terms of foraging value. We propose here that both the inverted U (i.e., negative surprise) and relative foraging value signals (i.e., proactive control) may coexist as distinct signals within ACC. Figure 3B shows the model Controller R-O conjunction activity as a function of relative foraging value. Of note, the activity shows a strong positive correlation with foraging value. This is because choices to engage are increasingly “bad” as the relative value of foraging is higher. This leads to greater learned activation of the Controller R-O conjunctions at higher relative foraging values. In turn, the higher Controller activity leads to a proportionally increased probability of foraging, as seen in Figure 2A. Thus, the model Controller signals the relative value of foraging, and the model Critic signals the negative surprise associated with the choice.

It is notable also that, in Figure 3B, the Controller R-O conjunction activity shows a slight drop-off of activity at the highest relative foraging value conditions, despite the overall positive correlation of activity and foraging value. This inverted-U shape results from the learning law (Equation 1), which is a product of both the magnitude of the error and the probability of the error. This trains proactive control representations more strongly for larger errors (e.g., choosing to engage despite a larger relative foraging value), but the learning effect is weaker overall at the highest relative foraging values, because then the model rarely makes an error. Additional simulations show that this inverted-U effect remains even when RTs are held artificially uniform in the model, so it is not due to correlations with RT as in Figure 2B (Grinband et al., 2011), nor is it due to differences in trial frequency, because the relative foraging values are uniformly distributed across new trials. It is also not due to changes in the learning rate as a function of surprise (Equation 5 of Alexander & Brown, 2011). This drop-off effect may partly account for why it has been difficult to disentangle the relative foraging value correlate from the difficulty correlate in some studies (Shenhav et al., 2014, 2016).

The model also predicts that relative foraging value and difficulty signals will be seen at different times in a given trial. Figure 3D begins with the model activation summed across the model Controller R-O conjunction activity and the negative surprise activity at each time point in a trial. This summed activity is regressed against relative foraging value and difficulty (and also RT as a

nuisance regressor). Figure 3D shows that the model activity correlates with relative foraging value early in the trial, as the controller generates proactive control signals. Later in the trial, the correlation with difficulty increases, as the model generates a response and signals negative surprise. In this way, signals that correlate with relative foraging value and difficulty can be dissociated in time by the model. This agrees with recent findings in Figure 2A and B of an imaging study of essentially the same task (Kolling et al., 2016).

### *Simulation 1 Results—Lesions*

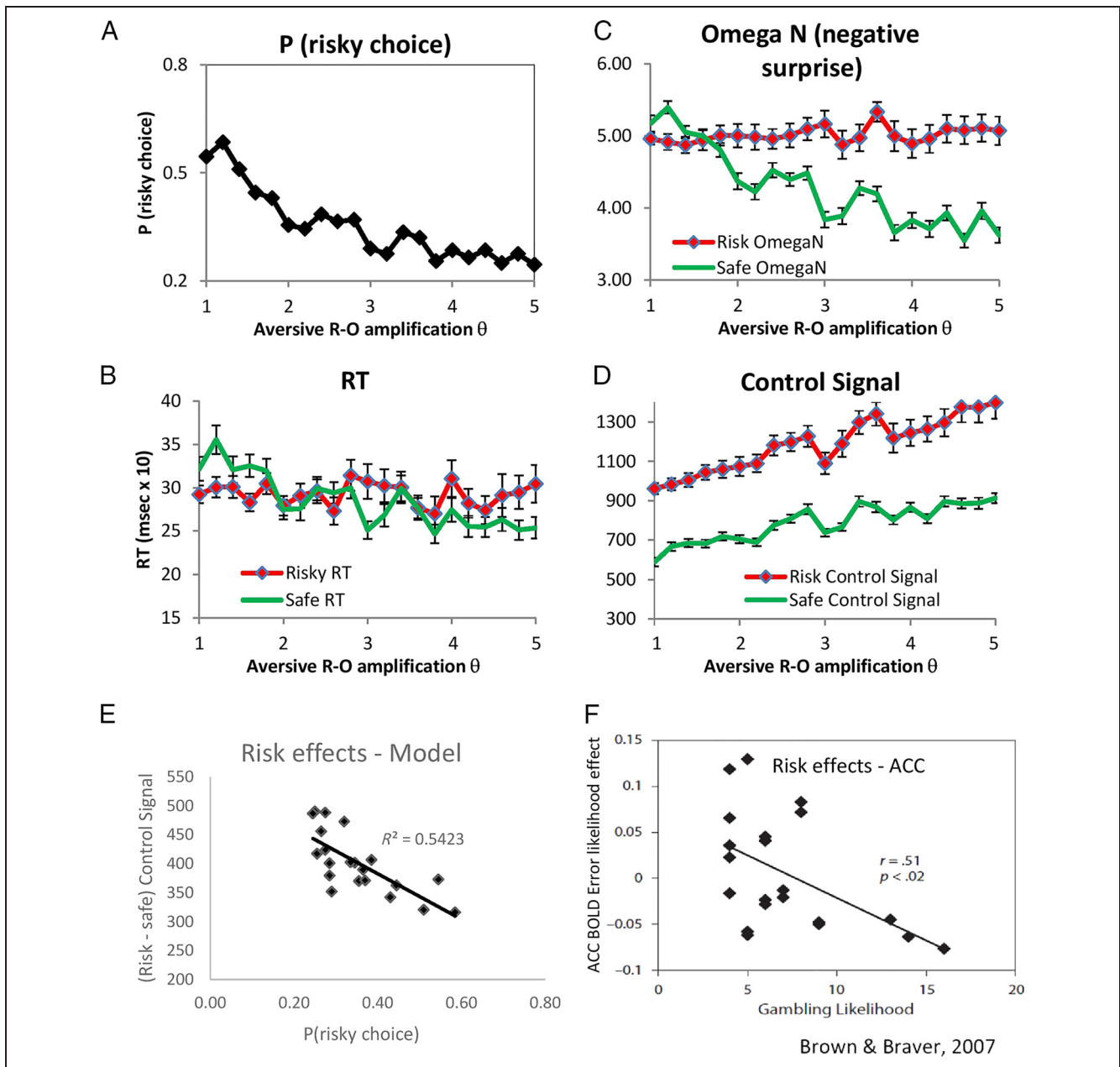
To explore the role of the Controller on decisions, we virtually lesioned the Controller R-O conjunction units by setting their activities to always be zero. The result was that subjects chose to forage only about 10% of the time, regardless of the relative value of foraging. This is in agreement with the reduced motivation to expend effort to obtain larger rewards, as observed in rats (Walton et al., 2002, 2009).

### **Simulation 2. ACC and Risk Avoidance**

In Simulation 1, the model Controller is shown to learn and signal when potentially undesirable outcomes are likely to occur, in proportion to both how likely the bad outcome is and how bad it would be. In turn, these representations provide a basis for a control signal that effectively biases against choices that would lead to such bad outcomes. This constitutes a kind of risk avoidance circuit. ACC has been suggested as a region that drives risk avoidance (Fukunaga et al., 2012; Krawitz et al., 2010; Brown & Braver, 2005, 2007, 2008) as a complementary function to vmPFC (Shenhav et al., 2016; Fukunaga et al., 2012), but it has until now been unclear how the mechanisms of ACC might drive risk avoidance. To explore how the new model might drive risk avoidance, we simulated a simple task in which a single stimulus can drive two responses. One response, the “safe” response, has smaller S-R weights and thus a lower default probability of being chosen. Still, it yields a correct outcome 100% of the time. The other response, the “risky” response, is more likely to be chosen by default, but it yields a correct outcome only 50% of the time. The maximum possible weight from stimuli to Controller R-O conjunctions is defined by the “Aversive R-O amplification” parameter ( $\theta$ ). Larger values of  $\theta$  lead to larger activations of Controller R-O conjunctions that are associated with errors. We parametrically manipulated  $\theta$  and evaluated the model for each value.

### *Simulation 2 Results*

The results are shown in Figure 4. As expected, Figure 4A shows that greater values of  $\theta$  are associated with a reduced probability of choosing the risky option, although there is



**Figure 4.** Simulations of risk avoidance. The PRO-control model was simulated on a 2AFC task with a “risky” option that provided a win outcome only 50% of the time and a “safe” option that provided a win 100% of the time. There was one stimulus, a cue to make a response. The S-R weights were adjusted so that in the absence of a control signal, the model would choose the risky option most of the time. The maximum possible weights from the stimulus to the Controller R-O conjunctions ( $\rho$ ) were manipulated parametrically. (A) Greater values of  $\rho$  led to stronger learned control signals as choosing the risky option led to poor outcomes, which in turn trained the Controller to become more active and suppress the “risky” response associated with the poor outcomes. (B) RT was little changed by changes in risk avoidance. (C) Greater learning of poor outcomes led to a greater difference in negative surprise response to risky versus safe choices. Essentially, the choice of the risky option became less frequent and therefore more surprising when it did happen. (D) The sum of the negative surprise and the Controller activity was stronger as  $\rho$  increased, and it was slightly stronger when the risky option was chosen. (E) Greater model ACC risk sensitivity correlates with reduced probability of choosing the riskier option. Each data point represents a unique value of  $\theta$ , corresponding to varying individual differences in risk avoidance. Activation is derived as the difference of the Risk minus Safe control signal in D, and the probability of risky choice is derived from A. (F) Empirical results consistent with model properties in E. Shown is a neuroimaging signal derived from the dorsal ACC in humans (adapted with permission from Brown & Braver, 2007). Each data point represents a subject. The horizontal axis represents increasing risk taking, and the vertical axis represents the dorsal ACC fMRI bold contrast of higher minus lower error likelihood.

little to no effect on RT (Figure 4B). When  $\theta$  is larger, each error leads to stronger learning signals in the controller (Equation 1), which leads to stronger inhibition of the riskier response that led to an error (Figure 4A). Figure 4C

shows that, as  $\theta$  increases, the negative surprise associated with choosing the safe option decreases, because the safe option is chosen more frequently. The net result is that with higher risk avoidance, the model negative surprise

signal shows a stronger activity contrast of risky minus safe. This agrees with a number of findings from neuroimaging (Fukunaga et al., 2012; Krawitz et al., 2010; Brown & Braver, 2005, 2007, 2008), in which risky option choices entail greater ACC activity, especially in those subjects who are more risk averse.

### Simulation 3. ACC and Reactive Control

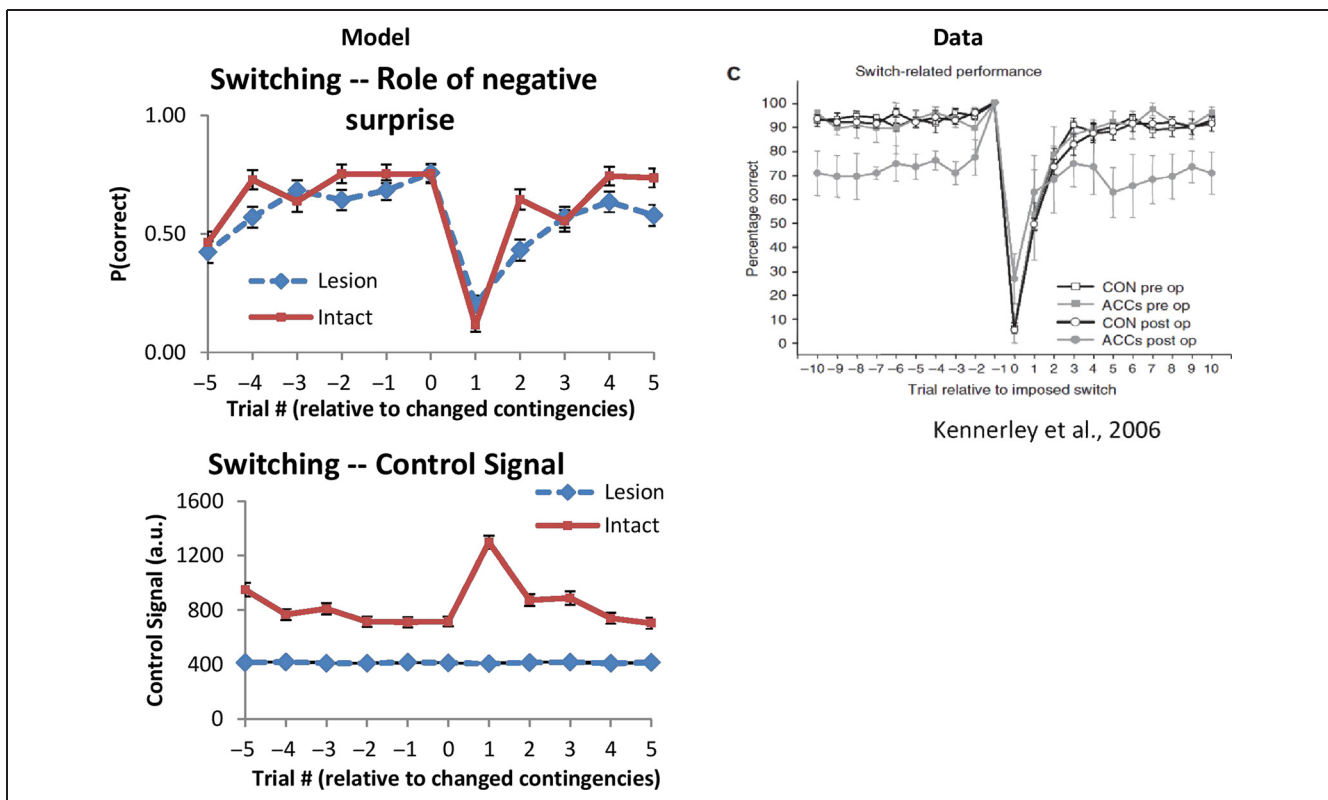
ACC has been cast as implementing reactive control (Braver et al., 2007). The distinction between proactive and reactive control is that proactive control implements control over decision-making before an error occurs (in anticipation that more control is needed to avoid an error). On the other hand, reactive control implements control after an error has been made to prevent further errors.

The PRO-control model builds on the PRO model by adding a direct reactive control signal from the negative surprise units to the response layer (Figure 1). Thus, when a rewarding outcome does not occur, the negative surprise signal indicates the omission of the expected outcome. In turn, the inhibitory connection from the active negative surprise unit to the active response unit is strengthened and provides a short-term inhibition of the

response that persists over the next few trials. This causes the model to rapidly switch away from responses that lead to poor outcomes. To investigate this function, we used a simple 2AFC task in which the correct, rewarded response stays the same for between 5 and 10 trials. Then without warning, the correct response changes (with probability  $p = .3$  per trial after five trials), and the previously correct response now yields reward. The model must detect the changed contingency from the absence of expected reward and then switch tasks accordingly. This is substantially similar to a task that has been shown to elicit ACC activity in monkeys (Kennerley, Walton, Behrens, Buckley, & Rushworth, 2006; Shima & Tanji, 1998) and humans (Bush et al., 2002).

### Simulation 3 Results

The model was able to perform the task with well over chance accuracy, switching rapidly and maintaining the new response set until another switch occurred (Figure 5). Still, it is possible that the Controller R-O conjunctions were rapidly learning to suppress the inappropriate response, so that the reactive control signal from negative surprise would be unnecessary. To test whether the reactive control pathway was necessary for performing



**Figure 5.** Simulation of how reduced reward leads to switching. In this 2AFC task, the rewarded response is the same for at least five trials, and then it switches with  $p = .3$ . Once a switch occurs, another five trials must elapse before another switch can occur. Left: In the model, reduced reward leads to a negative surprise signal, indicating that an expected reward did not occur. This in turn rapidly suppresses the just-executed response and maintains the suppression for the next few trials. When the reactive control signal from the negative surprise unit is lesioned, then the model is unable to adapt rapidly to changing contingencies, and the performance is closer to chance. Right: Behavioral data from monkeys showing impaired switching following ACC sulcus lesions. Adapted with permission from Kennerley et al. (2006).

the task, we virtually lesioned the reactive control pathway by setting all of the weights in this pathway to zero. Figure 5 shows that, in the lesioned case, the model performs poorly compared with the intact model, with weaker maintenance of the new task over several trials. Overall, this demonstrates that negative surprise is in principle sufficient to drive reactive control, such that once negative surprise exceeds a certain level of activity, the model switches away from its current set. This is consistent with findings from monkey neurophysiology showing that animals switch their strategy when certain ACC cell activities exceed a fixed threshold (Hayden et al., 2011).

#### **Simulation 4: Hot-hand Bias and Post-correct Speeding**

In the context of value-based decision-making, it is generally assumed that a behaving animal will select behaviors that maximize the frequency and magnitude of reward while minimizing aversive outcomes. Given this, it is ambiguous whether ACC signals related to reactive control (i.e., those generated following feedback) during value-based decision-making are due to the behavioral import of feedback or due to its surprising nature. In the PRO model, ACC activity is interpreted as being affectively neutral, and this interpretation is maintained in the PRO-control model. In the previous simulation, it was demonstrated that aversive feedback coupled with negative surprise signals generated by ACC could drive rapid adjustments in behavioral strategies to avoid future error. As a corollary to this, it may be expected that rewarding feedback coupled with reactive ACC signals could serve to increase the probability of repeating a rewarded behavior—the so-called hot-hand bias. The hot-hand bias refers to the tendency to perceive positive serial autocorrelations in independent sequential events and has been demonstrated in monkeys (Blanchard & Hayden, 2014; Blanchard, Wilke, & Hayden, 2014). In particular, the hot-hand bias is evidenced by a win-stay strategy despite the independence of a sequence of events.

Similarly, reactive control is thought to contribute to post-error slowing—the tendency of subjects to respond less rapidly following behavioral error (Soshi et al., 2015; Narayanan, Cavanagh, Frank, & Laubach, 2013; Cavanagh, Gründler, Frank, & Allen, 2010). Recent evidence has suggested a corollary effect—post-correct speeding—in which responses following correct trials are more rapid. To investigate the model's ability to capture these effects, we simulated the model on a version of the Correlated Outcomes Task (Blanchard & Hayden, 2014). In our implementation, the model repeatedly selected between two options. On each trial, one of the options was rewarded with a probability of 1. However, the identity of the rewarded option could change with a fixed, condition-dependent probability following each trial. A total of nine conditions were simulated in which the probability of a change occurring was manipulated (from .1 to

.9). Two versions of the model were simulated as in Simulation 3, with reactive control weights either intact or artificially lesioned. The Correlated Outcomes Task was previously used to identify behavioral correlates of a hot-hand bias—the tendency to overestimate the likelihood that a successful trial will be followed by another successful trial for the same behavioral response—in rhesus monkeys.

#### *Simulation 4 Results*

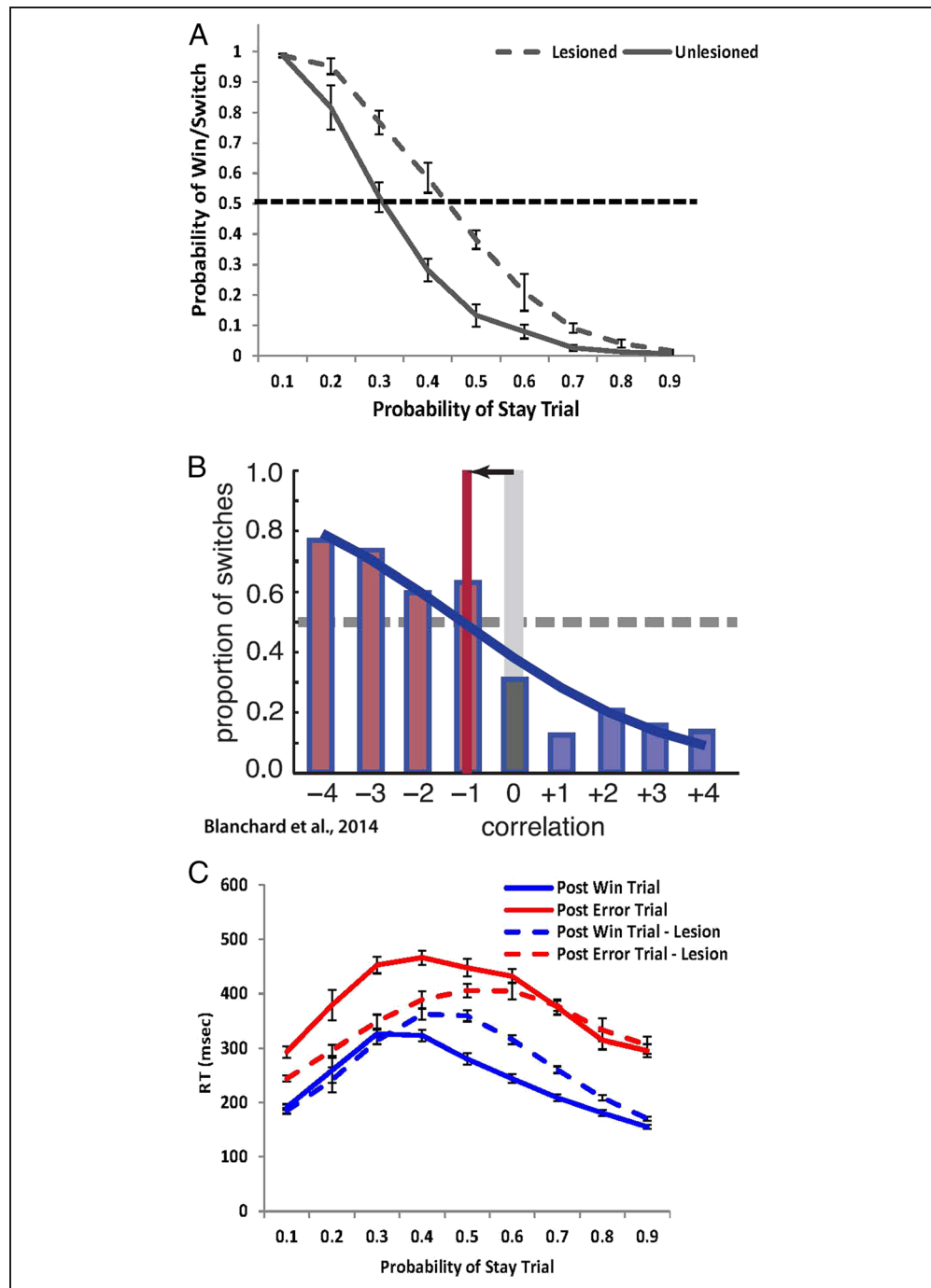
In our simulations (Figure 6), the hot-hand bias emerges from the transient allocation of reactive control following the unexpected omission of an aversive outcome. Following a rewarded trial, both proactive control weights (Equation 14 of Alexander & Brown, 2011) and reactive control weights (Equation 4) are updated to reflect long-term S-RO contingencies as well as immediate control demands, respectively. The bias in favor of repeating a previously rewarded response arises due to the combination of negative surprise with the affective import of the unexpected omission of an aversive outcome: The behavior that contributed to the rewarded outcome is actively promoted by reactive control.

The model further suggests that reactive control may contribute both to post-error slowing as well as post-correct speeding. Relative to a model in which reactive control weights have been lesioned, the intact PRO-control model exhibits increased RTs following error and decreased RTs following correct responses. Interestingly, this pattern is dependent on the probability that the same response will be required on the subsequent trial; when the identity of the correct response changes frequently (probability of a stay trial < .4), reactive control appears to have a stronger influence on post-error slowing, whereas RTs for the post-win trial are identical for the lesioned and nonlesioned simulations. However, when the identity of the correct response is likely to remain the same (probability of a stay trial > .6), RTs following error trials are the same for the lesioned and nonlesioned simulations, whereas RTs decreased for the nonlesioned model relative to the lesioned model following correct trials.

#### **Backward Compatibility with Original PRO Model**

The new model's ability to capture new effects above raises the question of whether the required modifications break the simulations of the original PRO model. In particular, the new reactive control mechanism driven by negative surprise allows rapid switching when contingencies change. In a probabilistic environment in which the correct option is not always rewarded, might this lead to erratic lose-shift behavior that would render the model unable to maintain correct performance? To address this question, we reran the Behrens et al. task simulation (Behrens et al., 2007) on the new model using the same

**Figure 6.** Hot-hand bias and post-correct speeding. In our implementation of the Correlated Outcomes Task, the probability of a previously rewarded response remaining the same on a subsequent trial is manipulated. (A) Reactive control in the model contributes to a bias to repeated previously rewarded behaviors, whereas lesions of control weights  $W^o$  eliminate this bias. (B) Behavioral results from monkeys performing the Correlated Outcome Task also reveal a hot-hand bias (adapted with permission from Blanchard et al., 2014). (C) Reactive control-dependent differences in RT are influenced by global environmental contingencies; in environments for which the identity of the rewarded response is likely to change frequently, the nonlesioned model responds more slowly than the lesioned model, whereas in environments for which the rewarded response is likely to remain the same, reactive control contributes to speeded RTs.



task design as in the original PRO model simulation (Alexander & Brown, 2011). In this two alternative forced-choice task, correct trials were rewarded only 80% of the time. In the simulation, there was an initial non-volatile period of 120 trials, followed by a volatile period of 280 trials in which the contingencies changed every 40 trials, followed by a nonvolatile period in which the contingencies were unchanged for 180 trials. We found that the intact model was able to perform the task above chance ( $p = .0002$ , Fisher exact, two-tailed), with 60.9% reward rate (chance = 50%, max possible performance = 80%).

This demonstrates that the model could learn the task even with periodic reward omission. Furthermore, the performance was only above chance during the nonvolatile periods (67.7%,  $p < .0001$ , Fisher exact, two-tailed), whereas it was reduced during the volatile period (54.6%,  $p = .31$ , Fisher exact, two-tailed). At the neural level, we found that, as in the original PRO model simulation and consistent with empirical findings (Behrens et al., 2007), the average negative surprise signals were greater during the volatile periods than during the nonvolatile periods ( $t(578) = 8.38$ ,  $p < .001$ ). This demonstrates that the

additions to the model do not substantially change the results relative to the original PRO model simulations.

## DISCUSSION

The above four simulations collectively suggest how ACC may control value-based decision-making. The PRO-control model casts ACC as learning both proactive and reactive control signals (Braver et al., 2007), which operate on longer and shorter timescales, respectively. This is consistent with the range of timescales reported in ACC (Wittmann et al., 2016; Bernacchia, Seo, Lee, & Wang, 2011). The control signals act directly on action representations rather than stimulus representations, consistent with a role in representing action values rather than stimulus values (Rudebeck et al., 2008).

In Simulation 1, the proactive control signals provide a representation that simulates the foraging value signals reported in a foraging task (Kolling et al., 2012), whereas the negative surprise representations simulate the inverted-U activation as a function of relative foraging value, as reported for a similar foraging task (Shenhav et al., 2014). The results suggest that both signals may coexist and provide distinct functions within ACC. Consistent with an ACC role in effortful behavior, the probability of foraging is significantly lower in the simulation if the proactive control signals that represent foraging value are lesioned.

Simulation 2 provides a mechanistic account of how ACC may drive risk avoidance, as a consequence of proactive control. When costs or losses occur, the net effect is to train proactive control signals to represent the situation in which such losses may occur. This idea has been proposed earlier (Brown & Braver, 2005; Holroyd et al., 2005). What has not been shown until now is how the same proactive control signals may both drive effortful behavior, as in the foraging decisions of Simulation 1, and also drive risk avoidance, as in Simulation 2. This approach is consistent with a recent proposal that ACC energizes overall strategy choices (Holroyd & McClure, 2015), biasing behavior against default but lower payoff options and instead toward larger gains that require increased effort to obtain.

Simulation 3 illustrates a dissociation between the longer-term proactive control signals and the shorter-term reactive control signals. In Simulation 3, rapid switching and updating due to reduced reward and changed contingencies was heavily dependent on negative surprise, such that the omission of an expected reward signaled the need to switch. Lesions of the reactive control pathway led to severe deficits in rapid switching so that behavior was near chance. Such lesions of the reactive control pathways had no substantial impact on the foraging and risky decision tasks of Simulations 1 and 2. This is consistent with earlier reports suggesting that ACC provides inhibition (Medalla & Barbas, 2009), which may lead to disengagement from less adaptive behavioral

sets (Donoso et al., 2014; Hochman, Vaidya, & Fellows, 2014), especially in the absence of clear external cues indicating the correct action (Kennerley, 2003; Shima & Tanji, 1998).

Finally, Simulation 4 demonstrates the dual role of reactive control in avoiding behavioral error as well as exploiting global environmental contingencies. The hot-hand bias—the tendency to repeat responses that have been rewarded in the immediate past—is observed to be a consequence of transient application of reactive control. In the model, reactive control is a product of negative surprise, the unexpected nonoccurrence of a predicted outcome. In Simulation 4, the unexpected nonoccurrence of an aversive outcome following a rewarded behavior contributes to the model's increased likelihood to select that behavior on subsequent trials. Moreover, effects related to RT such as post-error slowing and post-correct speeding are observed to depend on global environment contingencies: For conditions in which the identity of a rewarded response is likely to change frequently, reactive control is deployed to slow responses, whereas in stable conditions reactive control promotes more rapid responses. The PRO-control model thus provides an account of how ACC may be involved in mediating the tradeoff between deliberative control of behavior and enhanced response vigor.

### Energizing Behavior/Expected Value of Control

The PRO-control model offers a novel perspective relative to existing theories of ACC function. A recent proposal argues that ACC has a role in “energizing” or otherwise driving motivated behavior by maintaining high-level representations of behavior directed toward specific longer-term goals (Holroyd & Yeung, 2012). According to this theory, ACC is not concerned with the minutiae of task implementation, which is instead handled in the lateral pFC and BG. Rather, dopaminergic reinforcement signals train ACC to represent “option values” that exert control to activate certain behaviors. Our simulations accord well with these notions. Specifically, pauses in the dopamine signal may serve to train both the proactive and reactive control signals to exert control over tasks. In the absence of such control signals, our model defaults to taking the immediately available reward (Walton et al., 2009) rather than exerting the effort required to obtain a potentially larger reward, and the influence of previous reward history is attenuated (Kennerley et al., 2006; Figure 5).

### Subregions of mPFC

There has been substantial recent controversy over the function of distinct subregions of the mPFC. In a recent response to the ongoing controversy over foraging value and difficulty effects in mPFC, Kolling et al. suggested that difficulty effects may be found more dorsally in the

mPFC, toward the pre-SMA, whereas foraging value effects may be found more ventrally (Kolling et al., 2016). If so, that would account for how both effects may exist simultaneously. Shenhav et al. have countered that, in a subsequent replication study, they find no evidence for foraging value effects (Shenhav et al., 2016). Our simulation results (Figure 3B) show that even in the regions that correlate with relative foraging value, the highest relative foraging values lead to somewhat weaker activation, leading the activity to look more like a difficulty signal (cf. Figure 3A). This may account for why positive correlations with relative foraging value have not always been found (Kolling et al., 2016; Shenhav et al., 2014, 2016), despite the existence of distinct signals that correlate with relative foraging value and difficulty effects in our model (Figure 3).

Another recent controversy has begun over whether ACC signals primarily pain (Lieberman & Eisenberger, 2015) or also other cognitive functions (Wager et al., 2016). Lieberman and Eisenberger argue that pain is represented more ventrally, whereas cognitive effects are found more dorsally. This accords well with the proposal by Kolling et al. (2016), and it suggests a mapping of the present PRO-control model components such that learned predictions and representations of negative affect (underlying proactive control) may be represented more ventrally in the dorsal ACC, whereas prediction errors (underlying reactive control) may be represented more dorsally, toward the pre-SMA. To help resolve this controversy, Jahn et al. (2016) directly compared pain, conflict, and prediction error in a full factorial fMRI design. They found that prediction error effects are represented more dorsally in mPFC, straddling ACC, and the pre-SMA, consistent with our account of difficulty effects as reflecting prediction error. Likewise, pain effects are more ventral, which could reflect a specific kind of negative affect, and this negative affect in the context of a current set of options might also drive foraging decisions.

### Proactive versus Reactive Mechanisms of Control

The PRO-control model provides a new degree of mechanistic clarity to the notions of proactive and reactive control signals, which have been explored both theoretically (Aron, 2011; Braver et al., 2007; De Pisapia & Braver, 2006) and empirically (Marini, Demeter, Roberts, Chelazzi, & Woldorff, 2016; Kawai, Yamada, Sato, Takada, & Matsumoto, 2015). Although we have cast proactive control as one of several functions of ACC (Kawai et al., 2015), we do not claim that proactive control is solely driven by ACC, as indeed there is evidence that proactive control can be exerted by a network of other regions (Marini et al., 2016; MacDonald, Cohen, Stenger, & Carter, 2000). Perhaps the closest computational model to the PRO-control model is a recent model of dual control processes (Ziegler et al., 2014). That model casts ACC as representing proactive and reactive control signals detected as conflict and surprise, respectively, but is posits

cholinergic and noradrenergic brainstem mechanisms, with abstract conflict and error computations, instead of the cortical mechanisms proposed here.

We do not claim that the proactive and reactive control signals simulated here are the only outputs generated by the mPFC. There is evidence that ACC modulates learning rates (Behrens et al., 2007), as we simulated in the original PRO model (Alexander & Brown, 2011). Also, prediction error signals may be useful to training working memory representations in lateral pFC (Alexander & Brown, 2015), which implies a role in learning, beyond the control signals simulated here. ACC may also modulate brainstem neuromodulator activity as in the locus coeruleus (Aston-Jones & Cohen, 2005). Also, ACC is one of a small set of regions that projects to the striosomes of the striatum, which may serve as a basis to drive prediction error signals in dopamine cells (Eblen & Graybiel, 1995). Overall, it appears that prediction and prediction error signals generated by ACC may serve multiple roles simultaneously.

### Relation to ERP Effects

The PRO-control model suggests how aversive outcomes may train proactive control signals, whereas unvalenced prediction error signals (especially negative surprise,  $\omega^N$ ) may be trained by reinforcement to drive reactive control (Braver et al., 2007; Ridderinkhof, Ullsperger, Crone, & Nieuwenhuis, 2004; Schall, Stuphorn, & Brown, 2002). This accords well with prior work arguing for a role of dopamine signal pauses on ACC activity, although there remains some question about whether ACC signals aversive signals, or unsigned prediction errors, or both. Some have argued that phasic pauses of the midbrain dopamine signals lead to disinhibition of ACC (Baker & Holroyd, 2011; Brown & Braver, 2005; Holroyd & Coles, 2002), which could account for greater ACC activity following an error such as in the feedback error-related negativity (Holroyd & Coles, 2002). Separately, the N200 has been interpreted as a conflict signal (Baker & Holroyd, 2011), although more generally, the N200 is increased in response to improbable events (Hajihosseini & Holroyd, 2013). This observation is consistent with the N200 reflecting at least in part the surprise signals generated in the PRO-control model as the negative surprise ( $\omega^N$ ) signal. Even the feedback error-related negativity may reflect at least in part a surprise signal rather than a reinforcement signal (Talmi, Atkinson, & El-Deredy, 2013), consistent with other reports that frontocentral negativities signal surprise rather than aversive outcomes (Garofalo et al., 2014; Ferdinand et al., 2012; Jessup et al., 2010; Oliveira et al., 2007).

### Clinical Implications

Our results provide new mechanistic perspective on some clinical and developmental issues involving proactive

and reactive control. For example, schizophrenia patients appear to have a deficit specifically in proactive control, as they fail to show increased activation in a proactive control task, but not in a reactive control task (Lesh et al., 2013). Proactive control seems to develop later than reactive control mechanisms in children (Chatham, Frank, & Munakata, 2009). This highlights the importance of proactive control, which has been emphasized elsewhere in general (Aron, 2011) and with regard to specific clinical disorders such as social anxiety (Schmid, Kleiman, & Amodio, 2015). Proactive control may also reduce gambling behavior (Verbruggen, Adams, & Chambers, 2012). Our simulations above suggest how such control mechanisms might be implemented at the neural level.

### Acknowledgments

The authors thank Nils Kolling, Matthew Rushworth, Amitai Shenhav, and Matthew Botvinick for helpful discussions and comments. J. W. B. thanks Matthew Rushworth for hosting him as a visiting fellow on sabbatical, during which part of this work was developed. W. H. A. was supported in part by FWO-Flanders Odysseus II Award #G.OC44.13N.

Reprint requests should be sent to Joshua W. Brown, Department of Psychological and Brain Sciences, 1101 E. Tenth St., Bloomington, IN 47405-7007, or via e-mail: jwmbrown@indiana.edu.

### REFERENCES

- Aarts, E., Roelofs, A., & van Turenout, M. (2008). Anticipatory activity in anterior cingulate cortex can be independent of conflict and error likelihood. *Journal of Neuroscience*, *28*, 4671–4678.
- Alexander, W. H., & Brown, J. W. (2010). Computational models of performance monitoring and cognitive control. *Topics in Cognitive Science*, *2*, 658–677.
- Alexander, W. H., & Brown, J. W. (2011). Medial prefrontal cortex as an action-outcome predictor. *Nature Neuroscience*, *14*, 1338–1344.
- Alexander, W. H., & Brown, J. W. (2014). A general role for medial prefrontal cortex in event prediction. *Frontiers in Computational Neuroscience*, *8*, 69.
- Alexander, W. H., & Brown, J. W. (2015). Hierarchical error representation: A computational model of anterior cingulate and dorsolateral prefrontal cortex. *Neural Computation*, *27*, 2354–2410.
- Alexander, W. H., Fukunaga, R., Finn, P., & Brown, J. W. (2015). Reward salience and risk aversion underlie differential ACC activity in substance dependence. *Neuroimage: Clinical*, *8*, 59–71.
- Amador, N., Schlag-Rey, M., & Schlag, J. (2000). Reward-predicting and reward-detecting neuronal activity in the primate supplementary eye field. *Journal of Neurophysiology*, *84*, 2166–2170.
- Amiez, C., Sophie, C. A., Charles, R. E. W., Procyk, E., & Petrides, M. (2015). A unilateral medial frontal cortical lesion impairs trial and error learning without visual control. *Neuropsychologia*, *75*, 314–321.
- Apps, M. A. J., Balsters, J. H., & Ramnani, N. (2012). The anterior cingulate cortex: Monitoring the outcomes of others' decisions. *Social Neuroscience*, *7*, 424–435.
- Aron, A. R. (2011). From reactive to proactive and selective control: Developing a richer model for stopping inappropriate responses. *Biological Psychiatry*, *69*, e55–e68.
- Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*, 403–450.
- Baker, T. E., & Holroyd, C. B. (2011). Dissociated roles of the anterior cingulate cortex in reward and conflict processing as revealed by the feedback error-related negativity and N200. *Biological Psychology*, *87*, 25–34.
- Barto, A. G., Sutton, R. S., & Anderson, C. W. (1983). Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Transactions on Systems, Man, and Cybernetics*, *13*, 834–846.
- Behrens, T. E., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*, 1214–1221.
- Bernacchia, A., Seo, H., Lee, D., & Wang, X.-J. (2011). A reservoir of time constants for memory traces in cortical neurons. *Nature Neuroscience*, *14*, 366–372.
- Blais, C., & Bunge, S. (2010). Behavioral and neural evidence for item-specific performance monitoring. *Journal of Cognitive Neuroscience*, *22*, 2758–2767.
- Blanchard, T. C., & Hayden, B. Y. (2014). Neurons in dorsal anterior cingulate cortex signal postdecisional variables in a foraging task. *Journal of Neuroscience*, *34*, 646–655.
- Blanchard, T. C., Wilke, A., & Hayden, B. Y. (2014). Hot-hand bias in rhesus monkeys. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 280–286.
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. C. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*, 624–652.
- Braver, T. S., Gray, J. R., & Burgess, G. C. (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. In A. Conway, C. Jarrold, M. Kane, A. Miyake, & J. Towse (Eds.), *Variation of working memory* (pp. 76–106). Oxford: Oxford University Press.
- Brown, J. W. (2009). Conflict effects without conflict in anterior cingulate cortex: Multiple response effects and context specific representations. *Neuroimage*, *47*, 334–341.
- Brown, J. W. (2014). The tale of the neuroscientists and the computer: Why mechanistic theory matters. *Frontiers in Neuroscience*, *8*, 349.
- Brown, J. W., & Braver, T. S. (2005). Learned predictions of error likelihood in the anterior cingulate cortex. *Science*, *307*, 1118–1121.
- Brown, J. W., & Braver, T. S. (2007). Risk prediction and aversion by anterior cingulate cortex. *Cognitive, Affective, & Behavioral Neuroscience*, *7*, 266–277.
- Brown, J. W., & Braver, T. S. (2008). A computational model of risk, conflict, and individual difference effects in the anterior cingulate cortex. *Brain Research*, *1202*, 99–108.
- Bryden, D. W., Johnson, E. E., Tobia, S. C., Kashatelyan, V., & Roesch, M. R. (2011). Attention for learning signals in anterior cingulate cortex. *Journal of Neuroscience*, *31*, 18266–18274.
- Bush, G., Vogt, B. A., Holmes, J., Dale, A. M., Greve, D., & Jenike, M. A. (2002). Dorsal anterior cingulate cortex: A role in reward-based decision making. *Proceedings of the National Academy of Sciences*, *99*, 507–512.
- Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D. C., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, *280*, 747–749.
- Carter, C. S., Macdonald, A. M., Botvinick, M., Ross, L. L., Stenger, A., Noll, D., et al. (2000). Parsing executive



- processes: Strategic versus evaluative functions of the anterior cingulate cortex. *Proceedings of the National Academy of Sciences*, *97*, 1944–1948.
- Cavanagh, J. F., Gründler, T. O. J., Frank, M. J., & Allen, J. J. B. (2010). Altered cingulate sub-region activation accounts for task-related dissociation in ERN amplitude as a function of obsessive-compulsive symptoms. *Neuropsychologia*, *48*, 2098–2109.
- Charnov, E. L. (1976). Optimal foraging, the marginal value theorem. *Theoretical Population Biology*, *9*, 129–136.
- Chatham, C. H., Frank, M. J., & Munakata, Y. (2009). Pupillometric and behavioral markers of a developmental shift in the temporal dynamics of cognitive control. *Proceedings of the National Academy of Sciences, U.S.A.*, *106*, 5529–5533.
- Crottaz-Herbette, S., & Menon, V. (2006). Where and when the anterior cingulate cortex modulates attentional response: Combined fMRI and ERP evidence. *Journal of Cognitive Neuroscience*, *18*, 766–780.
- Crosson, P. L., Walton, M. E., O'Reilly, J. X., Behrens, T. E., & Rushworth, M. F. (2009). Effort-based cost-benefit valuation and the human brain. *Journal of Neuroscience*, *29*, 4531–4541.
- De Pisapia, N., & Braver, T. S. (2006). A model of dual control mechanisms through anterior cingulate and prefrontal cortex interactions. *Neurocomputing*, *69*, 1322–1326.
- Devinsky, O., Morrell, M. J., & Vogt, B. (1995). Contributions of anterior cingulate cortex to behavior. *Brain*, *118*, 279–306.
- Donoso, M., Collins, A. G. E., & Koehlin, E. (2014). Human cognition. Foundations of human reasoning in the prefrontal cortex. *Science*, *344*, 1481–1486.
- Eblen, F., & Graybiel, A. M. (1995). Highly restricted origin of prefrontal cortical inputs to striosomes in the macaque monkey. *Journal of Neuroscience*, *15*, 5999–6013.
- Ferdinand, N. K., Mecklinger, A., Kray, J., & Gehring, W. J. (2012). The processing of unexpected positive response outcomes in the mediofrontal cortex. *Journal of Neuroscience*, *32*, 12087–12092.
- Fitzgerald, K. D., Welsh, R. C., Gehring, W. J., Abelson, J. L., Himle, J. A., Liberzon, I., et al. (2005). Error-related hyperactivity of the anterior cingulate cortex in obsessive-compulsive disorder. *Biological Psychiatry*, *57*, 287–294.
- Forster, S. E., & Brown, J. W. (2011). Medial prefrontal cortex predicts and evaluates the timing of action outcomes. *Neuroimage*, *55*, 253–265.
- Fukunaga, R., Brown, J. W., & Bogg, T. (2012). Decision making in the Balloon Analogue Risk Task (BART): Anterior cingulate cortex signals loss aversion but not the infrequency of risky choices. *Cognitive, Affective, & Behavioral Neuroscience*, *12*, 479–490.
- Garofalo, S., Maier, M. E., & di Pellegrino, G. (2014). Mediofrontal negativity signals unexpected omission of aversive events. *Scientific Reports*, *4*, 4816.
- Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science*, *4*, 385–390.
- Gemba, H., Sasaki, K., & Brooks, V. B. (1986). "Error" potentials in limbic cortex (anterior cingulate area 24) of monkeys during motor learning. *Neuroscience Letters*, *70*, 223–227.
- Grinband, J., Savitskaya, J., Wager, T. D., Teichert, T., Ferrera, V. P., & Hirsch, J. (2011). The dorsal medial frontal cortex is sensitive to time on task, not response conflict or error likelihood. *Neuroimage*, *57*, 303–311.
- Hajihosseini, A., & Holroyd, C. B. (2013). Frontal midline theta and N200 amplitude reflect complementary information about expectancy and outcome evaluation. *Psychophysiology*, *50*, 550–562.
- Hayden, B. Y., Pearson, J. M., & Platt, M. L. (2011). Neuronal basis of sequential foraging decisions in a patchy environment. *Nature Neuroscience*, *14*, 933–939.
- Hochman, E. Y., Vaidya, A. R., & Fellows, L. K. (2014). Evidence for a role for the dorsal anterior cingulate cortex in disengaging from an incorrect action. *PLoS One*, *9*, e101126.
- Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*, 679–709.
- Holroyd, C. B., & McClure, S. M. (2015). Hierarchical control over effortful behavior by rodent medial frontal cortex. *Psychological Review*, *122*, 54–83.
- Holroyd, C. B., & Yeung, N. (2012). Motivation of extended behaviors by anterior cingulate cortex. *Trends in Cognitive Sciences*, *16*, 122–128.
- Holroyd, C. B., Yeung, N., Coles, M. G., & Cohen, J. D. (2005). A mechanism for error detection in speeded response time tasks. *Journal of Experimental Psychology: General*, *134*, 163–191.
- Ide, J. S., Shenoy, P., Yu, A. J., & Li, C. R. (2013). Bayesian prediction and evaluation in the anterior cingulate cortex. *Journal of Neuroscience*, *33*, 2039–2047.
- Jahn, A., Nee, D. E., Alexander, W. H., & Brown, J. W. (2014). Distinct regions of anterior cingulate cortex signal prediction and outcome evaluation. *Neuroimage*, *95*, 80–89.
- Jahn, A., Nee, D. E., Alexander, W. H., & Brown, J. W. (2016). Distinct regions within medial prefrontal cortex process pain and cognition. *Journal of Neuroscience*, *36*, 12385–12392.
- Jessup, R. K., Bussemeyer, J. R., & Brown, J. W. (2010). Error effects in anterior cingulate cortex reverse when error likelihood is high. *Journal of Neuroscience*, *30*, 3467–3472.
- Kawai, T., Yamada, H., Sato, N., Takada, M., & Matsumoto, M. (2015). Roles of the lateral habenula and anterior cingulate cortex in negative outcome monitoring and behavioral adjustment in nonhuman primates. *Neuron*, *88*, 792–804.
- Kennerly, S. W. (2003). Organization of action sequences and the role of the pre-SMA. *Journal of Neurophysiology*, *91*, 978–993.
- Kennerly, S. W., Behrens, T. E., & Wallis, J. D. (2011). Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nature Neuroscience*, *14*, 1581–1589.
- Kennerly, S. W., Walton, M. E., Behrens, T. E., Buckley, M. J., & Rushworth, M. F. (2006). Optimal decision making and the anterior cingulate cortex. *Nature Neuroscience*, *9*, 940–947.
- Kolling, N., Behrens, T., Wittmann, M., & Rushworth, M. (2016). Multiple signals in anterior cingulate cortex. *Current Opinion in Neurobiology*, *37*, 36–43.
- Kolling, N., Behrens, T. E. J., Mars, R. B., & Rushworth, M. F. S. (2012). Neural mechanisms of foraging. *Science*, *336*, 95–98.
- Koyama, T., Kato, K., Yanaka, Y. Z., & Mikami, A. (2001). Anterior cingulate activity during pain-avoidance and reward tasks in monkeys. *Neuroscience Research*, *39*, 421–430.
- Krawitz, A., Fukunaga, R., & Brown, J. W. (2010). Anterior insula activity predicts the influence of positively framed messages on decision making. *Cognitive, Affective, & Behavioral Neuroscience*, *10*, 392–405.
- Lesh, T. A., Westphal, A. J., Niendam, T. A., Yoon, J. H., Minzenberg, M. J., Ragland, J. D., et al. (2013). Proactive and reactive cognitive control and dorsolateral prefrontal cortex dysfunction in first episode schizophrenia. *Neuroimage: Clinical*, *2*, 590–599.
- Lieberman, M. D., & Eisenberger, N. I. (2015). The dorsal anterior cingulate cortex is selective for pain: Results from large-scale reverse inference. *Proceedings of the National Academy of Sciences, U.S.A.*, *112*, 15250–15255.
- Ljungberg, T., Apicella, P., & Schultz, W. (1992). Responses of monkey dopamine neurons during learning of behavioral reactions. *Journal of Neurophysiology*, *67*, 145–163.

- MacDonald, A. W., Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000). Dissociating the role of the dorsolateral prefrontal cortex and anterior cingulate cortex in cognitive control. *Science*, *288*, 1835–1838.
- Marini, F., Demeter, E., Roberts, K. C., Chelazzi, L., & Woldorff, M. G. (2016). Orchestrating proactive and reactive mechanisms for filtering distracting information: Brain-behavior relationships revealed by a mixed-design fMRI study. *Journal of Neuroscience*, *36*, 988–1000.
- Medalla, M., & Barbas, H. (2009). Synapses with inhibitory neurons differentiate anterior cingulate from dorsolateral prefrontal pathways associated with cognitive control. *Neuron*, *61*, 609–620.
- Narayanan, N. S., Cavanagh, J. F., Frank, M. J., & Laubach, M. (2013). Common medial frontal mechanisms of adaptive control in humans and rodents. *Nature Neuroscience*, *16*, 1888–1895.
- Oliveira, F. T., McDonald, J. J., & Goodman, D. (2007). Performance monitoring in the anterior cingulate is not all error related: Expectancy deviation and the representation of action-outcome associations. *Journal of Cognitive Neuroscience*, *19*, 1994–2004.
- Parvizi, J., Rangarajan, V., Shirer, W. R., Desai, N., & Greicius, M. D. (2013). The will to persevere induced by electrical stimulation of the human cingulate gyrus. *Neuron*, *80*, 1359–1367.
- Paulus, M. P., & Frank, L. R. (2006). Anterior cingulate activity modulates nonlinear decision weight function of uncertain prospects. *Neuroimage*, *30*, 668–677.
- Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science*, *306*, 443–447.
- Rudebeck, P. H., Behrens, T. E., Kennerley, S. W., Baxter, M. G., Buckley, M. J., Walton, M. E., et al. (2008). Frontal cortex subregions play distinct roles in choices between actions and stimuli. *Journal of Neuroscience*, *28*, 13775–13785.
- Rushworth, M. F. S., Kolling, N., Sallet, J., & Mars, R. B. (2012). Valuation and decision-making in frontal cortex: One or many serial or parallel systems? *Current Opinion in Neurobiology*, *22*, 946–955.
- Sallet, J., Quilodran, R., Rothe, M., Vezoli, J., Joseph, J. P., & Procyk, E. (2007). Expectations, gains, and losses in the anterior cingulate cortex. *Cognitive Affective & Behavioral Neuroscience*, *7*, 327–336.
- Schall, J. D., Stuphorn, V., & Brown, J. W. (2002). Monitoring and control of action by the frontal lobes. *Neuron*, *36*, 309–322.
- Schmid, P. C., Kleiman, T., & Amodio, D. M. (2015). Neural mechanisms of proactive and reactive cognitive control in social anxiety. *Cortex*, *70*, 137–145.
- Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593–1599.
- Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron*, *79*, 217–240.
- Shenhav, A., Straccia, M. A., Botvinick, M. M., & Cohen, J. D. (2016). Dorsal anterior cingulate and ventromedial prefrontal cortex have inverse roles in both foraging and economic choice. *Cognitive, Affective & Behavioral Neuroscience*, *16*, 1127–1139.
- Shenhav, A., Straccia, M. A., Cohen, J. D., & Botvinick, M. M. (2014). Anterior cingulate engagement in a foraging context reflects choice difficulty, not foraging value. *Nature Neuroscience*, *17*, 1249–1254.
- Shima, K., & Tanji, J. (1998). Role of cingulate motor area cells in voluntary movement selection based on reward. *Science*, *282*, 1335–1338.
- Skvortsova, V., Palminteri, S., & Pessiglione, M. (2014). Learning to minimize efforts versus maximizing rewards: Computational principles and neural correlates. *Journal of Neuroscience*, *34*, 15621–15630.
- Soshi, T., Ando, K., Noda, T., Nakazawa, K., Tsumura, H., & Okada, T. (2015). Post-error action control is neurobehaviorally modulated under conditions of constant speeded response. *Frontiers in Human Neuroscience*, *8*, 1072.
- Talmi, D., Atkinson, R., & El-Dereedy, W. (2013). The feedback-related negativity signals salience prediction errors, not reward prediction errors. *Journal of Neuroscience*, *33*, 8264–8269.
- Verbruggen, F., Adams, R., & Chambers, C. D. (2012). Proactive motor control reduces monetary risk taking in gambling. *Psychological Science*, *23*, 805–815.
- Wager, T. D., Atlas, L. Y., Botvinick, M. M., Chang, L. J., Coghill, R. C., Davis, K. D., et al. (2016). Pain in the ACC? *Proceedings of the National Academy of Sciences, U.S.A.*, *113*, E2474–E2475.
- Walton, M. E., Bannerman, D. M., & Rushworth, M. F. (2002). The role of rat medial frontal cortex in effort-based decision making. *Journal of Neuroscience*, *22*, 10996–11003.
- Walton, M. E., Groves, J., Jennings, K. A., Croxson, P. L., Sharp, T., Rushworth, M. F. S., et al. (2009). Comparing the role of the anterior cingulate cortex and 6-hydroxydopamine nucleus accumbens lesions on operant effort-based decision making. *European Journal of Neuroscience*, *29*, 1678–1691.
- Wessel, J. R., Danielmeier, C., Morton, J. B., & Ullsperger, M. (2012). Surprise and error: Common neuronal architecture for the processing of errors and novelty. *Journal of Neuroscience*, *32*, 7528–7537.
- Wittmann, M. K., Kolling, N., Akaishi, R., Chau, B. K., Brown, J. W., Nelissen, N., et al. (2016). Predictive decision making driven by multiple time-linked reward representations in the anterior cingulate cortex. *Nature Communications*, *7*, 12327.
- Yeung, N., & Nieuwenhuis, S. (2009). Dissociating response conflict and error likelihood in anterior cingulate cortex. *Journal of Neuroscience*, *29*, 14506–14510.
- Ziegler, M. D., Chelian, S. E., Benvenuto, J., Krichmar, J. L., O'Reilly, R., & Bhattacharyya, R. (2014). A model of proactive and reactive cognitive control with anterior cingulate cortex and the neuromodulatory system. *Biologically Inspired Cognitive Architectures*, *10*, 61–67.