nature neuroscience

Medial prefrontal cortex as an action-outcome predictor

William H Alexander & Joshua W Brown

The medial prefrontal cortex (mPFC) and especially anterior cingulate cortex is central to higher cognitive function and many clinical disorders, yet its basic function remains in dispute. Various competing theories of mPFC have treated effects of errors, conflict, error likelihood, volatility and reward, using findings from neuroimaging and neurophysiology in humans and monkeys. No single theory has been able to reconcile and account for the variety of findings. Here we show that a simple model based on standard learning rules can simulate and unify an unprecedented range of known effects in mPFC. The model reinterprets many known effects and suggests a new view of mPFC, as a region concerned with learning and predicting the likely outcomes of actions, whether good or bad. Cognitive control at the neural level is then seen as a result of evaluating the probable and actual outcomes of one's actions.

The medial prefrontal cortex (mPFC) is critically involved in both higher cognitive function and psychopathology[1], yet the nature of its function remains in dispute. No one theory has been able to account for the variety of mPFC effects observed with a broad range of methods. Initial event-related potential (ERP) findings of an error-related negativity (ERN)[2,3] have been reinterpreted with human neuroimaging studies to reflect a response conflict detector[4], and the conflict model[5] has been influential despite some controversy. Nonetheless, monkey neurophysiological studies have found mixed evidence for pure conflict detection[6,7] and have instead highlighted reinforcement-like reward and error signals[7–11]. Theories of mPFC function have multiplied beyond response conflict theories to include detecting discrepancies between actual and intended responses[12] or outcomes[7,13], predicting error likelihood[14,15], detecting environmental volatility[16] and predicting the value of actions[17,18]. The diversity of findings and theories has led some to question whether the mPFC is functionally equivalent across humans and monkeys[19], despite the similar effects seen with functional magnetic resonance imaging (fMRI) for comparable tasks in monkey and human mPFC[20]. Thus, a central open question is whether all of these varied findings can be accounted for by a single theoretical framework. If so, the strongest test of a theory is whether it can provide a rigorous quantitative account and yield useful predictions. In this paper we aim to provide such a quantitative model account.

The model begins with the premise that the medial prefrontal cortex (mPFC), and especially the dorsal aspects, may be central to forming expectations about actions and detecting surprising outcomes[21]. A growing body of literature casts mPFC as learning to anticipate the value of actions. This requires both a representation of possible outcomes and a training signal to drive learning as contingencies change[16]. New evidence suggests that mPFC represents the various likely outcomes of actions, whether positive[9], negative[14,15] or both[22,23], and signals a composite cost-benefit analysis[24,25]. This proposed function of mPFC as anticipating action values[17,18] is distinct from the role of orbitofrontal cortex in signaling stimulus values[26]. For mPFC to learn outcome predictions in a changing environment, a mechanism is needed to detect discrepancies between actual and predicted outcomes and update the outcome predictions appropriately. Several studies suggest that mPFC, and anterior cingulate cortex (ACC) in particular, signal such discrepancies[7,10,27,28]. Recent work further suggests that distinct effects of error detection, prediction and conflict are localized to the anterior and posterior rostral cingulate zones[29].

Given the above, we propose a new theory and model of mPFC function, the predicted response–outcome (PRO) model (**Fig. 1a**), to reconcile these findings. The model suggests that individual neurons generate signals reflecting a learned prediction of the probability and timing of the various possible outcomes of an action. These prediction signals are inhibited when the corresponding predicted outcome actually occurs. The resulting activity is therefore maximal when an expected outcome fails to occur, which suggests that what mPFC signals, in part, is the unexpected non-occurrence of a predicted outcome.

At its core, the PRO model is a generalization of standard reinforcement learning algorithms

$$\delta_t = r_{t+1} + \gamma V_{t+1} - V_t \quad (1)$$

that compute a temporal prediction error, δ, reflecting the discrepancy between a reward prediction, V, on successive time steps t and $t + 1$, and the actual amount of reward, r. The temporal discount factor $\gamma (0 < \gamma < 1)$ describes how the value of delayed rewards is reduced.

The PRO model builds on reinforcement learning as a representative learning law, but this should not be taken to imply that mPFC does reinforcement learning *per se*. The PRO model differs from standard reinforcement learning algorithms in four ways. First, in contrast to typical reinforcement learning algorithms, the PRO model does not primarily train stimulus–response mappings. Instead, it maps existing action plans in a stimulus context to predictions of the responses and outcomes that are likely to result—that is, response–outcome learning. This change to standard reinforcement learning conforms well

Department of Psychological and Brain Sciences, Indiana University, Bloomington, Indiana, USA. Correspondence should be addressed to J.W.B. (jwmbrown@indiana.edu).

Received 18 January; accepted 20 July; published online 18 September 2011; doi:10.1038/nn.2921

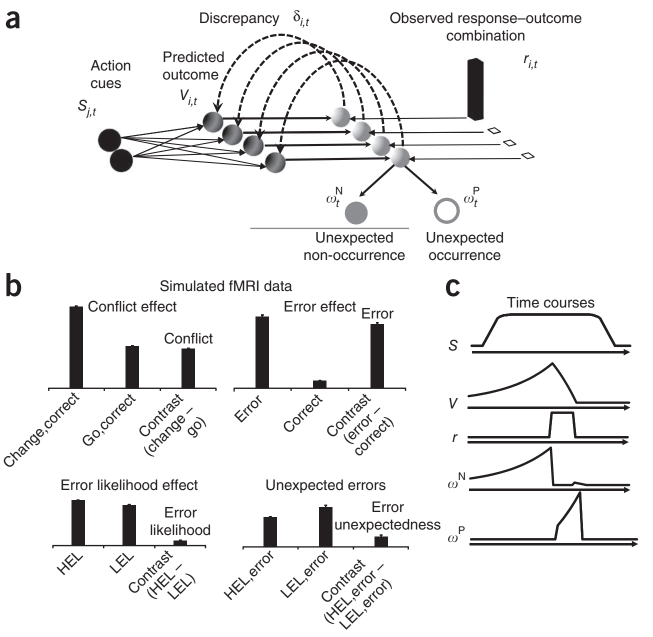


Figure 1 The PRO model. (**a**) In an idealized experiment, a task-related stimulus (*S*) signaling the onset of a trial is presented. Over the course of a task, the model learns a timed prediction (*V*) of possible responses and outcomes (*r*). The temporal difference learning signal (δ) is decomposed into its positive and negative components (ω^P and ω^N, respectively), indicating unpredicted occurrences and unpredicted non-occurrences, respectively. (**b**) ω^N accounts for typical effects observed in mPFC from human imaging studies. Conflict and error likelihood panels show activity magnitude aligned on trial onset; error and error unexpectedness panels show activity magnitude aligned on feedback. Model activity (vertical axis) is in arbitrary units. HEL, high error likelihood; LEL, low error likelihood. Error bars indicate s.e.m. Contrasts indicate the difference in model activity between two conditions. (**c**) Typical time courses for components of the PRO model.

to reports of single units in macaque ACC that learn action–outcome relationships[10,18,30]. Second, instead of a typical scalar prediction of future rewards and scalar prediction error, the PRO model implements a vector-valued prediction, V_i, and prediction error, δ_i, reflecting the hypothesized mPFC role in monitoring multiple potential outcomes, indexed by *i*. This allows multiple possible action outcomes to be predicted simultaneously, each with a corresponding probability. Previous influential models of mPFC[13,31], similarly derived from reinforcement learning, use scalar value and error signals that represent, respectively, a prediction and subsequent prediction error of reward. In these models, and in reinforcement learning in general, positive value and error signals represent affectively positive outcomes, whereas negative value and error signals represent affectively negative outcomes. In contrast, the PRO model maintains separate predictions of all possible outcomes, including both rewarding and aversive outcomes. The signed vector prediction error, then, represents unexpected occurrences (positive) or unexpected non-occurrences (negative)—regardless of whether these events are rewarding or aversive—and the purpose of these prediction error signals is to provide a training signal to update the predictions of response outcomes. Third, rather than the typical reward signal used in standard reinforcement learning, the model uses a vector signal r_i that reflects the actual response and outcome combination, again whether good or bad. This enables the PRO model to predict response–outcome conjunctions in proportion to the probability of their occurrence, similarly to the error likelihood model[15], with the addition that the PRO model learns representations not only of rewarding but also of aversive events (for more detail, see **Supplementary Note**). Fourth, and most crucial to the model's ability to account for a wide range of empirical findings, the model specifically detects the rectified negative prediction error, defined as the signal generated when an expected event fails to occur (whether good or bad); for example, a reward that is unexpectedly absent. To detect such events, the model computes negative surprise, ω^N, which reflects the probability of an expected outcome that nevertheless did not occur (that is, unexpected non-occurrence):

$$\omega_t^N = \sum_i \mathrm{MAX}(\mathrm{Expected} - \mathrm{Actual}, 0) = \sum_i \mathrm{MAX}(V_{i,t} - r_{i,t}, 0) \qquad (2)$$

The quantity ω^N reflects the aggregate activity of individual units that compare actual outcomes against the probability of expected response–outcome conjunctions. In equation (2), when the probability of an expected event is higher, its failure to occur leads to a larger negative surprise signal. mPFC activity, then, indexes the extent to which experienced outcomes fail to correspond with outcomes that are predicted—that is, negative surprise.

Although several of the ideas underlying the PRO model have been presented previously in some form, we are not aware of any effort that has brought these ideas to bear simultaneously on the diverse effects observed in mPFC. The contribution of this paper, then, is twofold. First, we propose a hypothesis that suggests that mPFC signals unexpected non-occurrences of predicted outcomes. Second, we demonstrate that the proposed role of mPFC in monitoring observed outcomes and comparing them against predicted outcomes can account for an unprecedented array of cognitive control, behavioral, neuroimaging, ERP and single-unit neurophysiological findings, and also provide a priori predictions for future empirical studies.

RESULTS

Representative tasks

To test the ability of the PRO model to account for a diverse range of empirical results, we selected two representative tasks to simulate: the change signal task and the Eriksen flanker task. These tasks have been widely used in the context of both behavioral and imaging methods, and they reliably elicit markers of cognitive control, including increases in reaction time and error rate in behavioral data, and increased activity in brain regions associated with control in imaging data.

At the start of a trial in the change signal task (simulations 1, 2, 4, 5 and 9), a subject is cued to make one of two behavioral responses. On a subset of trials, a second change cue will be displayed shortly following the original cue, instructing the subject to cancel the original response and instead make the alternative response. By manipulating the delay between the original cue and the change cue, specific overall error rates can be obtained.

In the Eriksen flanker task (simulations 3 and 7), subjects are cued to make one of two behavioral responses by a central target stimulus. Distractor cues are presented simultaneously on both sides of the central stimulus. On congruent trials, the distractors cue the same response as the target cue, whereas on incongruent trials, the distractors cue the alternative response.

Additionally, to test the sensitivity of the PRO model to environmental volatility effects[16], we simulate the model in a two-armed bandit task (simulation 6) similar to a previous report. In the two-armed bandit task, subjects repeatedly choose from one of two options that yield rewards at preset rates for each option. In the task simulated, this rate shifts over the course of the experiment, with each option alternately yielding rewards at a high frequency or low frequency.

Table 1 Model parameters

Parameter	Description	Value	Equation
α	Learning rate	0.012	7
Γ	Response threshold	0.313	14
ρ	Input scaling factor	1.764	12
ϕ	Control signal scaling factor	2.246	13
ψ	Mutual inhibition scaling factor	0.724	13
β	Rate coding scaling factor	1.038	11
σ	Variance of noise in control units	0.005	11

Our first goal was to ensure that the PRO model could replicate the basic effects observed in mPFC with these tasks and captured by competing models, including error, conflict and error likelihood effects, as well as the error-related negativity and its relation to speed–accuracy tradeoffs. Second, we sought to show that the PRO model can account for additional data that are not addressed by competing models, including single-unit activity from monkey neurophysiological studies. To ensure that the effects observed in the PRO model do not depend on a specific, manually tuned parameterization, we initially fit the model to behavioral data from the change signal task. Because the model was only fit to behavioral data, all model predictions of ERP, fMRI and monkey neurophysiology results should be considered qualitative predictions rather than quantitative fits. Except where noted, all simulations reported derive from the model with this single parameter set (**Table 1**). More details regarding the simulations are given in the Online Methods.

Simulation 1: error, conflict and error likelihood effects

In our first simulation, we showed that the PRO model could reproduce effects of error, error likelihood and conflict using the change signal task. Over the course of the simulation, the PRO model generates a negative surprise signal corresponding to these effects (**Fig. 1b**,**c**). The intuition behind error effects is that a correct outcome was predicted, but that that prediction signal was not suppressed by signals of an actual correct outcome. Hence the error effect reflects negative surprise—that is, an unexpected non-occurrence of a correct outcome. Moreover, error effects in the model were stronger for errors made in conditions of low error likelihood, consistent with fMRI results not accounted for by previous models[14,15]. The PRO model accounts for this effect because activity predicting a correct response is greater when error likelihood is low. Thus the absence of a correct outcome when a correct outcome is very likely yields stronger negative surprise.

This reasoning applies equally well to findings that the ERN is observed to be larger on error trials in congruent conditions in an Eriksen flanker task[12]. For conflict effects, the intuition is that incongruent stimuli signal a prediction of responding to the distractor, in addition to the already strong prediction of a correct response, hence greater aggregate prediction-related activity. The same logic accounts for error likelihood effects: activity representing the prediction of a correct response button-press is already high, and as the probability of an error increases, the activity predicting an additional button-press of the incorrect response also increases proportionally, hence greater aggregate prediction-related activity. Of note, the model suggests a reinterpretation of response conflict effects as not reflecting conflict *per se*. Rather, conflict effects in the model are due to the presence of a greater prediction of multiple responses, namely the correct and incorrect responses (simulation 5 below).

Simulation 2: error-related negativity

One of the earliest findings in medial prefrontal cortex is the ERN[2,3,13] and the related feedback ERN (fERN)[13,32], in which the scalp potential overlying mPFC is significantly more negative for errors than for correct responses or outcomes. The PRO model simulates the difference-wave fERN, which is not confounded with the P300 (a positive-going ERP component with a 300 ms latency; ref. 31), as the negative surprise at each time step during a trial. **Figure 2a** shows the simulated fERN compared with an actual ERN[31]. The model not only qualitatively simulates the fERN but also simulates the increasing size of the fERN in proportion to the unexpectedness of the error.

Simulation 3: speed-accuracy tradeoff and the N2

Recent attempts to distinguish between conflict and error likelihood accounts of mPFC function find that the amplitude of the N2 component of the ERP, associated with increased cognitive demand and originating in ACC, reflects the widely observed speed–accuracy tradeoff[33]. The conflict account of the N2 suggests trials with longer reaction times reflect longer ongoing competition between potential responses, resulting in higher levels of conflict than for trials with short reaction times (although this explanation is not without controversy[34]). In contrast, the PRO model intuition for this effect is that

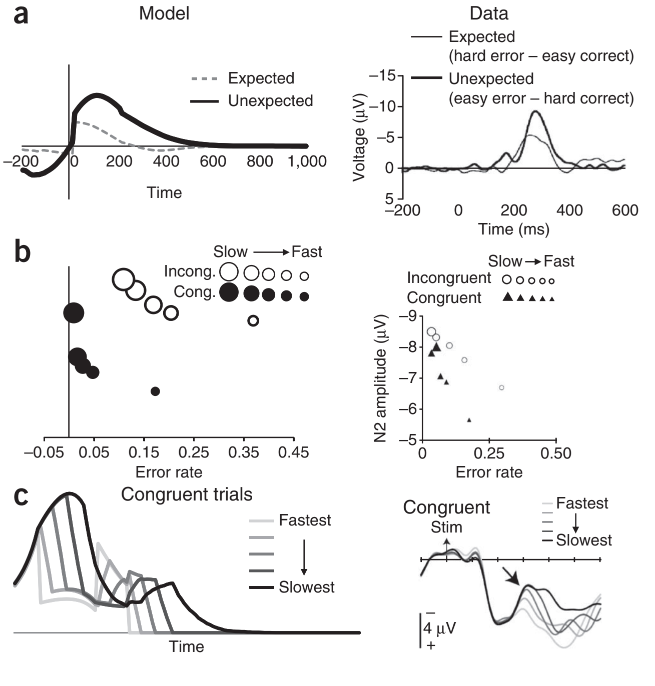


Figure 2 ERP simulations. (**a**) Left: simulated fERN difference wave. Effects of surprising outcomes (low error likelihood,error – high error likelihood,correct) were larger than outcomes that were predictable (high error likelihood,error minus low error likelihood,correct). Right: observed ERP difference wave, adapted with permission from ref. 31, consistent with simulation results. ("Hard" and "easy" indicate task difficulty). (**b**) The effects of speed–accuracy tradeoffs on ERP amplitude are observed in the PRO model (left). Trials for incongruent (incong.) and congruent (cong.) conditions were divided into quintile bins by reaction time (large markers, slow reaction times; small markers, fast reaction times), and activity of the PRO model was calculated for correct trials in each bin. Accuracy and activity of the model were highest for trials with long reaction times and lowest for trials with short reaction times, consistent with human EEG data (right; adapted with permission from ref. 33). (**c**) The simulated activity of the PRO model (left) reflects amplitude and duration of the N2 component observed in humans EEG studies (right; aligned on stimulus onset (Stim); adapted with permission from ref. 33). Model activity (vertical axis) is in arbitrary units.

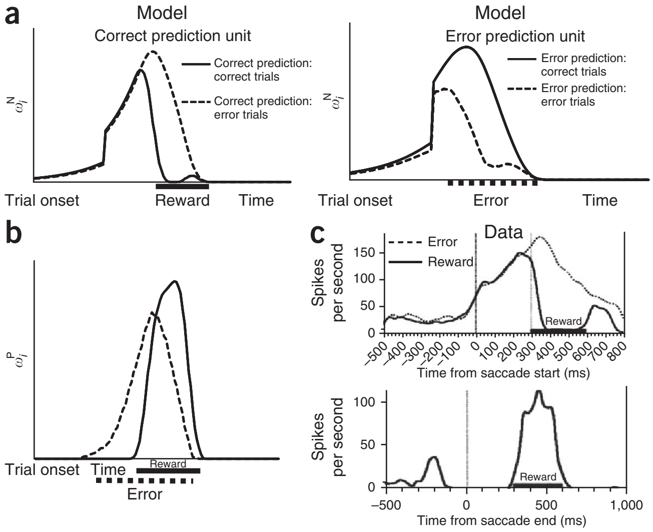


Figure 3 Single-unit neurophysiology simulation. (**a**) Calculation of the negative surprise signal ω^N was performed for individual outcome predictions (indexed as *i*). For predictions of, for example, reward, the surprise signal increases steadily to the time at which the reward is predicted. The signal is suppressed on the occurrence of the predicted reward. Single units predicting error follow a similar pattern, with increased variance in the timing of the error. (**b**) The complement of negative surprise (namely, positive surprise ω^P) indicates unpredicted occurrences. Model activity (vertical axis) is in arbitrary units. (**c**) Reward-predicting and reward-detecting cells recorded in monkey mPFC consistent with simulation results. Top: activity of a single unit consistent with the prediction of a reward. On error trials, activity peaks and gradually attenuates, potentially signaling an unsatisfied prediction of reward. Bottom: single-unit activity related to the detection of a rewarding event. Adapted with permission from ref. 28.

longer reaction times also entail a greater period of time during which the expectation of a correct response is unmet, which in turn yields larger N2 signals. Thus, the model accounts for N2 amplitude effects as a simple positive correlation with reaction time. The PRO model simulates the speed–accuracy tradeoff in a simulated version of a flanker task (**Fig. 2b**); the negative surprise component of the PRO model is greatest for trials with a relatively long reaction time and is higher for incongruent than congruent trials, as in simulation 1. The correlation of the simulated amplitude with error rate for the congruent ($r = -0.725$) and incongruent ($r = -0.863$) trials corresponds well with the pattern observed in previously reported data from humans[33]. The model further captures how the temporal profile of the N2 component varies with reaction time[33] (**Fig. 2c**).

Simulation 4: monkey single-unit performance monitoring data

Using the change signal task above, we also compared the model predictions with monkey single-unit neurophysiological data. A key challenge to the conflict model of mPFC has been the lack of evidence showing single-unit activity related to conflict in monkey ACC[7]. In contrast, by maintaining multiple predictions of specific response–outcome combinations, single units in the PRO model show activity similar to that of reward- and error-predicting neurons observed in single-unit neurophysiological data. **Figure 3** shows the average time course of negative surprise (ω^N) and its complement, positive surprise (ω^P, the unexpected occurrence of an outcome; see Online Methods), which can each reflect predictions of either reward or error outcomes. Like activity in monkey supplementary eye field[28] (**Fig. 3c**), signals related to the prediction of reward increased steadily before the expected time of reward (**Fig. 3a**, left). On trials in which the reward was delivered as expected, the negative surprise was suppressed, whereas on trials in which the reward was not delivered, ω^N peaked around the time of expected outcome and gradually decayed. Surprise related to error prediction showed a similar pattern (**Fig. 3a**, right). Owing to the nature of learned temporal predictions in the model, at equilibrium, activity in reward-predicting cells will be proportional to the average probability of predicted reward associated with an outcome[27,35], and activity of error-predicting cells will be proportional to the average probability of error associated with an action. Regarding positive surprise, neurons in mPFC seem to respond to the detection of unpredicted events (**Fig. 3b**), and the strength with which they respond moderates as the event becomes more predictable[10,28,36].

Simulation 5: conflict effects as due to multiple responses

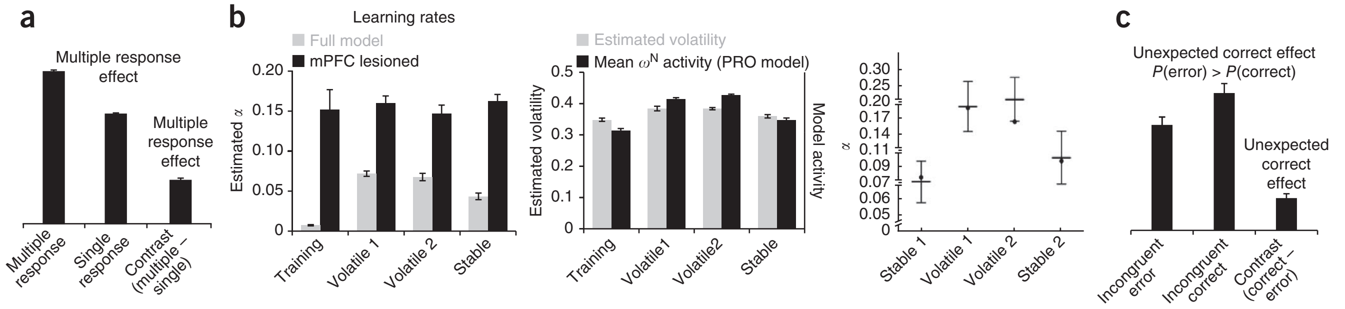
The computation underlying response conflict effects in mPFC has been disputed. Early models cast conflict as a multiplication of two mutually incompatible response processes[5]. More recent studies suggest that conflict may arise from having a greater number of responses—regardless of mutual incompatibility[37,38]. In a recent study[37], both the Eriksen flanker task and the change signal task[15] were modified to require simultaneous responses to both distracters and target stimuli. The results showed similar ACC activation in the same region for conditions in which the two possible responses were mutually incompatible to that seen when the responses were required to be executed simultaneously. This suggests that mPFC may signal a greater number of predicted or actual responses or outcomes instead of a response conflict *per se*, as found previously with neurophysiological studies[38].

The PRO model simulates these findings (**Fig. 4a**) with a modification of the change signal task in which lateral inhibition between response units is removed (see Online Methods), allowing both responses to be generated simultaneously when a change signal is presented. The PRO model then learns to associate 'go' signals with a high probability of the corresponding anticipated left or right motor response. On trials with a change signal, the PRO model generates an additional prediction of the other motor response, which yields an overall net increase in signals predicting the correspondingly greater number of motor responses.

Simulation 6: volatility

A recent Bayesian model of ACC[16] suggests that ACC activity reflects the estimated volatility (non-stationarity) of reinforcement contingencies of an environment. Subjects choosing between two gambles were found to more quickly adapt their strategies (that is, they learned faster) when the probabilities underlying the gambles changed frequently. Moreover, activity in ACC tracked environmental volatility and was higher for subjects with higher estimated learning rates.

The PRO model fits the observed pattern of greater mPFC activity in volatile environments (ω^N, **Fig. 4b**, bottom left). Essentially, as contingencies change, the outcome predictions based on the previous contingency persist even as new predictions form based on the new contingencies. As reversals occur, predictions of outcomes made by the PRO model are frequently upset, leading to a state of constant surprise and resulting in more frequent but weaker ω^N signals. This pattern indicates environmental volatility and also serves to drive increased learning during periods of shifting environmental contingencies (**Fig. 4b**, top left).



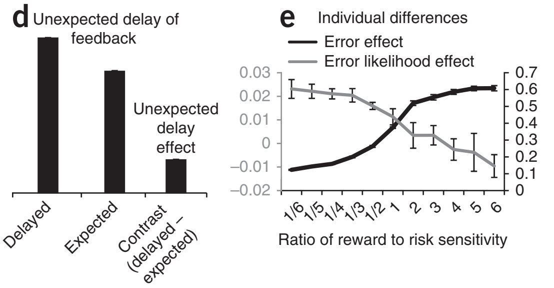


Figure 4 fMRI simulations. Except where noted below, vertical axes represent model activity in arbitrary units. Contrasts indicate the difference in model activity between two conditions. (**a**) Multiple response effects. The change signal task is modified to require both change and go responses simultaneously when a change signal cue is presented. Change trials lead to greater prediction layer activity (aligned on trial onset) compared with go trials, even though response conflict is by definition absent. The incongruency effect in the absence of conflict is the multiple response effect[23]. (**b**) Volatility effects. When environmental contingencies change frequently (Volatile 1 & 2), mPFC shows greater activity than in non-volatile conditions (Training & Stable). Vertical axes in the left panel represent the equivalent learning rate of a reinforcement learning model (see **Supplementary Note**). Vertical axes in the center panel indicate a Bayesian estimate of volatility (left axis, gray bars) and model activity in arbitrary units (right axis, black bars). This has been interpreted with a Bayesian model in which mPFC signals the expected volatility (right panel; bars indicate human behavior, circles indicate behavior of a Bayesian model, and the vertical axis represents the equivalent learning rate of a reinforcement learning model; adapted with permission from ref. 16). In the PRO model, greater volatility in a block led to greater mean ω^N (center). Surprise signals, in turn, dynamically modulated the effective learning rate of the model (left), yielding lower effective learning rates (see **Supplementary Note**) during periods of greater stability ($F_{1,3} = 70.3$, $P = 4.0 \times 10^{-15}$). In the mPFC-lesioned model, learning rates did not significantly change between periods ($F_{1,3} = 0.23$, $P = 0.88$). (**c**) mPFC signals discrepancies between actual and expected outcomes. If errors occur more frequently than correct trials (in this case, 70% error rate), mPFC is predicted to show an inversion of the error effect—that is, greater activity (aligned on feedback) for correct than for error trials. (**d**) Delayed feedback effect. Feedback that is delayed an extra 400 ms on a minority of trials (20% here) leads to timing discrepancies and greater surprise activation (aligned on feedback). (**e**) Effects of reward salience on error prediction and detection. As rewarding events influence learning to a greater degree, error likelihood effects (aligned on trial onset) decrease while error effects (aligned on feedback) increase. The error and error likelihood effects are calculated as contrasts (as in **a**) and given in arbitrary units. All error bars indicate s.e.m.

Simulation 7: mPFC activity reflects unexpected outcomes

The PRO model reinterprets error effects in mPFC as unexpected outcomes, as distinct from outcomes that are merely undesired. In most human studies, error rates are low. This confounds the interpretation of errors as unintended outcomes with errors as unexpected outcomes. These theories can be distinguished by a manipulation that causes error outcomes to be more likely than correct outcomes. In that case, an error may be expected as the most likely outcome even though it is unintended. If errors reflect unexpected outcomes, then error signals should reverse if correct outcomes are infrequent and therefore unexpected, and correct trials should instead yield greater 'error'-related activation in mPFC than error trials, and in the same mPFC regions that show error effects.

Using a flanker task in which the error rate for incongruent trials was much higher than the rate of correct responses, we tested this prediction and found a notable reversal of the error effect (**Fig. 4c**), consistent with recent findings[39,40]. This result presents a clear challenge to both the conflict account of mPFC function and models of mPFC that are based on standard formulations of reinforcement learning. It is not clear how the conflict account of the ERN can accommodate increased activity in mPFC after correctly executed trials in which behavioral conflict is presumed to be lower than for incorrect trials. Similarly, previous models based on reinforcement learning suggest that mPFC activity reflects only the detection and processing of errors. It is unclear how such a model could account for increased activity in response to correct trials relative to error trials.

Simulation 8: ACC signals unexpected timing of feedback

Single units have been observed in ACC that show precisely timed patterns of activation before the occurrence of an outcome[28,41]. The PRO model can show activity consistent with such timed predictions (for example, **Fig. 3a**). A further prediction of the model, then, is that outcomes that occur at unexpected times, even if the outcomes themselves are predicted, will lead to increased ACC activity (**Fig. 4d**). This prediction suggests another means by which the PRO model may be differentiated from the conflict account, and further experimental work is needed to test this prediction of the PRO model.

Simulation 9: individual differences

We tested the effect of the salience of rewarding versus aversive outcomes by parametrically adjusting the relative influence on learning of error and correct outcomes in the change signal task. The PRO model predicts that individuals who are particularly attentive to rewarding outcomes will show greater mPFC activity in response to error trials (**Fig. 4e**) than individuals who are sensitive to aversive outcomes, whereas reward-sensitive individuals will show less activity related to error likelihood (**Fig. 4e**). In the course of learning, the reward-sensitive model learns predictions primarily about rewarding outcomes and so shows weaker anticipation of errors. Consequently, more activity occurs when, on error trials, the strong prediction of reward is not counteracted by the actual reward outcome.

DISCUSSION

Overall, the model suggests a unified account of monkey and human mPFC that builds on widely accepted learning models. The simulation results demonstrate that a single term, ω^N, reflecting the surprise related to the non-occurrence of a predicted event, can capture a broad range of cognitive control and performance monitoring effects from various research methodologies. These effects have previously been marshaled as evidence in favor of competing theories, especially of conflict and error monitoring in humans and, conversely, reward prediction and value in monkeys. Thus the PRO model suggests a reconciliation of debates in the literature based on different modalities. The model reinterprets several well known effects: error effects may represent a comparison of actual versus expected outcomes, whereas conflict effects may result from the prediction of multiple possible responses and their outcomes rather than response conflict *per se*. Notably, the model derives these effects from a single mechanism, unexpected non-occurrence, which reflects the rectified negative component of a prediction error signal for both aversive and rewarding events. Furthermore, in the present model, the negative surprise signals consist of rich and context-specific predictions and evaluations[37]. These might drive correspondingly specific proactive and reactive[42] cognitive control adjustments that are appropriate to the specific context. Finally, the PRO model suggests that, within the brain, temporal difference learning signals may be decomposed into their positive and negative components.

The PRO model builds on or relates to several existing model concepts, such as the Bayesian volatility model of ACC simulated above[16]. The negative surprise signal resembles the unexpected uncertainty signal that has been proposed to drive norepinephrine signals[43], although unexpected uncertainty has not been proposed as a signal related to mPFC. The PRO model also resembles models of reinforcement learning in which the value of future states is determined by both the predicted amount of reward and the potential actions available to a learning agent. Indeed, others have simulated ERP data related to mPFC with reinforcement learning models[13,44]. Examples of other related reinforcement learning models include Q learning and SARSA[45,46]. However, these models use a scalar learning signal that combines predicted rewards and possible actions (which may in turn lead to further rewards) into a composite value prediction. In contrast, our model represents individual rather than aggregate outcome probabilities and includes distinct representations of possible aversive as well as rewarding outcomes. The PRO model further diverges from models of reinforcement learning in that it learns a joint probability of responses and their outcomes for a given stimulus context, $P(R,O|S)$, in contrast to reinforcement learning models that aim to learn the probability of an outcome given a response, $P(O|R)$, to select appropriate behaviors. Other reinforcement learning models have been developed with vector rather than scalar learning signals[47]. Although these models are generally concerned with subdividing task control and learning among distinct reinforcement learning controls, the use of a vector-valued learning signal similar to ours has been previously recognized as being necessary for model-based reinforcement learning[48]. However, unlike this previous work, the PRO model suggests that positive and negative components of such a learning signal are maintained independently within the brain. Further comparisons with related models are drawn in the **Supplementary Discussion**.

The mPFC signals representing outcome prediction and negative surprise might have several effects on brain mechanisms and behavior. The PRO model currently simulates surprise signals ω^N and ω^P as modulating the effective learning rate for associating a stimulus with its likely responses and outcomes[16,49]. The prediction and surprise signals may also serve other functions not simulated here. As an impetus for proactive control, mPFC predictions of multiple likely outcomes may provide a basis for evaluating candidate actions and decisions before execution, weighing their anticipated risks[14] against benefits[24], especially in novel situations. Similarly, negative surprise signals may provide an important reactive control signal to other brain regions to drive a change in strategy when the current behavioral strategy is no longer appropriate[8,50].

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/natureneuroscience/.

Note: Supplementary information is available on the Nature Neuroscience website.

ACKNOWLEDGMENTS

We thank L. Pessoa, S. Padmala, T. Braver, V. Stuphorn, A. Krawitz, D. Nee and the J. Schall laboratory for critical feedback. Supported in part by Air Force Office of Scientific Research FA9550-07-1-0454 (J.W.B.), R03 DA023462 (J.W.B.), R01 DA026457 (J.W.B.), a NARSAD Young Investigator Award (J.W.B.) and the Sidney R. Baer Jr. Foundation (J.W.B.). Supported by the Intelligence Advanced Research Projects Activity (IARPA) through Department of the Interior (DOI) contract D10PC20023. The US Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DOI or the US Government.

AUTHOR CONTRIBUTIONS

J.W.B. and W.H.A. conceptualized the model. W.H.A. implemented the model and ran the simulations. J.W.B. and W.H.A. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/natureneuroscience/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Carter, C.S., MacDonald, A.W. III, Ross, L.L. & Stenger, V.A. Anterior cingulate cortex activity and impaired self-monitoring of performance in patients with schizophrenia: an event-related fMRI study. *Am. J. Psychiatry* **158**, 1423–1428 (2001).
2. Gehring, W.J., Coles, M.G.H., Meyer, D.E. & Donchin, E. The error-related negativity: An event-related potential accompanying errors. *Psychophysiology* **27**, S34 (1990).
3. Falkenstein, M., Hohnsbein, J., Hoorman, J. & Blanke, L. Effects of crossmodal divided attention on late ERP components: II. Error processing in choice reaction tasks. *Electroencephalogr. Clin. Neurophysiol.* **78**, 447–455 (1991).
4. Carter, C.S. *et al.* Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science* **280**, 747–749 (1998).
5. Botvinick, M.M., Braver, T.S., Barch, D.M., Carter, C.S. & Cohen, J.C. Conflict monitoring and cognitive control. *Psychol. Rev.* **108**, 624–652 (2001).
6. Olson, C.R. & Gettner, S.N. Neuronal activity related to rule and conflict in macaque supplementary eye field. *Physiol. Behav.* **77**, 663–670 (2002).
7. Ito, S., Stuphorn, V., Brown, J. & Schall, J.D. Performance monitoring by anterior cingulate cortex during saccade countermanding. *Science* **302**, 120–122 (2003).
8. Shima, K. & Tanji, J. Role of cingulate motor area cells in voluntary movement selection based on reward. *Science* **282**, 1335–1338 (1998).
9. Matsumoto, K., Suzuki, W. & Tanaka, K. Neuronal correlates of goal-based motor selection in the prefrontal cortex. *Science* **301**, 229–232 (2003).
10. Matsumoto, M., Matsumoto, K., Abe, H. & Tanaka, K. Medial prefrontal cell activity signaling prediction errors of action values. *Nat. Neurosci.* **10**, 647–656 (2007).
11. Amiez, C., Joseph, J.P. & Procyk, E. Anterior cingulate error-related activity is modulated by predicted reward. *Eur. J. Neurosci.* **21**, 3447–3452 (2005).
12. Scheffers, M.K. & Coles, M.G. Performance monitoring in a confusing world: error-related brain activity, judgments of response accuracy, and types of errors. *J. Exp. Psychol. Hum. Percept. Perform.* **26**, 141–151 (2000).
13. Holroyd, C.B. & Coles, M.G. The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychol. Rev.* **109**, 679–709 (2002).
14. Brown, J.W. & Braver, T.S. Risk prediction and aversion by anterior cingulate cortex. *Cogn. Affect. Behav. Neurosci.* **7**, 266–277 (2007).
15. Brown, J.W. & Braver, T.S. Learned predictions of error likelihood in the anterior cingulate cortex. *Science* **307**, 1118–1121 (2005).
16. Behrens, T.E., Woolrich, M.W., Walton, M.E. & Rushworth, M.F. Learning the value of information in an uncertain world. *Nat. Neurosci.* **10**, 1214–1221 (2007).
17. Walton, M.E., Devlin, J.T. & Rushworth, M.F. Interactions between decision making and performance monitoring within prefrontal cortex. *Nat. Neurosci.* **7**, 1259–1265 (2004).
18. Rudebeck, P.H. *et al.* Frontal cortex subregions play distinct roles in choices between actions and stimuli. *J. Neurosci.* **28**, 13775–13785 (2008).

19. Cole, M.W., Yeung, N., Freiwald, W.A. & Botvinick, M. Cingulate cortex: diverging data from humans and monkeys. *Trends Neurosci.* **32**, 566–574 (2009).
20. Ford, K.A., Gati, J.S., Menon, R.S. & Everling, S. BOLD fMRI activation for anti-saccades in nonhuman primates. *Neuroimage* **45**, 470–476 (2009).
21. Haggard, P. Human volition: towards a neuroscience of will. *Nat. Rev. Neurosci.* **9**, 934–946 (2008).
22. Aarts, E., Roelofs, A. & van Turennout, M. Anticipatory activity in anterior cingulate cortex can be independent of conflict and error likelihood. *J. Neurosci.* **28**, 4671–4678 (2008).
23. Brown, J.W. Conflict effects without conflict in anterior cingulate cortex: multiple response effects and context specific representations. *Neuroimage* **47**, 334–341 (2009).
24. Kennerley, S.W., Dahmubed, A.F., Lara, A.H. & Wallis, J.D. Neurons in the frontal lobe encode the value of multiple decision variables. *J. Cogn. Neurosci.* **21**, 1162–1178 (2009).
25. Croxson, P.L., Walton, M.E., O'Reilly, J.X., Behrens, T.E. & Rushworth, M.F. Effort-based cost-benefit valuation and the human brain. *J. Neurosci.* **29**, 4531–4541 (2009).
26. Schoenbaum, G., Setlow, B., Saddoris, M.P. & Gallagher, M. Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala. *Neuron* **39**, 855–867 (2003).
27. Sallet, J. *et al.* Expectations, gains, and losses in the anterior cingulate cortex. *Cogn. Affect. Behav. Neurosci.* **7**, 327–336 (2007).
28. Amador, N., Schlag-Rey, M. & Schlag, J. Reward-predicting and reward-detecting neuronal activity in the primate supplementary eye field. *J. Neurophysiol.* **84**, 2166–2170 (2000).
29. Nee, D.E., Kastner, S. & Brown, J.W. Functional heterogeneity of conflict, error, task-switching, and unexpectedness effects within medial prefrontal cortex. *Neuroimage* **54**, 528–540 (2011).
30. Procyk, E., Tanaka, Y.L. & Joseph, J.P. Anterior cingulate activity during routine and non-routine sequential behaviors in macaques. *Nat. Neurosci.* **3**, 502–508 (2000).
31. Holroyd, C.B. & Krigolson, O.E. Reward prediction error signals associated with a modified time estimation task. *Psychophysiology* **44**, 913–917 (2007).
32. Miltner, W.H.R., Braun, C.H. & Coles, M.G.H. Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a 'generic' neural system for error-detection. *J. Cogn. Neurosci.* **9**, 788–798 (1997).
33. Yeung, N. & Nieuwenhuis, S. Dissociating response conflict and error likelihood in anterior cingulate cortex. *J. Neurosci.* **29**, 14506–14510 (2009).
34. Burle, B., Roger, C., Allain, S., Vidal, F. & Hasbroucq, T. Error negativity does not reflect conflict: a reappraisal of conflict monitoring and anterior cingulate cortex activity. *J. Cogn. Neurosci.* **20**, 1637–1655 (2008).
35. Amiez, C., Joseph, J.P. & Procyk, E. Reward encoding in the monkey anterior cingulate cortex. *Cereb. Cortex* **16**, 1040–1055 (2006).
36. Quilodran, R., Rothe, M. & Procyk, E. Behavioral shifts and action valuation in the anterior cingulate cortex. *Neuron* **57**, 314–325 (2008).
37. Brown, J.W. Multiple cognitive control effects of error likelihood and conflict. *Psychol. Res.* **73**, 744–750 (2009).
38. Nakamura, K., Roesch, M.R. & Olson, C.R. Neuronal activity in macaque SEF and ACC during performance of tasks involving conflict. *J. Neurophysiol.* **93**, 884–908 (2005).
39. Oliveira, F.T., McDonald, J.J. & Goodman, D. Performance monitoring in the anterior cingulate is not all error related: expectancy deviation and the representation of action-outcome associations. *J. Cogn. Neurosci.* **19**, 1994–2004 (2007).
40. Jessup, R.K., Busemeyer, J.R. & Brown, J.W. Error effects in anterior cingulate cortex reverse when error likelihood is high. *J. Neurosci.* **30**, 3467–3472 (2010).
41. Shidara, M. & Richmond, B.J. Anterior cingulate: single neuronal signals related to degree of reward expectancy. *Science* **296**, 1709–1711 (2002).
42. Braver, T.S., Gray, J.R. & Burgess, G.C. Explaining the many varieties of working memory variation: dual mechanisms of cognitive control. in *Variation of Working Memory* (eds. Conway, C.J.A., Kane, M., Miyake, A. & Towse, J.) 76–106 (Oxford University Press, 2007).
43. Yu, A.J. & Dayan, P. Uncertainty, neuromodulation, and attention. *Neuron* **46**, 681–692 (2005).
44. Holroyd, C.B., Yeung, N., Coles, M.G. & Cohen, J.D. A mechanism for error detection in speeded response time tasks. *J. Exp. Psychol. Gen.* **134**, 163–191 (2005).
45. Singh, S.P. & Sutton, R.S. Reinforcement learning with replacing eligibility traces. *Mach. Learn.* **22**, 123–158 (1996).
46. Watkins, C.J.C.H. & Dayan, P. Q-learning. *Mach. Learn.* **8**, 279–292 (1992).
47. Doya, K., Samejima, K., Katagiri, K.-i. & Kawato, M. Multiple Model-Based Reinforcement Learning. *Neural Comput.* **14**, 1347–1369 (2002).
48. Gläscher, J., Daw, N., Dayan, P. & O'Doherty, J.P. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* **66**, 585–595 (2010).
49. Pearce, J.M. & Hall, G. A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol. Rev.* **87**, 532–552 (1980).
50. Bush, G. *et al.* Dorsal anterior cingulate cortex: a role in reward-based decision making. *Proc. Natl. Acad. Sci. USA* **99**, 523–528 (2002).

ONLINE METHODS

Computational model. The PRO model consists of three main components (see **Supplementary Fig. 1**). The model constitutes a bridge between cognitive control and reinforcement learning theories in that the structure of the model resembles an actor-critic model, with a module responsible for generating actions (the 'actor') architecturally segregated from a module that generates predictions and signals prediction errors (the 'critic'). An additional module learns a prediction of the frequency with which composite events are observed to occur within a task context ('outcome representation'). Unlike typical actor-critic architectures, the critic component is not involved directly in training the actor; rather, the critic indirectly influences the actor's policy by modulating the rate at which predictions of response–outcome conjunctions, which serve as direct input into the actor component, are learned.

Representing events. The outcome representation component of the PRO model (**Supplementary Fig. 1**) learns to associate observed conjunctions of responses and outcomes with the task-related stimuli that predict them. The number of total conjunctions that are available for learning may vary from task to task depending on the particular responses required and potential outcomes. In the change signal task described below, for example, subjects may either make a 'go' or 'change' response, resulting in 'correct' or 'error' outcomes, for a total of four possible response–outcome conjunctions.

The PRO model (**Supplementary Fig. 1**) learns a prediction of response–outcome conjunctions ($S_{i,t}$) that may occur, specifically in the current task, as a function of incoming task stimuli ($D_{j,t}$)

$$S_{i,t} = \sum_j D_{j,t} W^S_{ij,t} \tag{3}$$

where D is a vector representing current task stimuli and W^S is a matrix of weights that maintain a prediction of response–outcome conjunctions. S can be thought of as proportional to a conditional probability of a particular response–outcome conjunction given the current trial conditions D. The role of S is to provide an immediate prediction of the likely outcomes of actions and inhibit those that are predicted to yield an undesirable outcome (see equation (13)). Prediction weights are updated according to

$$W^S_{ij,t+1} = W^S_{ij,t} + A_{i,t}(O_{i,t} - S_{i,t})G_t D_j \tag{4}$$

where O is a vector of actual response–outcome conjunctions occurring at time t, G is a neuromodulatory gating signal equal to 1 if a behaviorally relevant event is observed and 0 otherwise, and A is a learning rate variable calculated as

$$A_{i,t} = \frac{\alpha}{1 + \left(\omega^{\mathrm{P}}_{i,t} + \omega^{\mathrm{N}}_{i,t}\right)} \tag{5}$$

where α is a baseline learning rate and $\omega^{\mathrm{P}}_{i,t}$ and $\omega^{\mathrm{N}}_{i,t}$ are measures of positive and negative surprise, respectively (see below).

Temporal difference model of outcome prediction. In addition to the immediate outcome prediction signals S above that can quickly control behavior, the critic unit (**Supplementary Fig. 1**) also learns a complementary timed prediction of the time at which an outcome is expected to occur. Unlike S, this timed prediction signal V is not immediately active but peaks at the time of the expected outcome. This in turn provides a critical basis for detecting when expected outcomes fail to occur, so that the outcome predictions S that control behavior can be updated. In general, the temporal difference error may be written as follows

$$\delta_t = r_t + \gamma V_{t+1} - V_t \tag{6}$$

$$\delta_{i,t} = r_{i,t} + \gamma V_{i,t+1} - V_{i,t} \tag{7}$$

Here $r_{i,t}$ is a function of observed response–outcome conjunctions $O_{i,t}$. For most simulations, $r_{i,t}$ was equal to $O_{i,t}$, except for simulation 9, in which $r_{i,t}$ was equal to $O_{i,t} \times F_i$, where F is a constant reflecting the salience of response–outcome conjunction i. In essence, equation (7) specifies a vector-valued temporal difference model that learns a prediction proportional to the likelihood of a given response–outcome conjunction at a given time. Except where noted, $\gamma = 0.95$ in all simulations.

As in previous formulations of temporal difference learning, the representation of task-related stimuli over time is modeled as a tapped delay chain, X, composed of multiple units, indexed by j, whose activity (value set to 1) tracks the number of model iterations ('time') elapsed since the presentation of a task-related stimulus. Each iteration (dt) represents 10 ms of real time. Value predictions are computed as

$$V_{i,t} = \sum_{j,k} X_{jk,t} \times U_{ijk,t} \tag{8}$$

where j is the delay unit corresponding to the current time elapsed since the onset of a stimulus k and U is the learned prediction weight. Weights are updated according to

$$U_{ijk,t+1} = U_{ijk,t} + \alpha\delta_{i,t}\bar{X}_{jk} \tag{9}$$

where α is a learning rate parameter and constrained by $U_{ijk} > 0$. $\bar{X}$ is an eligibility trace computed as

$$\bar{X}_{jk,t+1} = X_{jk,t} + 0.95\bar{X}_{jk,t} \tag{10}$$

Stimulus-response architecture. In the actor unit (**Supplementary Fig. 1**), activity in response units C is modeled as

$$C_{i,t+1} = C_{i,t} + \beta \mathrm{d}t(E_{i,t}(1 - C_{i,t}) - (C_{i,t} + 0.05)(I_{i,t} + 1)) + N(0,\sigma) \tag{11}$$

where dt is a time constant, β is a multiplicative factor and N is Gaussian noise with mean 0 and variance σ. E is the net excitatory input to the response units and I is the net inhibitory input to response units. Excitatory input to the response units is determined by

$$E_{i,t} = \rho\sum_j D_j W^C_{ij} \tag{12}$$

where D are task-related stimuli, W^C are prespecified weights describing hard-wired responses indicated by task stimuli and ρ is a scaling factor. Note that weights W^C implement stimulus-response mappings that are the usual target of (model-free) reinforcement learning in other models. Here learning in the PRO model instead updates outcome predictions S, which provide model-based control of actions C. The model is considered to have generated a behavioral responses when the activity of any response unit exceeds a response threshold Γ. Subsequent response unit activity in a trial that exceeds the threshold is ignored (that is, is not considered to be a behavioral response), whether it is a different response unit or the same response unit whose activity has returned below threshold owing to processing noise.

Cognitive control signal architecture. *Proactive control.* The simulation of the change signal task requires a cognitive control signal based on outcome predictions S, which inhibits the model units that generate responses. The vector-valued control signal derived from predicted outcomes could be extended to provide a variety of different control signals in different conditions. In the present model, inhibition to the response units is determined by

$$I_{i,t} = \psi\left(\sum_j C_j W^{\mathrm{I}}_{ij}\right) + \phi\left(\sum_k S_k W^{\mathrm{F}}_{ik}\right) \tag{13}$$

where W^{I} are fixed weights describing mutual inhibition between response units, W^{F} are adjustable weights describing learned, top-down control from predicted response-outcome representations, and ψ and ϕ are scaling factors. O is the vector of experienced response-outcome representations (equations (3) and (4)). Adjustable weights W^{F} are learned by

$$W^{\mathrm{F}}_{ik,t+1} = W^{\mathrm{F}}_{ik,t} + 0.01 C_{i,t} T_{i,t} O_{k,t} G_t Y_t \tag{14}$$

where Y_t is an affective evaluation of the observed outcome. For errors, $Y_t = 1$; for correct responses, $Y_t = -0.1$. The variable $T_{i,t}$ implements a thresholding function such that $T_{i,t} = 1$ if $C_{i,t} > \Gamma$ and 0 otherwise.

Reactive control. Reactive control signals in the model are generated whenever an actual outcome differs from an expected outcome. Their magnitude is greatest when an outcome is most unexpected. Signals from the PRO model corresponding to the two forms of surprise described in the main text are calculated as follows. For the first type, unexpected occurrences, the signal is calculated as

$$\omega^{\mathrm{P}}_{i,t} = \sum_i [O_{i,t} - V_{i,t}]^+ \tag{15}$$

doi:10.1038/nn.2921

and the second type of surprise, unexpected non-occurrence, is calculated as

$$\omega^{N}_{i,t} = \sum_{i} [V_{i,t} - O_{i,t}]^{+} \quad (16)$$

As noted above, ω^P and ω^N are used to modulate the effective learning rate for predictions of response–outcome conjunctions. The formulation of equation (5) modulates the learning rate of the model in proportion to uncertainty. In stable environments, infrequent surprises result in large values for ω^P and ω^N, which in turn reduce the effective learning rate, whereas in situations in which the model has only weak predictions of likely outcomes, ω^P and ω^N are relatively weak, resulting in increased learning rates. The rationale underlying this arrangement is that infrequent events, which are associated with increased ACC activity, are likely to represent noise rather than a behaviorally significant shift in environmental contingencies, and therefore an individual should be slow to adjust behavior.

Model fitting. Model parameters (**Table 1**) were adjusted by gradient descent to optimize the least-squares fit between human behavioral and model reaction time and error rate data. The model was fit using a change signal task using previously reported behavioral data[15]. There are seven free parameters in the model in **Table 1** and ten data points from the change signal task (eight for reaction time and two for error rate). These parameters allowed the model to simulate the reaction time and error rate effects in the change signal data. The parameters were then fixed for the remaining simulations unless explicitly stated otherwise. Because the model was only fit to human behavioral data, the key model predictions of fMRI, ERP and single-unit neurophysiology effects result from the qualitative properties of the model rather than from *post hoc* data fits.

The best-fit parameters yielded model behavior that corresponded well with human results. The model was trained on 400 trials of the change signal task. Error rates for the model were 49.97% and 5.64% for the high and low error-likelihood conditions, respectively, in line with human data. Effects of previous trial type on current trial reaction time were in agreement with human performance. For go trials in which the previous and current trial were correct, the eight conditions yielded a correlation of $r = 0.96$ ($t_{1,6} = 27.17$, $P = 0.00021$) between human and model responses times, indicating that the model captured relevant behavioral effects observed in human data.

Simulation details. In each simulation, trials were presented at intervals of 3 s of simulated time. Trials were initiated with the onset of a stimulus presented to the input vector D. All results presented in the main text were averaged over ten separate runs for each simulated task and reflect the derived measure of negative surprise ω^N, except for **Figure 3b**, which reflects positive surprise (ω^P). For results presented in bar graph form or results in which data were otherwise concatenated (simulations 1, 3, 5–8), the value of ω^N for the first 120 iterations (1.2 s) of a trial were averaged together when trials were aligned on stimulus onset. When data were aligned on feedback, the value of ω^N was taken from the 20 iterations preceding feedback and 80 iterations following feedback.

Simulations 1, 2 and 4: change signal task. In the change signal task, participants must press a button corresponding to an arrow pointing left or right. On one-third of the trials, a second arrow is presented above the first, indicating that the subject must withhold the response to the first arrow and instead make the opposite response. The color of the arrows is an implicit cue that predicts the likelihood of error as follows: for conditions with high error likelihood, the onset delay of the second arrow is dynamically adjusted to enforce a high rate of error commission (50%). On trials with low error likelihood, the onset of the second arrow is shortened to allow a lower error rate of 5%. The error effect is the difference in ω^N between change,error versus change,correct trials; the conflict effect is the contrast between change,correct versus go,correct trials, and the error likelihood effect is the contrast of correct,go trials between high and low error likelihood color cues.

The model was trained for 400 trials, presented randomly. Four task stimuli were used, indicating trial condition: high error likelihood,go; high error likelihood,change; low error likelihood,go; low error likelihood,change. On go trials in either error likelihood condition, the stimulus unit (D) corresponding to the go cue in that condition became active (D(go) = 1) at 0 s and remained active for a total of 1,000 ms. On change trials, a second input unit became active at either 130 ms (low error likelihood) or 330 ms (high error likelihood). On change trials, units representing both go and change cues were active simultaneously when the change signal was presented.

Simulation 3: speed-accuracy tradeoff. The model architecture and parameters were the same as in simulation 1 except that connection weights from stimulus units corresponding to the central cue in an Eriksen flanker task were set to 1, and weights corresponding to distractor cues were set to 0.4, the noise parameter was set to 0.02 and the temporal discount factor was set to 0.85. The model was trained for 1,000 trials on the flanker task. In this task, subjects are asked to make a response as cued by a central target stimulus. On 'congruent' trials in the task, additional stimuli that cue the same response as the target are presented to either side of the target stimulus. On 'incongruent' trials, the additional stimuli cue an alternative response. Incongruent and congruent trials were presented to the model pseudorandomly, with approximately half of all trials being congruent.

Simulation 5: multiple response effect. The model architecture remained the same as in simulation 1 except that lateral inhibition between response units (equation (13)) was removed to allow simultaneous generation of response. Two input representations were used to represent task stimuli, a 'single response' cue and a 'both response' cue. Hard-wired connections from stimulus representations to response units ensured that the single response cue could only result in generation of the appropriate solitary response, while the both response cue activated both response units at approximately the same rate. The model was trained for 400 trials, with approximately half of the trials being single-response trials.

Simulation 6: volatility. The model was trained on a two-armed bandit task[16] in which two responses, each representing a different gamble with different payoff frequencies, were possible. The model was trained in a series of nine stages, divided into four epochs (**Fig. 4b**). In the first stage of 120 trials, the payoff frequencies of the two gambles were fixed such that one gamble paid off on 80% of the trials in which it was chosen, and the alternative gamble paid off on 20% of the trials in which it was chosen. Starting on trial 121, these payoff contingencies were switched, so that the first gamble paid off at a rate of 20% and the alternate gamble paid off at a rate of 80%. These contingencies were switched every 40 trials a total of seven times. Finally, the payoff contingencies were returned to their initial values for the final 180 trials. Top-down control weights, W^C, were fixed such that weights associated with errors were 0.15 and weights associated with correct outcomes were −0.05. This was done so that estimates of learning rates were influenced by updates of response-outcome representations alone and were not influenced by learning related to control. The PRO model's choices and experienced outcomes were recorded and used as input to a Bayesian learner[16] to derive measures of volatility in each phase, and to a simple reinforcement learning model in order to estimate model learning rates during each phase (see **Supplementary Note**).

Choice behavior from the PRO model, as well as a version of the PRO model in which surprise signals were suppressed ('lesioned'), was used as input to a reinforcement learning model (see **Supplementary Note**) to derive effective learning rates. When surprise signals generated by the PRO model were used to modulate learning rates, the model adapted more quickly to changing environmental contingencies than during more stable periods. In contrast, the lesioned model maintained the same learning rate regardless of environmental instability.

Simulation 7: unexpected outcomes. The model architecture, task and parameters were the same as described in simulation 3, except that weights from stimulus input units to response units were set to 0.5 and 2 for the responses associated with, respectively, the central target cue and distractor cues in the Eriksen flanker task. This manipulation is analogous to increasing the saliency of distractor cues to promote increased error rate. The model was simulated for 1,000 trials a total of 10 times, and error rates for incongruent trials averaged about 70%.

Simulation 8: unexpected timing. The PRO model simulation predicts that mPFC signals not only unexpected outcomes but also expected outcomes that occur at an unexpected time. The model architecture was the same as for simulation 5. However, instead of manipulating the number of responses, feedback to the model (always correct) was given either after a short delay (200 ms) on 80% of the trials, whereas for the remaining 20% of the trials, feedback was given 600 ms after a response was generated. The model was trained on this task for 1,000 trials. **Figure 4d** shows ω^N averaged over trials for long and short delay intervals.

Simulation 9: individual differences. The model, task and parameters were the same as described for simulation 1, except that the effective salience to events was parametrically manipulated to explore the effect of sensitivity to rewarding and aversive events in the model. The salience factor F (see above) was varied from 0.2857 to 1.7143 for rewarding events, and the factor for aversive events was varied from 1.7143 to 0.2857, resulting in 11 conditions for which the ratio of reward to risk sensitivity ranged from 1/6 (risk sensitive) to 6 (reward sensitive). For each condition, ten simulated runs were included in calculating the mean for each data point.

doi:10.1038/nn.2921