

RESEARCH ARTICLE

Is Model Fitting Necessary for Model-Based fMRI?

Robert C. Wilson^{1*}, Yael Niv²

1 Department of Psychology and Cognitive Science Program, University of Arizona, Tucson Arizona, United States of America, **2** Princeton Neuroscience Institute, Princeton University, Princeton, New Jersey, United States of America

* bob@email.arizona.edu



OPEN ACCESS

Citation: Wilson RC, Niv Y (2015) Is Model Fitting Necessary for Model-Based fMRI?. *PLoS Comput Biol* 11(6): e1004237. doi:10.1371/journal.pcbi.1004237

Editor: Erie Dell Boorman, University of Oxford, UNITED KINGDOM

Received: June 20, 2014

Accepted: March 12, 2015

Published: June 18, 2015

Copyright: © 2015 Wilson, Niv. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Relevant data from Niv et al. (2012) are provided as part of the Supporting Information. Because the data from Daw et al. (2006) was acquired by a third party it is not part of the Supporting Information but can be obtained by contacting Nathaniel Daw (ndd204@nyu.edu) and John O'Doherty (joherty@caltech.edu).

Funding: This work was supported by National Institute of Health grant R01MH098861 (<http://www.nih.gov/>) to YN and National Institute of Mental Health and National Institute of Drug Abuse, training grant 5T32MH065214 (<http://www.drugabuse.gov/>) to the Princeton Neuroscience Institute. The funders had no

Abstract

Model-based analysis of fMRI data is an important tool for investigating the computational role of different brain regions. With this method, theoretical models of behavior can be leveraged to find the brain structures underlying variables from specific algorithms, such as prediction errors in reinforcement learning. One potential weakness with this approach is that models often have free parameters and thus the results of the analysis may depend on how these free parameters are set. In this work we asked whether this hypothetical weakness is a problem in practice. We first developed general closed-form expressions for the relationship between results of fMRI analyses using different regressors, e.g., one corresponding to the true process underlying the measured data and one a model-derived approximation of the true generative regressor. Then, as a specific test case, we examined the sensitivity of model-based fMRI to the learning rate parameter in reinforcement learning, both in theory and in two previously-published datasets. We found that even gross errors in the learning rate lead to only minute changes in the neural results. Our findings thus suggest that precise model fitting is not always necessary for model-based fMRI. They also highlight the difficulty in using fMRI data for arbitrating between different models or model parameters. While these specific results pertain only to the effect of learning rate in simple reinforcement learning models, we provide a template for testing for effects of different parameters in other models.

Author Summary

In recent years, model-based fMRI has emerged as a powerful technique in psychology and neuroscience. With this method, computational models of behavior can be leveraged to identify where, whether and how different algorithms are implemented in the brain. Yet this approach seems to have an Achilles heel in that the models frequently have free parameters, and errors in setting these parameters could lead to errors in interpretation of the data. Here we asked whether this potential weakness, in theory, is an actual weakness in practice. In particular, we tested whether errors in estimating participants' learning rate in a trial-and-error reinforcement learning setting would have adverse effects on

role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

identifying the neural substrates of the learning process. Amazingly, it turns out that even gross errors in the learning rate lead to only minute changes in the neural results. The good news is that precise identification of free parameters is not always necessary; the corollary bad news is that it may be harder to identify the precise computational roles of different brain areas than we had previously appreciated. Based on our analytical results, we offer suggestions for designing experiments that maximize or minimize sensitivity to model parameters, as needed.

Introduction

The advent of fMRI revolutionized psychology as it allowed, for the first time, the noninvasive mapping of human cognition. Despite this progress, traditional fMRI analyses are limited in that they can, for the most part, only ascertain the involvement of an area in a task but not its precise *role* in that task. Recently, model-based fMRI methods have been developed to overcome this limitation by using computational models of behavior to shed light on latent variables of the models (such as prediction errors) and their mapping to neural structures. This approach has led to important insights into the algorithms employed by the brain and has been particularly successful in understanding the neural basis of reinforcement learning (e.g. [1–13]).

In a typical model-based fMRI analysis, one first specifies a model that describes the hypothesized cognitive processes underlying the behavior in question. Typically these models have one or more free parameters (e.g. learning rate in a model of trial-and-error learning). These parameters must be set to fully specify the model, which is commonly done by fitting them to the observed behavior [14]. For instance, given the model, one can find subject-specific learning rates that best explain the subject's behavioral choices. The fully specified model is then used to generate trial-by-trial measures of latent variables in the model (e.g. action values and prediction errors) that can be regressed against neural data in order to find areas whose activity correlates with these variables in the brain.

One potential weakness of this approach is the requirement for model fitting. In many cases, the data are insufficient to precisely identify the parameter values. This can be due to limited number of trials, interactions between parameters that make them hard to disentangle [14] or lack of behavior that can be used for the fitting process (e.g., in some Pavlovian conditioning experiments). Thus a key question is: How important is the model fitting step? In other words, to what extent is model-based fMRI sensitive to errors in parameter estimation? The answer to this question will determine how hard we should work to obtain the best possible parameter fits, and will affect not only how we analyze data, but also how we design experiments in the first place.

Here we show how this question can be addressed, by analyzing the sensitivity of model-based fMRI to the learning rate parameter in simple reinforcement learning tasks. We provide analytical bounds on the sensitivity of the model-based analysis to errors in estimating the learning rate, and show through simulation how value and prediction error signals generated with one learning rate would be interpreted by a model-based analysis that used the wrong learning rate. Amazingly, we find that the results of model-based fMRI are remarkably robust to settings of the learning rate to the extent that, in some situations, setting the parameters of the model as far as possible from their true value barely affects the results. This theoretical prediction of robustness is borne out by analysis of fMRI data from two recent experiments.

Our findings are both good and bad news for model-based fMRI. The good news is that it is robust, thus errors in the learning rate will not dramatically change the results of studies

seeking to localize a particular signal. The bad news, however, is that model-based fMRI is insensitive to differences in parameters, which means that one should use extreme caution when attempting to determine the computational role of a neural area (e.g., when asking whether a brain area corresponds to an outcome signal or a prediction error signal). In the Discussion we consider the extent to which this result generalizes to other parameters and other models and offer suggestions to diagnose parameter sensitivity in other models.

Methods

Ethics statement

Both experiments were approved by their respective institutions. The experiment in [10] was approved by the Institutional Review Board of the California Institute of Technology. The experiment in [3] was approved by Ethics Committee at University College London. In both cases participants gave informed consent in writing.

Theoretical analysis

We begin by laying out a formal analysis of the sensitivity of model-based fMRI to model parameters. The rationale behind the mathematical derivations below is as follows. Assume that there is some signal in the brain (corresponding to some ‘ground truth’ regressor \mathbf{x}_g) that we have a noisy measurement of (e.g., via fMRI). We first derive the somewhat intuitive result that if we analyze the brain data with a different, incorrect regressor \mathbf{x}_f (where the subscript, f , denotes that the regressor is derived from our model with *fit* parameter values), the quality of our results depends on the correlation between the ground truth regressor and the incorrect regressor, $\rho(\mathbf{x}_g, \mathbf{x}_f)$.

To assess the sensitivity of model-based fMRI to errors in parameter estimation, we then focus on trial-and-error learning tasks. We assume a ground truth regressor derived from a reinforcement learning model with the learning rate parameter set to its true (though unknown) value, and analyze the correlation between this regressor and one that is derived from the same model but with a different setting of the learning rate, for some of the most commonly used task designs. Finally, we illustrate and flesh out the implications of these analytical results using both simulated and empirical data in the Results.

The effect of an incorrect regressor on fMRI analysis. Assume a ‘ground truth’ regressor $\mathbf{x}_g = (x_{g1}, x_{g2}, \dots, x_{gT})$ (where x_{gt} is the size of the variable of interest at time point t) that underlies the activity in a brain region, such that the measured signal in this region takes the form

$$\mathbf{Y} = \beta \mathbf{x}_g + \epsilon, \tag{1}$$

with β being a coefficient that controls the size of the effect and ϵ being zero-mean noise. What would be the magnitude of the estimated regression coefficient $\hat{\beta}$ if we analyzed the brain data using an incorrect regressor, \mathbf{x}_f (for example one that is derived from an incorrect model, or from the correct model with the wrong setting of the free parameters)? Using ordinary least squares regression, we have

$$\begin{aligned} \hat{\beta} &= (\mathbf{x}_f^T \mathbf{x}_f)^{-1} \mathbf{x}_f^T \mathbf{Y} \\ &= (\mathbf{x}_f^T \mathbf{x}_f)^{-1} (\mathbf{x}_f^T \beta \mathbf{x}_g) \\ &= \beta \rho(\mathbf{x}_g, \mathbf{x}_f) \frac{\sigma(\mathbf{x}_g)}{\sigma(\mathbf{x}_f)} \end{aligned} \tag{2}$$

where $\sigma(\mathbf{x}) = \sqrt{\frac{\mathbf{x}^T \mathbf{x}}{T}}$ is the standard deviation of regressor \mathbf{x} , $\rho(\mathbf{x}_g, \mathbf{x}_f) = \frac{\mathbf{x}_g^T \mathbf{x}_f}{T\sigma(\mathbf{x}_g)\sigma(\mathbf{x}_f)}$ is the correlation coefficient between \mathbf{x}_g and \mathbf{x}_f , and T is the number of data points in the regression. Thus, if we normalize the regressors to have unit variance, i.e. $\sigma(\mathbf{x}_g) = \sigma(\mathbf{x}_f) = 1$, then $\hat{\beta}$ is related to the ground truth regression coefficient through the correlation between the two regressors:

$$\hat{\beta} = \beta \rho(\mathbf{x}_g, \mathbf{x}_f) \tag{3}$$

Thus, the more correlated the fit regressor is to the true regressor, the larger the regression coefficient for the fit regressor, $\hat{\beta}$. How does this affect the statistical significance of $\hat{\beta}$, that is, the results of a statistical analysis that asks whether $\hat{\beta}$ is reliably different from zero? To answer this question, we must compute \hat{t} , the Student t statistic of $\hat{\beta}$ relative to the null hypothesis $\hat{\beta} = 0$. Making the further simplifying assumption that the fMRI noise, ϵ , is Gaussian with variance σ_{fMRI}^2 , we have

$$\hat{t} = \frac{\hat{\beta}}{s(\hat{\beta})} \tag{4}$$

where $s(\hat{\beta})$ is standard error of $\hat{\beta}$. For simple regression, $s(\hat{\beta})$ can be written in terms of the standard deviation of the regression residuals, $\hat{\epsilon}$, as

$$s(\hat{\beta}) = \frac{1}{\sqrt{T-2}} \frac{\sigma(\hat{\epsilon})}{\sigma(\mathbf{x}_f)} \tag{5}$$

To compute the standard deviation of the residuals we first note that, by definition,

$$\begin{aligned} \hat{\epsilon} &= \mathbf{Y} - \hat{\beta} \mathbf{x}_f \\ &= \beta \mathbf{x}_g + \epsilon - \beta \rho(\mathbf{x}_g, \mathbf{x}_f) \mathbf{x}_f. \end{aligned} \tag{6}$$

Thus, because the residuals have zero mean

$$\begin{aligned} \sigma(\hat{\epsilon})^2 &= \frac{\hat{\epsilon}^T \hat{\epsilon}}{T} \\ &= \frac{1}{T} (\beta^2 \mathbf{x}_g^T \mathbf{x}_g + \epsilon^T \epsilon + \beta^2 \rho(\mathbf{x}_g, \mathbf{x}_f)^2 \mathbf{x}_f^T \mathbf{x}_f - 2\beta^2 \rho(\mathbf{x}_g, \mathbf{x}_f) \mathbf{x}_f^T \mathbf{x}_g + \cancel{2\beta \mathbf{x}_g^T \epsilon} - \cancel{2\beta \rho(\mathbf{x}_g, \mathbf{x}_f) \mathbf{x}_f^T \epsilon}) \\ &= \sigma_{fMRI}^2 + \beta^2 (1 - \rho(\mathbf{x}_g, \mathbf{x}_f)^2) \end{aligned} \tag{7}$$

where we have used the fact that $\mathbf{x}_g^T \epsilon \approx 0$ and $\mathbf{x}_f^T \epsilon \approx 0$ to cancel out the terms in the second line, and the definition of the correlation coefficient to make the simplification in the third line. Combining this expression with Eqs 4 and 5 and keeping in mind that $\sigma(\mathbf{x}) = 1$ allows us to write down the t statistic as

$$\begin{aligned} \hat{t} &= \beta \rho(\mathbf{x}_g, \mathbf{x}_f) \sqrt{\frac{T-2}{\sigma_{fMRI}^2 + \beta^2 (1 - \rho(\mathbf{x}_g, \mathbf{x}_f)^2)}} \\ &= \rho(\mathbf{x}_g, \mathbf{x}_f) \text{CNR} \sqrt{\frac{T-2}{1 + \text{CNR}^2 (1 - \rho(\mathbf{x}_g, \mathbf{x}_f)^2)}} \end{aligned} \tag{8}$$

where $\text{CNR} = \beta/\sigma_{fMRI}$ denotes the ‘contrast-to-noise’ ratio [15]—the ratio between the strength of the fMRI signal, β , and the standard deviation of the fMRI noise, σ_{fMRI} . Note that this t statistic, like the regression coefficient, $\hat{\beta}$, is a function of the correlation between the ground truth

and incorrect regressors, $\rho(\mathbf{x}_g, \mathbf{x}_f)$, as well as the contrast-to-noise ratio, CNR, and the number of data points in the regression, T .

Correlations between regressors for different parameter settings in a Pavlovian task.

We now turn to analyzing the effects of incorrectly specifying free parameters on model-based fMRI. For this, we concentrate on the particular case of reinforcement learning models. These models, which have formed the bulk of model-based fMRI, attempt to predict how values of different options, as computed by the subject, change through experience. One important free parameter in such learning scenarios is the *learning rate* by which values are updated as a result of prediction errors [16]. We assume that there is some ground truth setting of this parameter that corresponds to the true learning rate of the subject, however, this setting is unknown and our fMRI analysis might utilize an incorrect learning rate. Thus to assess the quality of the results we may hope to obtain from the fMRI analysis, we analyze the properties of $\rho(\mathbf{x}_g, \mathbf{x}_f)$ assuming two regressors that are generated from the same model using different learning rates.

For simplicity, we concentrate on Pavlovian tasks in which subjects experience the rewards associated with different options passively, and use the Rescorla-Wagner learning rule [17]. On each trial, t , subjects are presented with a reward r_t and learn to estimate a value V_{t+1} that represents the reward predicted on the next trial. These values are updated on every trial according to

$$V_{t+1} = V_t + \alpha \delta_t \tag{9}$$

where α is the learning rate and

$$\delta_t = r_t - V_t \tag{10}$$

is the prediction error quantifying the difference between the actual and predicted reward on the current trial. We denote the ground truth learning rate as α_g , and the fit learning rate as α_f . Accordingly, we denote the vector of values learned with learning rate α_i as $\mathbf{V}_i = (V_{i1}, V_{i2}, \dots, V_{iT})$ and the corresponding vector of prediction errors as δ_i .

Model-based fMRI studies of reinforcement learning typically use values and prediction errors as regressors for neural activity [18]. Thus, in order to determine the sensitivity of the results to the accuracy of the parameter fits, we need to compute the correlation coefficients $\rho(\mathbf{V}_g, \mathbf{V}_f)$ and $\rho(\delta_g, \delta_f)$. To do this, we first note that in the general case, the values and prediction errors computed according to Eqs 9 and 10 will not have zero mean or unit variance. This does not invalidate the previously derived results but does require us to work with the more general form of the correlation coefficient

$$\rho(\mathbf{a}, \mathbf{b}) = \frac{\text{cov}(\mathbf{a}, \mathbf{b})}{\sigma(\mathbf{a})\sigma(\mathbf{b})} \tag{11}$$

where $\text{cov}(\mathbf{a}, \mathbf{b})$ is the covariance between vectors \mathbf{a} and \mathbf{b} defined by

$$\text{cov}(\mathbf{a}, \mathbf{b}) = \frac{\mathbf{a}^T \mathbf{b}}{T} - \mu(\mathbf{a})\mu(\mathbf{b}). \tag{12}$$

Here $\mu(\mathbf{a})$ is the mean of vector \mathbf{a} , and $\sigma(\mathbf{a})$, the standard deviation of an uncentered regressor, is given by

$$\sigma(\mathbf{a}) = \sqrt{\frac{\mathbf{a}^T \mathbf{a}}{T} - \mu(\mathbf{a})^2}. \tag{13}$$

Thus to compute the correlation coefficients $\rho(\mathbf{V}_g, \mathbf{V}_f)$ and $\rho(\delta_g, \delta_f)$ we must compute the mean, variance and covariance of the value and prediction error regressors. These statistics

turn out to be completely determined by the learning rates and the statistics of the reward vector, \mathbf{r} .

In particular, for the value regressors it can be shown that

$$\begin{aligned} \mu(\mathbf{V}_i) &\approx \mu(\mathbf{r}) \\ \sigma(\mathbf{V}_i)^2 &\approx \frac{\alpha_i}{2 - \alpha_i} \left(\mu(\mathbf{r}^2) + \frac{2}{T} \sum_{\Delta=1}^{T-1} (T - \Delta)(1 - \alpha_i)^\Delta R_\Delta(\mathbf{r}) \right) - \mu(\mathbf{r})^2 \\ \text{cov}(\mathbf{V}_i, \mathbf{V}_j) &\approx \frac{\alpha_i \alpha_j}{\alpha_i + \alpha_j - \alpha_i \alpha_j} \left(\mu(\mathbf{r}^2) + \frac{1}{T} \sum_{\Delta=1}^{T-1} \left((1 - \alpha_i)^\Delta + (1 - \alpha_j)^\Delta \right) (T - \Delta) R_\Delta(\mathbf{r}) \right) - \mu(\mathbf{r})^2 \end{aligned} \tag{14}$$

where the approximations hold in the limit of many trials and $R_\Delta(\mathbf{r})$ is the (uncentered) auto-correlation of \mathbf{r} at delay Δ defined as

$$R_\Delta(\mathbf{r}) = \frac{1}{T - \Delta} \sum_{a=1}^{T-\Delta} \mathbf{r}_a \mathbf{r}_{a+\Delta}. \tag{15}$$

Equivalently, for the prediction errors we have

$$\begin{aligned} \mu(\delta_i) &\approx 0 \\ \sigma(\delta_i)^2 &\approx \frac{2}{2 - \alpha_i} \left(\mu(\mathbf{r}^2) - \sum_{\Delta=1}^{T-1} \alpha_i (1 - \alpha_i)^{\Delta-1} \left(1 - \frac{\Delta}{T} \right) R_\Delta(\mathbf{r}) \right) \\ \text{cov}(\delta_i, \delta_j) &\approx \frac{1}{\alpha_i + \alpha_j - \alpha_i \alpha_j} \left((\alpha_i + \alpha_j) \mu(\mathbf{r}^2) - \sum_{\Delta=1}^{T-1} \left(\alpha_i^2 (1 - \alpha_i)^{\Delta-1} + \alpha_j^2 (1 - \alpha_j)^{\Delta-1} \right) \left(1 - \frac{\Delta}{T} \right) R_\Delta(\mathbf{r}) \right) \end{aligned} \tag{16}$$

See the Supplementary Information for detailed derivation of these equations. It is important to note that the approximations in Eqs 14 and 16 only hold when the number of trials, T , is large relative to the reciprocal of the run lengths ($1/\alpha_g$ and $1/\alpha_f$). This approximation simplifies the expressions greatly, by removing the dependence on initial values, but we urge caution in interpreting the results when both T and the learning rates are small. In particular, this implies that that our analysis holds for non-zero learning rates only.

Eqs 14 and 16 imply that to compute the required correlations we only need the statistics of the rewards: $\mu(\mathbf{r})$, $\mu(\mathbf{r}^2)$ and $R_\Delta(\mathbf{r})$. The exact form of these averages depends on the dynamics of the reward-generating process in the experiment. In the Results section we consider two commonly used experimental designs.

fMRI experiments

To test our theoretical predictions we used data from two different experiments corresponding to two different reward dynamics: fixed and drifting. In the following sections we briefly describe the two experiments along with details of our analyses. More precise descriptions of each experiment can be found in the original papers, also available as part of the supplementary information ([S1 Dataset](#)).

Fixed reward distribution. We used data from Niv, Edlund, Dayan & O’Doherty [10]. Preprocessing of the fMRI data were performed by the original authors, as described in [10]. In this experiment, 16 subjects made a series of choices between stimuli that paid out different amounts of monetary reward. There were five possible stimuli: two options paid out 0¢ with 100% probability, one paid out 20¢, one paid out 40¢, and one paid out either 0¢ or 40¢ with 50% probability (henceforth the risky 0/40 stimulus). The experiment involved two types of trials, intermingled: on ‘choice trials’ subjects were required to choose between two stimuli, while on ‘forced trials’ subjects were presented with only one of the five stimuli and had to choose it.

These forced trials ensured that subjects continued to experience all of the stimuli regardless of their subjective value (for example, the 0¢ stimuli were very rarely chosen on choice trials). Choices were made immediately after the stimuli appeared on screen and reward feedback was given 5s after the choice.

Reinforcement learning theory predicts that there would be two prediction errors on each trial: one at the time of stimulus onset/choice, equal to the value of the to-be-chosen stimulus V , and one at the time of reward, given by the difference between reward outcome and the value of the chosen option, $r - V$. Because subjects were trained on the task prior to the scan, we assumed that they knew the values for the constantly rewarding stimuli, and therefore concentrated on learning estimated values for the 0/40 stimulus, for which rewards were probabilistic. Note that the simple reinforcement-learning model we use is slightly different from the best-fitting model used in [10], which involved two different learning rates for positive and negative prediction errors. We used the simpler model to maintain consistency with our theoretical analysis. However, in any case our results show that such a difference in the model would not affect the regressors and the neural results to a large extent.

We focused our analysis on fMRI activations in the nucleus accumbens (NAc), an area whose BOLD activity has been repeatedly shown to correlate with prediction errors for both primary [2, 18–24] and monetary rewards [3, 10, 25–29], putatively due to the strong dopaminergic afferents to that area. We used average BOLD signals extracted by the original authors from anatomically defined regions of interest (ROIs) in the left and right NAc, and regressed this signal vector, Y , against parametric regressors for value and prediction error of the risky 0/40 stimulus as well as regressors for variables of less interest such as event onsets, value of the certain options and nuisance variables such as head motion and scanner drift. In keeping with our theoretical analysis, we mean-centered all model-based regressors and normalized them to have a standard deviation of 1. We repeated this analysis using different settings of the learning rate between 0.01 and 1, in steps of 0.01.

Drifting reward distribution. For the drifting reward distribution, we analyzed data from Daw, O’Doherty, Dayan, Seymour & Dolan [3]. Preprocessing of the fMRI data was done by the original authors, as described in [3]. In this experiment, 16 subjects performed two blocks of 150 choices between four options. Each option paid out probabilistic rewards sampled from a Gaussian distribution with a drifting mean and a noise standard deviation $\sigma_n = 4$. The drifting mean m_t on trial t was updated according to

$$m_{t+1} = \gamma m_t + 50(1 - \gamma) + n_t \quad (17)$$

where n_t was Gaussian random noise with drift standard deviation, $\sigma_d = 2.8$ and $\gamma = 0.9836$ was the decay rate that ensured that the mean decayed towards 50.

As in the experiment with fixed reward distributions [10], presentation of the stimulus and outcome were separated in time, and as before, we expected the signal at the stimulus to reflect the value of the to-be-chosen stimulus, V , and the signal at the outcome to reflect the prediction error $r - V$. We thus used a GLM with 5 regressors: two stick regressors, one at the onset of stimuli and one at the onset of the outcomes, a V parametric modulation on stimulus onset, a prediction error parametric modulation on outcome onset, and a parametric modulation on outcome onset by the magnitude of outcome itself, r . Because we were primarily interested in the value signal, we focused on data from an ROI in the ventromedial prefrontal cortex (vmPFC). Specifically, we used an ROI centered at (-3, 33, -6), a location that was reported in [3] to correlate strongly with choice probability, which is closely related to chosen value. As before, we analyzed this ROI with GLMs created using a variety of different learning rates between 0.01 and 1, in steps of 0.01.

Results

In the Methods, we developed general expressions that describe how changes in the learning rate affect model-based fMRI in simple reinforcement learning tasks. In particular, we showed that regression coefficients when using a model-based regressor, and their corresponding Student t values, depend on three factors: the contrast-to-noise ratio (CNR) in the signal, the number of data points in the regression, and $\rho(\mathbf{x}_g, \mathbf{x}_f)$ —the correlation between the regressor generated with the fit parameter values, \mathbf{x}_f , and a regressor generated with the ground truth parameter values, \mathbf{x}_g .

We now investigate how these factors play out in two cases with qualitatively different reward dynamics: a reward distribution that is fixed throughout the experiment, or one that changes over time. In both cases we show that model-based fMRI analysis of value and prediction error signals is relatively insensitive to the setting of the learning rate parameter and that this insensitivity can be, to a certain extent, manipulated by altering the design of the task.

Fixed reward distribution

We first consider a situation in which the reward distribution is fixed throughout the experiment. An example of such a distribution with mean m and variance σ_n^2 is shown in Fig 1A. In panels B and C of the same figure we show rewards from Gaussian and Bernoulli distributions, but it is important to note that the following theoretical results apply to *any* fixed reward distribution with finite variance.

Analytic results. Based on the methods developed above, the sensitivity of a model-based analysis looking for a value signal in the brain depends on the correlation, $\rho(\mathbf{V}_g, \mathbf{V}_f)$, between the value computed using the ground truth learning rate, \mathbf{V}_g , and the value computed using the

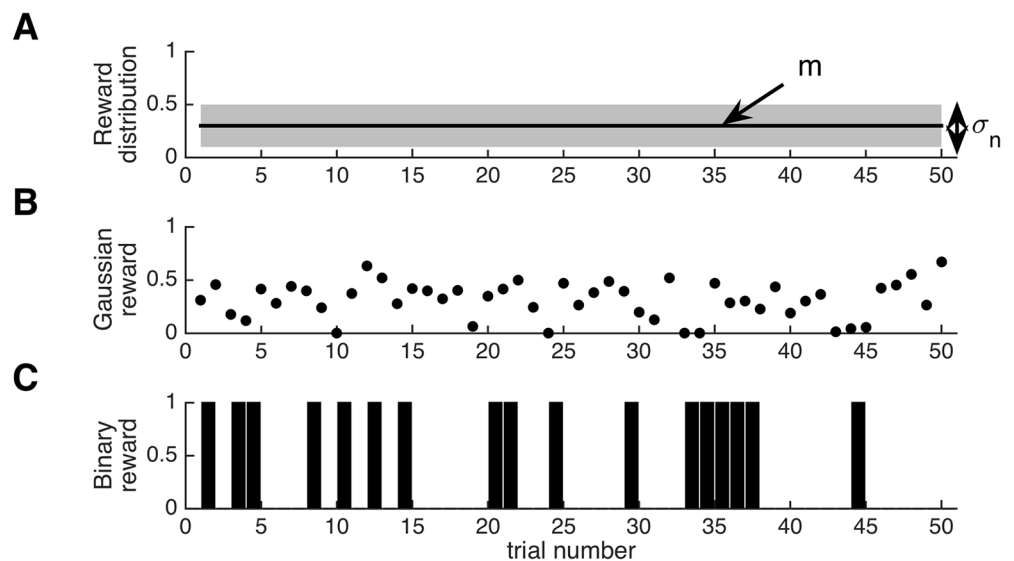


Fig 1. Illustration of a fixed reward distribution. (A) Schematic of a general reward distribution that is fixed over time, showing the relevant parameters: the mean m and standard deviation σ_n . (B) Example of continuous rewards sampled from a Gaussian distribution with mean $m = 0.3$ and standard deviation $\sigma_n = 0.2$. (C) Bar plot showing an example of binary reward data sampled from a Bernoulli distribution with the same mean $m = 0.3$ and standard deviation $\sigma_n = \sqrt{m(1-m)} \approx 0.46$.

doi:10.1371/journal.pcbi.1004237.g001

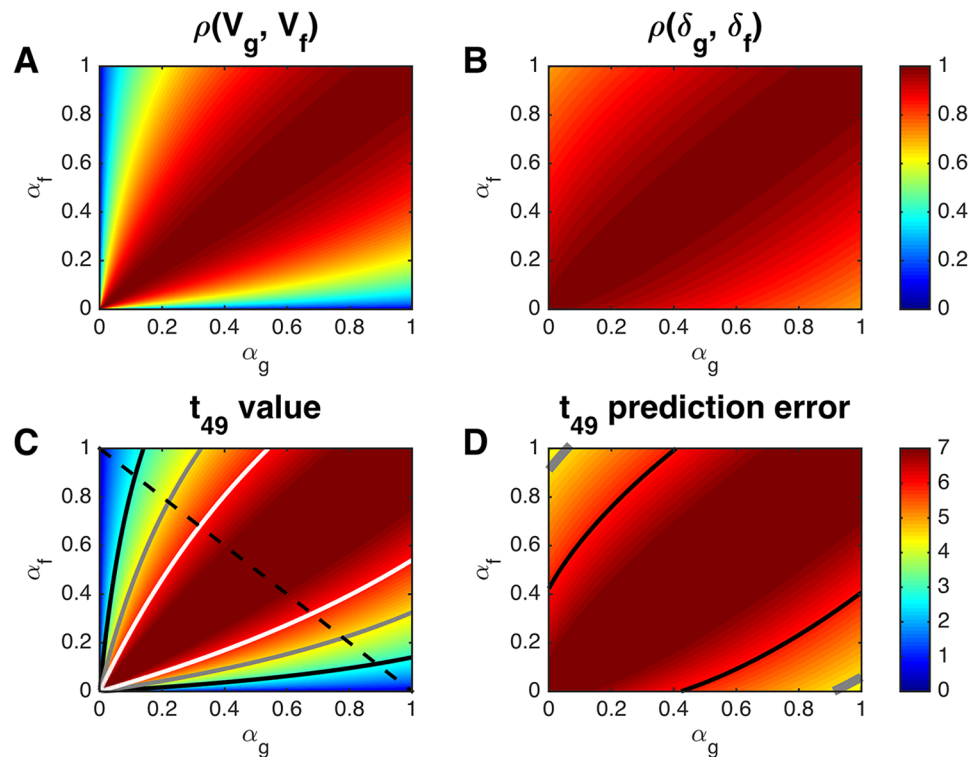


Fig 2. Correlations and t statistics for experiments with a fixed reward distribution, computed by evaluating Eq (18) at learning rates between 0.001 and 1, in steps of 0.001. (A) Correlation, $\rho(\mathbf{V}_g, \mathbf{V}_f)$, between regressors for value as a function of the true, α_g , and fitted, α_f , learning rates. (B) Correlation between the regressors for prediction error, $\rho(\delta_g, \delta_f)$. (C,D) Single-subject t statistics (assuming 49 degrees of freedom) as a function of the two learning rates for value (C) and prediction error (D). Black, gray and white contours denote significance at $p = 0.01$, $p = 10^{-4}$ and $p = 10^{-6}$, respectively. Dashed black line in C: values that will be analyzed in more detail in Fig 3.

doi:10.1371/journal.pcbi.1004237.g002

fit learning rate, \mathbf{V}_f . Likewise prediction error signals depend on $\rho(\delta_g, \delta_f)$. Moreover, expressions for these correlations rely only on a few statistics of the reward distribution: its mean $\mu(\mathbf{r}) = m$, the mean of the squared rewards $\mu(\mathbf{r}^2) = m^2 + \sigma_n^2$, and the reward autocorrelation $R_\Delta(\mathbf{r}) = m^2$. Given these expressions for the reward statistics, we can compute the sums in Eqs 14 and 16 exactly (as sums of geometric series), leading to the following expressions for the value and prediction error correlations using different learning rates α_g and α_f :

$$\begin{aligned} \rho(\mathbf{V}_g, \mathbf{V}_f) &= \frac{\sqrt{\alpha_g \alpha_f (2 - \alpha_g)(2 - \alpha_f)}}{\alpha_g + \alpha_f - \alpha_g \alpha_f} \\ \rho(\delta_g, \delta_f) &= \frac{(\alpha_g + \alpha_f) \sqrt{(2 - \alpha_g)(2 - \alpha_f)}}{2(\alpha_g + \alpha_f - \alpha_g \alpha_f)} \end{aligned} \tag{18}$$

In Fig 2A and 2B we plot these correlations as functions of the two learning rates. Strikingly, the correlations for both value and prediction error are relatively insensitive to mismatch in learning rates. Indeed, for prediction errors, the minimum possible value of the correlation (at $\alpha_g \rightarrow 0$ and $\alpha_f \rightarrow 1$ or vice versa) is $1/\sqrt{2} \approx 0.7$. This implies that even in the worst case

scenario, when the true learning rate had an extreme value of 0 or 1 and the fit learning rate was as far as possible from the true learning rate, the resulting prediction error regressor would still be highly correlated with the true signal. This result has important implications for the qualitative interpretation of prediction error signals, since with a true learning rate of 0, the ‘prediction error’ is simply the reward signal. Therefore, when the reward distribution is fixed throughout the experiment, prediction errors will always be strongly correlated with the reward signal, making it difficult to tease apart these two neural signals using linear regression.

In Fig 2C and 2D we show the corresponding t statistics assuming 49 degrees of freedom (or a very conservative total number of trials, $T = 50$) and a contrast-to-noise ratio, CNR of 1 (a fairly typical value in fMRI experiments [15] and also consistent with the range of CNR values seen in the data sets analyzed here, where CNR was 0.4 in one case and 11 in the other). As can be seen, the dependence of the t statistic for the regressor on the true and fit learning rates closely matches that of the regressor correlations. This is because at low CNR the t statistic in Eq 8 is approximately proportional to $\rho(\mathbf{x}_g, \mathbf{x}_f)$. For reward prediction error in particular, all possible values of α_g and α_f result in a significant t statistic at $p < 0.001$ (a commonly used uncorrected threshold for significance of prediction error signals in the brain). This result further exemplifies both the strength and the limitation of such an analysis: a neural signal will likely be identified even if model fitting produced especially poor learning rate parameter settings, however, from this regression alone we cannot know for sure that the identified signal is indeed a prediction error signal rather than a reward signal.

In Fig 3 we investigate the effect of CNR and T on the t statistic more explicitly. For exposition, we focus on the diagonal $\alpha_g + \alpha_f = 1$ (i.e. along the dashed line in Fig 2C) and plot the results as a function of the difference in learning rates, $\alpha_g - \alpha_f$, with the constraint that $\alpha_g + \alpha_f =$

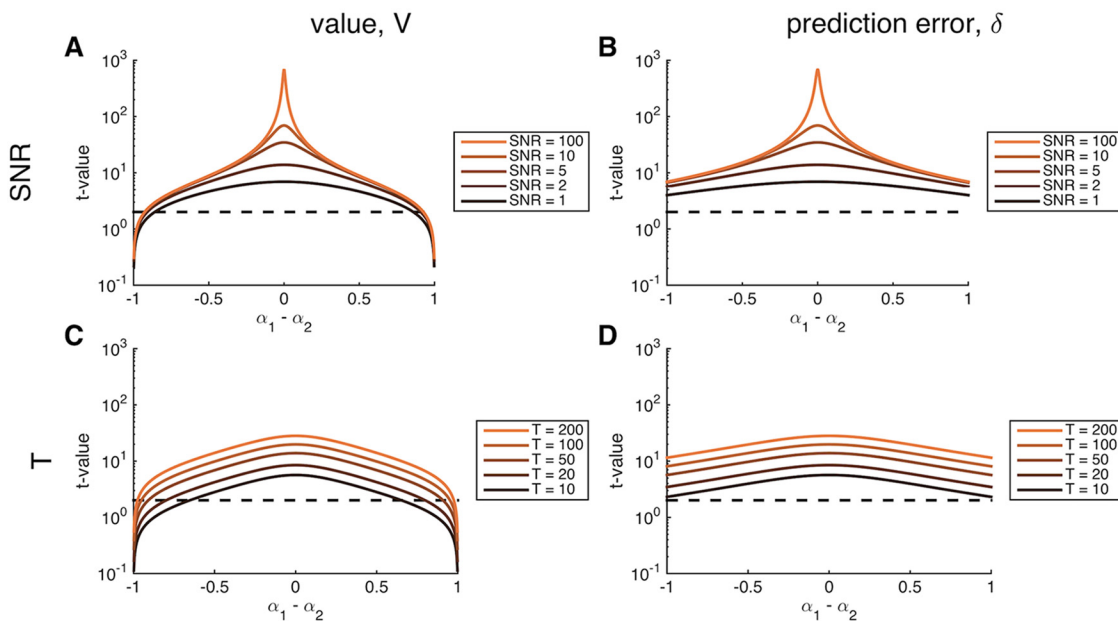


Fig 3. Effect of contrast-to-noise ratio, CNR, (A,B) and number of trials, T , (C,D) on the t values as a function of the difference between $\alpha_g - \alpha_f$, when $\alpha_g + \alpha_f = 1$, for experiments with a fixed reward distribution. The range of differences is from -0.999 to +0.999 in steps of 0.001. Dashed line: significance at $p < 0.05$ at the single subject level.

doi:10.1371/journal.pcbi.1004237.g003

1. Panels A and B show the effect of changing the contrast-to-noise ratio from 1 to 100 at fixed $T = 50$ for the value and prediction error regressors, respectively. In both cases, as the contrast-to-noise ratio increases, the curves become increasingly peaked at 0 (i.e. at $\alpha_g = \alpha_f$) indicating greater sensitivity of the resulting t statistics to accuracy of the fit learning rate. This shows that, when the underlying signal is strong, model-based fMRI will be much more sensitive to parameter settings than when the signal is weak. Of course, for a given statistical threshold, a higher CNR also results in a wider range of fit learning rates achieving statistical significance. Thus, not surprisingly, high CNR is always better than low CNR in that it makes the result more robust to model-fitting errors in addition to making it easier to use the fMRI signal to inform the parameter search.

In panels C and D we illustrate the effect of changing the number of trials, T , from 10 to 200 assuming a fixed CNR = 1. Again, as T increases, the t statistic becomes increasingly peaked around $\alpha_g = \alpha_f$ (note the logarithmic scale on the y-axis). These results show that the sensitivity of a model-based analysis can also be increased by increasing the number of trials in the experiment.

fMRI data. To test these theoretical predictions, we used data from from Niv, Edlund, Dayan & O’Doherty [10]. As Fig 4 shows, consistent with the theory, for both value and prediction error regressors, the regression coefficient depended relatively weakly on the learning rate used to generate the regressor. This was true both for single subjects (Fig 4A and 4B) and at the group level (Fig 4C and 4D).

Although the plots in Fig 4A and 4B are relatively insensitive to parameter value, many of the curves for individual subjects do appear to have a maximum, suggesting that there is a learning rate that best describes the neural data. To evaluate the extent to which these best-fitting learning rates could be estimated, we computed the log likelihood of the fMRI data for a linear model assuming that the data are generated by the combination of the model-based regressors for value and prediction error plus additive Gaussian noise, for each value of the learning rate parameter. For a linear regression model, the log likelihood has a closed form as follows:

$$LL(\alpha) = -T \left(\log(\sqrt{2\pi}\sigma_r(\alpha)) + \frac{1}{2} \right) \tag{19}$$

where $\sigma_r(\alpha)$ denotes the standard deviation of the residuals of the linear model with model-based regressors generated using learning rate α .

Fig 5 shows $LL(\alpha)$ relative to its maximum value, for each subject, $\Delta LL(\alpha) = LL(\alpha) - \max_{\alpha} LL(\alpha)$. This analysis further highlights the insensitivity of these results to learning rate—for most subjects the range of log likelihood between best and worst fits being less than 2, a difference in fit usually considered ‘barely worth mentioning’ [30].

Drifting reward distribution

Our approach can also be applied to scenarios in which the reward distribution is not fixed. To illustrate, we analyze experiments with rewards that are drawn from a Gaussian distribution whose mean, m_t , is generated by a discretized Ornstein-Uhlenbeck process (Fig 6) [31]. Specifically, m_t , undergoes a random walk defined by

$$m_{t+1} = \gamma m_t + n_t \tag{20}$$

where n_t is zero mean noise with drift variance σ_d^2 , and $\gamma (< 1)$ is a decay parameter. Because γ is smaller than one, the mean tends to decay to zero over time (illustrated by the arrows in

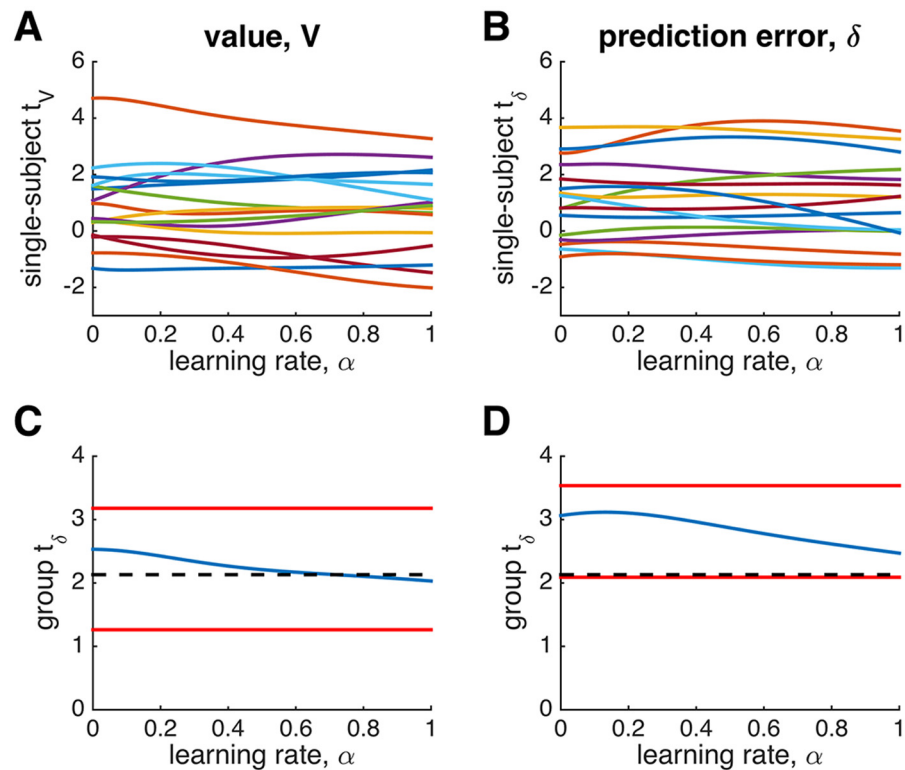


Fig 4. The effect of learning rate on the fMRI signal in the NAC in [10], an experiment with a fixed reward distribution (binary rewards with $p = 0.5$, see Methods). (A) Regression coefficients for value of the chosen option at the time of stimulus onset, as a function of learning rate. Each curve represents a single subject. (B) Single subject regression coefficients for prediction error at the time of reward. Note the relative lack of modulation of the regression coefficient by the value of the learning rate parameter. (C) Group analysis at the time of stimulus showing the group t statistic as a function of (group-wise) learning rate (blue). Red lines denote the best and worst case scenarios obtained by taking the value of the learning rate that either maximizes or minimizes the t statistic for each subject. Dashed black line: $p = 0.05$ threshold. (D) Group analysis at the time of reward. In all cases, as predicted, the effect of the learning rate parameter is small.

doi:10.1371/journal.pcbi.1004237.g004

Fig 6A). This helps to keep the means of different options from diverging too far as the experiment progresses.

Analytic results. Again, to compute the correlations we require the statistics of the reward distribution. For simplicity, we focus on situations in which T is large, where the reward statistics are asymptotically

$$\mu(\mathbf{r}) = 0 \tag{21}$$

$$\mu(\mathbf{r}^2) = \sigma_n^2 + \frac{\sigma_d^2}{1 - \gamma^2} \tag{22}$$

$$R_\Delta(\mathbf{r}) = \frac{\gamma^\Delta \sigma_d^2}{1 - \gamma^2}. \tag{23}$$

These reward statistics allow us to compute the sums in Eqs 14 and 16 exactly, leading to the

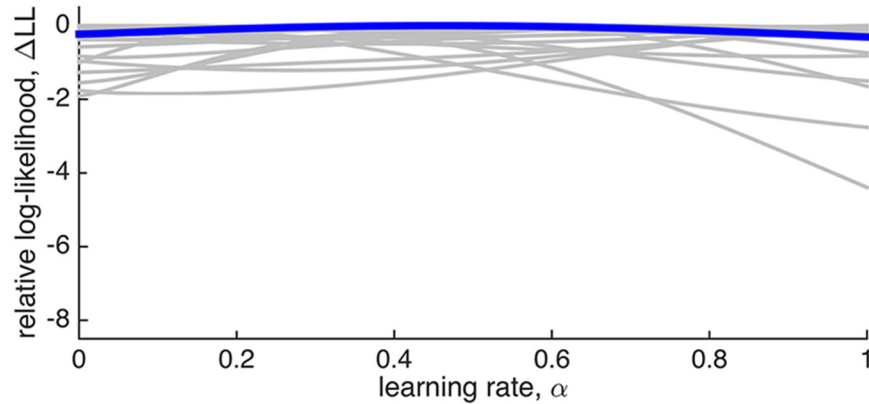


Fig 5. Fit of a linear model comprised of a single model-based regressor generated with different learning rates α , to fMRI data in the fixed reward probability experiment [10] as a function of learning rate. Each grey curve corresponds to a different subject and in blue is the mean across subjects. All curves are shifted to have a maximum value of 0. For most subjects the quality of the fit depends only weakly on the learning rate. Note that the y-axis is the same as in Fig 10, to highlight the differences between the sensitivities of the two experiments to the setting of the learning rate parameter.

doi:10.1371/journal.pcbi.1004237.g005

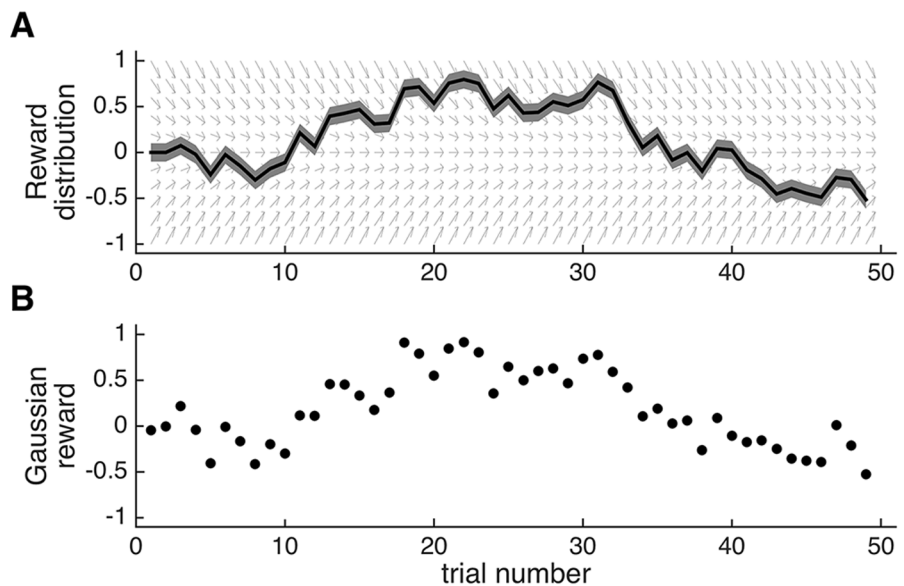


Fig 6. An example drifting reward distribution. (A) Evolution of the mean m_t over time, t , diffusing with a drift standard deviation σ_d . The decay, γ , is indicated by the gray arrows and the shaded region indicates the standard deviation of the Gaussian noise distribution, σ_n . (B) A set of rewards sampled from the distribution in panel A.

doi:10.1371/journal.pcbi.1004237.g006

following (slightly messy, but nonetheless fully tractable) expressions for $\rho(\mathbf{V}_g, \mathbf{V}_f)$ and $\rho(\delta_g, \delta_f)$:

$$\rho(\mathbf{V}_g, \mathbf{V}_f) = \frac{\left(1 + \left(\frac{1}{1 - \gamma + \alpha_g \gamma} + \frac{1}{1 - \gamma + \alpha_f \gamma} - 1\right) \frac{\sigma_d^2}{(1 - \gamma^2)\sigma_n^2}\right)}{\alpha_g + \alpha_f - \alpha_g \alpha_f} \sqrt{\frac{\alpha_g \alpha_f (2 - \alpha_g)(2 - \alpha_f)}{\left(1 + \frac{2}{(1 - \gamma + \alpha_g \gamma)} \frac{\sigma_d^2}{(1 - \gamma^2)\sigma_n^2}\right) \left(1 + \frac{2}{1 - \gamma + \alpha_f \gamma} \frac{\sigma_d^2}{(1 - \gamma^2)\sigma_n^2}\right)}} \quad (24)$$

$$\rho(\delta_g, \delta_f) = \frac{(\alpha_g + \alpha_f) + \left(\frac{\alpha_g}{1 - \gamma + \alpha_g \gamma} + \frac{\alpha_f}{1 - \gamma + \alpha_f \gamma}\right) \frac{\sigma_d^2}{(1 + \gamma)\sigma_n^2}}{2(\alpha_g + \alpha_f - \alpha_g \alpha_f)} \sqrt{\frac{(2 - \alpha_g)(2 - \alpha_f)}{\left(1 + \frac{1}{1 - \gamma + \alpha_g \gamma} \frac{\sigma_d^2}{(1 + \gamma)\sigma_n^2}\right) \left(1 + \frac{1}{1 - \gamma + \alpha_f \gamma} \frac{\sigma_d^2}{(1 + \gamma)\sigma_n^2}\right)}}$$

Note that these are functions of only two parameters of the reward distribution: the decay, γ , and the ratio of drift variance to noise variance, σ_d^2/σ_n^2 . When using a drifting reward probability, these experimentally determined parameters exert much greater control over the form of the correlations than can be achieved when reward probability is fixed. This control is demonstrated in Fig 7 where we tuned these two parameters to achieve sensitivity to prediction errors (panels A,B), value (panels E,F), or both (panels C,D). The parameters in panels A and B, $\gamma = 0.98$ and $\sigma_d/\sigma_n = 0.7$, closely match those in the experiment of Daw et al. [3], whose neural results we discuss below.

To explore the parameter space more thoroughly, we quantified the ‘insensitivity to learning rate’ as the fraction of (α_g, α_f) -space in which the correlations are greater than 0.7. This metric is 1 when the correlations are only weakly dependent on learning rate (as for the prediction error in the case of fixed rewards) and 0 when they are exquisitely sensitive. Fig 8 shows this metric as a function of the two parameters, γ and σ_d/σ_n , for the value and prediction error regressors. The plot demonstrates the somewhat reciprocal relationship between $\rho(\mathbf{V}_g, \mathbf{V}_f)$ and $\rho(\delta_g, \delta_f)$: when prediction errors have higher sensitivity to learning rate, the values tend to have lower sensitivity, and vice versa. Thus, while the sensitivity to learning rate can be tuned, there is a tradeoff between sensitivity to model-based regressors for prediction errors and value.

fMRI results. To test the theoretical predictions, we analyzed BOLD data from Daw et al. [3], performing a GLM analysis with a variety of different learning rates and examining value and prediction error regressors. The resulting regression coefficients and group *t* statistics are shown in Fig 9. Here we see much less sensitivity to the learning rate of the chosen value signal than the prediction error signal at both the single subject and group levels. This is in line with our predictions, as the reward parameters in the experiment ($\gamma \approx 0.98$ and $\sigma_d/\sigma_n = 0.7$) place it in the upper left of Fig 8, where values are more sensitive to the fit learning rate than are prediction errors.

A notable feature of these results is that the different regressors are significantly correlated with the neural signal in different regions of reward-parameter space, with prediction errors significantly correlated with BOLD signals when using low learning rates and values significant at higher learning rate. This reflects the fact that with low learning rates value changes slowly, and so prediction errors are more correlated with the (surprising and drifting) outcomes, whereas for high learning rates it is the value that closely tracks the drifting outcomes. It may not be surprising that the correlation between the different regressors and trial outcomes drives the significance of the regression result in vmPFC, as this area has been repeatedly associated with encoding of outcome magnitude [32, 33]. However, this result highlights again an important (and worrying) point: while the overall regression coefficients can be remarkably robust to

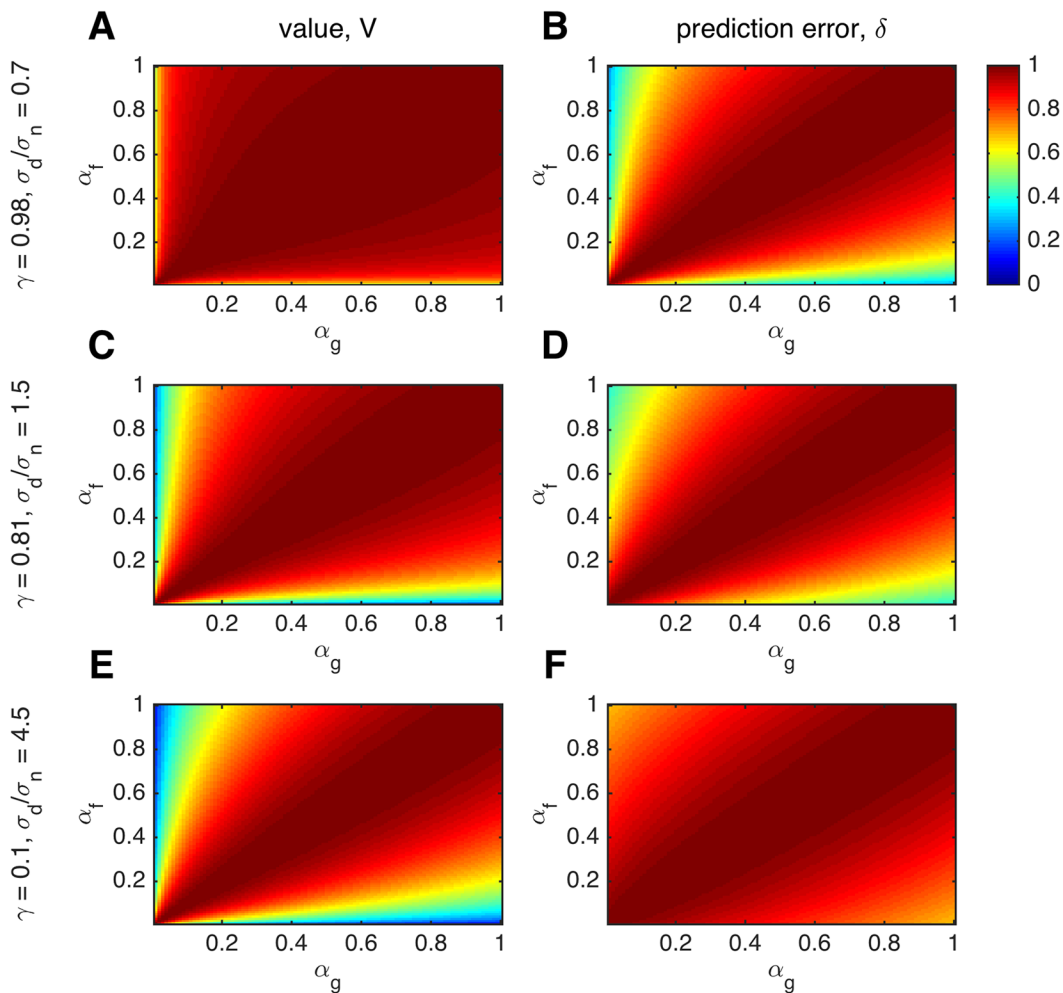


Fig 7. Correlations between model based regressors derived using different learning rates, in an experiment with drifting rewards, for three different settings of the decay of the reward mean to 0, γ , and the drift-to-noise ratio of the reward mean, σ_d/σ_n . Plots were generated by evaluating Eq (24) for learning rates between 0.001 and 1 in steps of 0.001. (A,B) When γ is high (0.98) and σ_d/σ_n is low (0.7), values are not sensitive to fit learning rate, but prediction errors are sensitive. (C,D) Intermediate γ and σ_d/σ_n lead to intermediate sensitivity of both value and prediction error to learning rate. (E, F) When γ is low (0.1) and σ_d/σ_n is high (4.5), the results mimic those obtained with a fixed reward distribution (values are more sensitive to fit learning rate than are prediction errors, compare Fig 2A and 2B).

doi:10.1371/journal.pcbi.1004237.g007

the fit learning rate, interpretation of what function a neural area fulfills can change significantly as the fit parameter values change.

To further investigate the sensitivity of our results to settings of the learning rate, we again computed the log likelihood of the neural data for linear regression models using model-based regressors with different learning rates. These results are shown in Fig 10. Unlike the case of constant reward probability (Fig 5), in this experiment we found much stronger dependence of the log likelihood on learning rate, likely due to the increased contrast-to-noise ratio for the larger vmPFC ROI (11 here, compared to 0.4 for the NAc ROI). This increased sensitivity also allows us to extract a potentially meaningful fit of the learning rate to the fMRI data—on average 0.36 (± 0.08 [s.e.m.]). Of course, this analysis is only suggestive and one should be carefully interpreting group-averaged statistics.

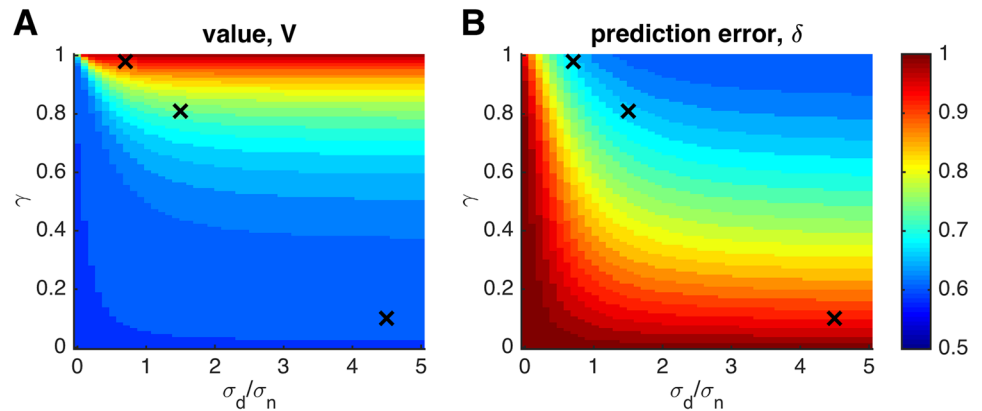


Fig 8. Insensitivity of value (A) and prediction error (B) regressors to the fit learning rate as a function of decay of the reward mean to zero, γ , and the drift variance to noise variance ratio of the reward mean, σ_d/σ_n , in experiments with drifting rewards. The three black crosses indicate the parameter values in the examples in Fig 7.

doi:10.1371/journal.pcbi.1004237.g008

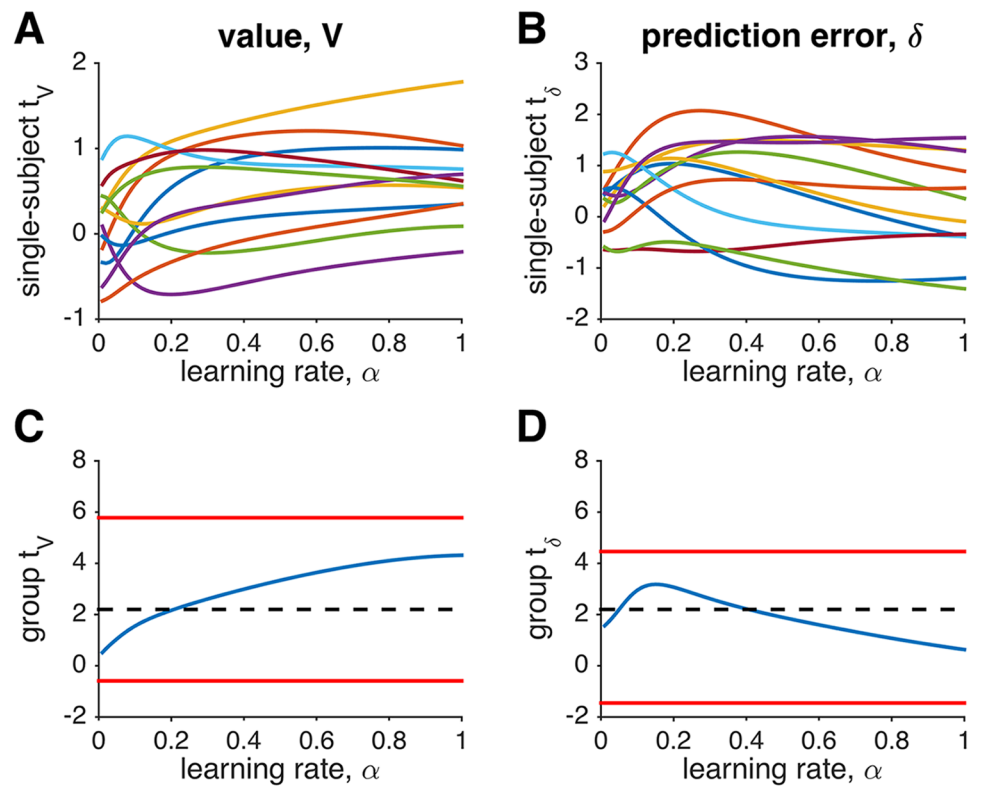


Fig 9. The effect of learning rate on the BOLD signal in vmPFC in [3], an experiment with drifting rewards. (A) Regression coefficients for value of the chosen option at the time of stimulus onset, as a function of learning rate. Each curve represents a single subject. (B) Single subject regression coefficients for prediction error at the time of reward. (C) Group analysis of value signals at the time of stimulus showing the group t statistic as a function of (group-wise) learning rate (blue). Red lines denote the best and worst case scenarios obtained by taking the value of the learning rate that either maximizes or minimizes the t statistic for each subject. Dashed black line: $p = 0.05$ threshold. (D) Group analysis of prediction error signals at the time of reward.

doi:10.1371/journal.pcbi.1004237.g009

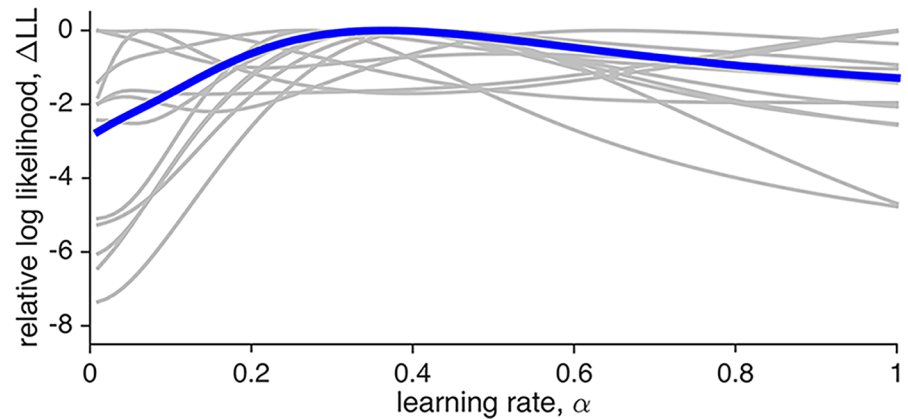


Fig 10. Fit of a linear model comprised of a single model-based regressor generated with different learning rates α , to fMRI data in the drifting reward experiment [3] as a function of learning rate. Each grey curve corresponds to a different subject and in blue is the mean across subjects. All curves are shifted to have a maximum value of 0. The y-axis is the same as that for Fig 5, to highlight the differences between the two experiments.

doi:10.1371/journal.pcbi.1004237.g010

To decide between different accounts of vmPFC activity—value, prediction error, or both—one could use a similar method to compare the goodness of fit of different models and assess, at the group level, which model fits the data best. In particular, one could compare three distinct linear models: one with a regressor for value but not prediction error, one with a regressor for prediction error but not value and one with both (for instance, generated using the best-fit learning rate). The log-likelihood measure (corrected for the different number of parameters, in this case, the number of regressors) could then be compared to determine the best model. We note that while such *model comparison* is closely related to the questions of parameter fitting and parameter estimation we consider here, it comes with none of the guarantees that we have established for parameter fitting.

Discussion

In this paper, we considered the extent to which errors in the estimation of model parameters impact model-based fMRI. We showed that, in general, the answer to this question depends crucially on the correlation between regressors derived from different parameterizations of the model, $\rho(\mathbf{x}_g, \mathbf{x}_f)$, and is further affected by the contrast-to-noise ratio in the data, CNR, and the number of trials, T , in the experiment. In the specific case where the fit parameter is the learning rate in a reinforcement learning model, we found that regressors for both value and prediction error signals were fairly insensitive to the fit learning rate, such that for realistic values of CNR and T , the results of the model-based analysis were predicted to be robust to different parameterizations. Indeed for an experiment with a fixed reward distribution, the estimated learning rate had close to no effect on the detection of prediction error signals in the NAc either in theory or in the experimental data. Similar results also held when rewards were drawn from a Gaussian distribution with a randomly drifting mean.

These findings are consistent with the report from one of the earliest model-based fMRI papers [18], in which changing the learning rate from 0.2 to 0.7 was found to have relatively little effect on the results. However, when either the contrast-to-noise ratio or number of trials is high, sensitivity of the model-based analysis to learning rate can increase. This might explain

the anecdotal finding (personal communication, J.P. O’Doherty) that the results reported in Bray & O’Doherty [34] were relatively sensitive to learning rate. In particular, this study had more trials ($T = 288$) than in either [18] or [10] and also used ‘natural’ rewards (in the form of good- and bad-looking faces) instead of monetary rewards, which might lead to a larger effect and hence greater CNR.

Our results hold important consequences for the interpretation of model-based fMRI experiments. As regards learning rate, the relative insensitivity to this parameter is both good news and bad news. For studies investigating what areas in the brain are involved in reinforcement learning, these results are good news as the robustness to the fit parameters will make errors in the fitting procedure inconsequential. In this sense, our philosophy diverges slightly from that of Forstmann and colleagues [35] who suggest redesigning either the model or the experiment if parameters cannot be estimated with sufficient accuracy. In contrast, we espouse the position that imperfect parameter recovery can be tolerated if the scientific question of interest can be answered without it, as it can, for example, when we wish to know *where* reinforcement learning signals are located in the brain.

For studies that ask more nuanced questions, such as whether a particular signal is a reward signal, a value signal or a prediction error signal, or whether different areas use different learning rates, the insensitivity of the neural analysis to learning rate means that a simple analysis is not sufficient. In these cases, there is special premium for clever task design [29], and a more detailed analysis, for instance requiring that a putative neural prediction error signal correlate significantly with all its theoretical subcomponents [10]. Our analysis also suggests a way to minimize this problem: changing the experiment, either by optimizing the dynamics of the reward distribution or increasing the number of trials, can substantially change the sensitivity to learning rate.

The analysis we are suggesting bears resemblance to calculations of statistical power. Statistical power refers to the probability that a specific experiment will be successful in detecting an effect that truly exists—it is obvious why this is an important quantity to optimize in experiment design. Indeed many of the manipulations that we suggest—such as increasing the number of trials—will also improve statistical power. For cases in which the effect one is looking for involves differences in model parameters, we suggest a formula for testing in advance whether these differences are likely to be detectable neurally.

Of course, the fact that parameter values may be difficult to infer from brain data does not mean that they are not inferable at all. In many (if not all) cases, suitable behavioral data can provide strong constraints on model selection and parameter fitting. The ‘power’ of this type of analysis can also be tested, for example by recovering parameters from simulated data [36] and using data simulated by different models to test for confusion between these models [37]. Nevertheless, it is not obvious that parameters that provide a good description of behavior will necessarily correspond to processes in any brain area. For example, behavior could be driven by a combination of several distinct processes each with different parameter values [38–40].

More generally, for parameters other than the learning rate (for example, the discount factor in inter-temporal choice, or the softmax parameter in bandits tasks) our results highlight the importance of testing parameter sensitivity *before* running the experiment. This need not be done analytically (as was the case here) but can be approximated easily using simulations. As our results show, it is often possible to increase or decrease sensitivity to a particular variable by changing the parameters of the task and, with a clear focus on the goal of the model-based analysis, one could use such simulations to optimize experiment design.

Finally, while in this paper we have focused on the sensitivity of model-based fMRI to the parameters of a single model, an important question for future work is the extent to which fMRI can be used to adjudicate between *different* models. Such model comparison would

involve computing goodness-of-fit measures (such as the log likelihoods we computed above) for each model and asking which model fit the fMRI data best. The extent to which models can be distinguished based on neural data is related to the degree of divergence of the predictions of the two models (i.e., the correlation between the regressors of the different models). However, it is also likely related to how close the compared models are to the ground-truth generative process that underlies the fMRI data, for which we unfortunately have no *a priori* guarantees.

Supporting Information

S1 Text. Detailed derivation of the statistics of the value and prediction error regressors.
(PDF)

S2 Text. Effect of number of trials, T , in real data.
(PDF)

S1 Dataset. NAc ROI data from [10].
(ZIP)

Acknowledgments

We thank Nathaniel Daw, Peter Dayan, John O’Doherty, and Ray Dolan for helpful comments and for generously sharing their data and Catherine Hartley, Reka Daniel, Nicolas Schuck and Amitai Shenhav for helpful suggestions and comments on the manuscript.

Author Contributions

Conceived and designed the experiments: RCW YN. Performed the experiments: RCW. Analyzed the data: RCW. Contributed reagents/materials/analysis tools: RCW YN. Wrote the paper: RCW YN.

References

1. Tanaka SC, Doya K, Okada G, Ueda K, Okamoto Y, et al. (2004) Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nat Neurosci* 7: 887–93. PMID: [15235607](#)
2. O’Doherty JP, Dayan P, Schultz J, Deichmann R, Friston K, et al. (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304: 452–4. doi: [10.1126/science.1094285](#) PMID: [15087550](#)
3. Daw ND, O’Doherty JP, Dayan P, Seymour B, Dolan RJ (2006) Cortical substrates for exploratory decisions in humans. *Nature* 441: 876–879. doi: [10.1038/nature04766](#) PMID: [16778890](#)
4. Preuschoff K, Bossaerts P, Quartz SR (2006) Neural differentiation of expected reward and risk in human subcortical structures. *Neuron* 51: 381–90. doi: [10.1016/j.neuron.2006.06.024](#) PMID: [16880132](#)
5. O’Doherty JP, Hampton A, Kim H (2007) Model-based fMRI and its application to reward learning and decision making. *Ann N Y Acad Sci* 1104: 35–53. doi: [10.1196/annals.1390.022](#) PMID: [17416921](#)
6. Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS (2007) Learning the value of information in an uncertain world. *Nature Neuroscience* 10: 1214–1221. doi: [10.1038/nn1954](#) PMID: [17676057](#)
7. Boorman ED, Behrens TEJ, Woolrich MW, Rushworth MFS (2009) How green is the grass on the other side? frontopolar cortex and the evidence in favor of alternative courses of action. *Neuron* 62: 733–43. doi: [10.1016/j.neuron.2009.05.014](#) PMID: [19524531](#)
8. Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011) Model-based influences on humans’ choices and striatal prediction errors. *Neuron* 69: 1204–15. doi: [10.1016/j.neuron.2011.02.027](#) PMID: [21435563](#)
9. Badre D, Doll BB, Long NM, Frank MJ (2012) Rostrolateral prefrontal cortex and individual differences in uncertainty-driven exploration. *Neuron* 73: 595–607. doi: [10.1016/j.neuron.2011.12.025](#) PMID: [22325209](#)

10. Niv Y, Edlund JA, Dayan P, O'Doherty JP (2012) Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. *The Journal of Neuroscience* 32: 551–562. doi: [10.1523/JNEUROSCI.5498-10.2012](https://doi.org/10.1523/JNEUROSCI.5498-10.2012) PMID: [22238090](https://pubmed.ncbi.nlm.nih.gov/22238090/)
11. Eppinger B, Schuck NW, Nystrom LE, Cohen JD (2013) Reduced striatal responses to reward prediction errors in older compared with younger adults. *J Neurosci* 33: 9905–12. doi: [10.1523/JNEUROSCI.2942-12.2013](https://doi.org/10.1523/JNEUROSCI.2942-12.2013) PMID: [23761885](https://pubmed.ncbi.nlm.nih.gov/23761885/)
12. Boorman ED, O'Doherty JP, Adolphs R, Rangel A (2013) The behavioral and neural mechanisms underlying the tracking of expertise. *Neuron* 80: 1558–1571. doi: [10.1016/j.neuron.2013.10.024](https://doi.org/10.1016/j.neuron.2013.10.024) PMID: [24360551](https://pubmed.ncbi.nlm.nih.gov/24360551/)
13. Donoso M, Collins AGE, Koehlin E (In Press) Foundations of human reasoning in the prefrontal cortex. *Science*.
14. Daw ND (2011) Trial by trial data analysis using computational models. In: Delgado M, Phelps E, Robbins T, editors, *Decision making, affect, and learning: attention and performance*, Oxford UP, Oxford, UK, XXIII. pp. 3–38.
15. Welvaert M, Rosseel Y (2013) On the definition of signal-to-noise ratio and contrast-to-noise ratio for fmri data. *PLoS One* 8: e77089. doi: [10.1371/journal.pone.0077089](https://doi.org/10.1371/journal.pone.0077089) PMID: [24223118](https://pubmed.ncbi.nlm.nih.gov/24223118/)
16. Niv Y, Schoenbaum G (2008) Dialogues on prediction errors. *Trends Cogn Sci* 12: 265–72. doi: [10.1016/j.tics.2008.03.006](https://doi.org/10.1016/j.tics.2008.03.006) PMID: [18567531](https://pubmed.ncbi.nlm.nih.gov/18567531/)
17. Rescorla RA, Wagner AR (1972) A theory of Pavlovian conditioning variations in the effectiveness of reinforcement and nonreinforcement.
18. O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ (2003) Temporal difference models and reward-related learning in the human brain. *Neuron* 38: 329–37. doi: [10.1016/S0896-6273\(03\)00169-7](https://doi.org/10.1016/S0896-6273(03)00169-7) PMID: [12718865](https://pubmed.ncbi.nlm.nih.gov/12718865/)
19. McClure SM, Berns GS, Montague PR (2003) Temporal prediction errors in a passive learning task activate human striatum. *Neuron* 38: 339–46. doi: [10.1016/S0896-6273\(03\)00154-5](https://doi.org/10.1016/S0896-6273(03)00154-5) PMID: [12718866](https://pubmed.ncbi.nlm.nih.gov/12718866/)
20. Abler B, Walter H, Erk S, Kammerer H, Spitzer M (2006) Prediction error as a linear function of reward probability is coded in human nucleus accumbens. *Neuroimage* 31: 790–5. doi: [10.1016/j.neuroimage.2006.01.001](https://doi.org/10.1016/j.neuroimage.2006.01.001) PMID: [16487726](https://pubmed.ncbi.nlm.nih.gov/16487726/)
21. Li J, McClure SM, King-Casas B, Montague PR (2006) Policy adjustment in a dynamic economic game. *PLoS One* 1: e103. doi: [10.1371/journal.pone.0000103](https://doi.org/10.1371/journal.pone.0000103) PMID: [17183636](https://pubmed.ncbi.nlm.nih.gov/17183636/)
22. Preuschoff K, Bossaerts P, Quartz SR (2006) Neural differentiation of expected reward and risk in human subcortical structures. *Neuron* 51: 381–90. doi: [10.1016/j.neuron.2006.06.024](https://doi.org/10.1016/j.neuron.2006.06.024) PMID: [16880132](https://pubmed.ncbi.nlm.nih.gov/16880132/)
23. Tobler PN, O'Doherty JP, Dolan RJ, Schultz W (2006) Human neural learning depends on reward prediction errors in the blocking paradigm. *J Neurophysiol* 95: 301–10. doi: [10.1152/jn.00762.2005](https://doi.org/10.1152/jn.00762.2005) PMID: [16192329](https://pubmed.ncbi.nlm.nih.gov/16192329/)
24. Seymour B, Daw N, Dayan P, Singer T, Dolan R (2007) Differential encoding of losses and gains in the human striatum. *J Neurosci* 27: 4826–31. doi: [10.1523/JNEUROSCI.0400-07.2007](https://doi.org/10.1523/JNEUROSCI.0400-07.2007) PMID: [17475790](https://pubmed.ncbi.nlm.nih.gov/17475790/)
25. Kuhnen CM, Knutson B (2005) The neural basis of financial risk taking. *Neuron* 47: 763–70. doi: [10.1016/j.neuron.2005.08.008](https://doi.org/10.1016/j.neuron.2005.08.008) PMID: [16129404](https://pubmed.ncbi.nlm.nih.gov/16129404/)
26. Knutson B, Taylor J, Kaufman M, Peterson R, Glover G (2005) Distributed neural representation of expected value. *J Neurosci* 25: 4806–12. doi: [10.1523/JNEUROSCI.0642-05.2005](https://doi.org/10.1523/JNEUROSCI.0642-05.2005) PMID: [15888656](https://pubmed.ncbi.nlm.nih.gov/15888656/)
27. Kim H, Shimojo S, O'Doherty JP (2006) Is avoiding an aversive outcome rewarding? neural substrates of avoidance learning in the human brain. *PLoS Biol* 4: e233. doi: [10.1371/journal.pbio.0040233](https://doi.org/10.1371/journal.pbio.0040233) PMID: [16802856](https://pubmed.ncbi.nlm.nih.gov/16802856/)
28. Schönberg T, Daw ND, Joel D, O'Doherty JP (2007) Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. *J Neurosci* 27: 12860–7. doi: [10.1523/JNEUROSCI.2496-07.2007](https://doi.org/10.1523/JNEUROSCI.2496-07.2007) PMID: [18032658](https://pubmed.ncbi.nlm.nih.gov/18032658/)
29. Hare TA, O'Doherty J, Camerer CF, Schultz W, Rangel A (2008) Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *J Neurosci* 28: 5623–30. doi: [10.1523/JNEUROSCI.1309-08.2008](https://doi.org/10.1523/JNEUROSCI.1309-08.2008) PMID: [18509023](https://pubmed.ncbi.nlm.nih.gov/18509023/)
30. Kass RE, Raftery AE (1995) Bayes factors. *Journal of the American Statistical Association* 430: 773–795. doi: [10.1080/01621459.1995.10476572](https://doi.org/10.1080/01621459.1995.10476572)
31. Uhlenbeck GE, Ornstein LS (1930) On the theory of Brownian motion. *Physical Review* 36: 823–841. doi: [10.1103/PhysRev.36.823](https://doi.org/10.1103/PhysRev.36.823)
32. McNamee D, Rangel A, O'Doherty JP (2013) Category-dependent and category-independent goal-value codes in human ventromedial prefrontal cortex. *Nat Neurosci* 16: 479–85. doi: [10.1038/nn.3337](https://doi.org/10.1038/nn.3337) PMID: [23416449](https://pubmed.ncbi.nlm.nih.gov/23416449/)

33. Bartra O, McGuire JT, Kable JW (2013) The valuation system: A coordinate-based meta-analysis of bold fmri experiments examining neural correlates of subjective value. *Neuroimage* 76: 412–427. doi: [10.1016/j.neuroimage.2013.02.063](https://doi.org/10.1016/j.neuroimage.2013.02.063) PMID: [23507394](https://pubmed.ncbi.nlm.nih.gov/23507394/)
34. Bray S, O'Doherty J (2007) Neural coding of reward-prediction error signals during classical conditioning with attractive faces. *J Neurophysiol* 97: 3036–45. doi: [10.1152/jn.01211.2006](https://doi.org/10.1152/jn.01211.2006) PMID: [17303809](https://pubmed.ncbi.nlm.nih.gov/17303809/)
35. Forstmann BU, Wagenmakers EJ, Eichele T, Brown S, Serences JT (2011) Reciprocal relations between cognitive neuroscience and formal cognitive models: opposites attract? *Trends Cogn Sci* 15: 272–9. doi: [10.1016/j.tics.2011.04.002](https://doi.org/10.1016/j.tics.2011.04.002) PMID: [21612972](https://pubmed.ncbi.nlm.nih.gov/21612972/)
36. Halpern D, Gureckis T (2013). On the identifiability of parameters in reinforcement learning models. URL <http://gureckislab.org/blog/?p=3450>.
37. Wilson RC, Niv Y (2011) Inferring relevance in a changing world. *Frontiers in Human Neuroscience* 5: 189. doi: [10.3389/fnhum.2011.00189](https://doi.org/10.3389/fnhum.2011.00189) PMID: [22291631](https://pubmed.ncbi.nlm.nih.gov/22291631/)
38. Bornstein AM, Daw ND (2012) Dissociating hippocampal and striatal contributions to sequential prediction learning. *Eur J Neurosci* 35: 1011–23. doi: [10.1111/j.1460-9568.2011.07920.x](https://doi.org/10.1111/j.1460-9568.2011.07920.x) PMID: [22487032](https://pubmed.ncbi.nlm.nih.gov/22487032/)
39. Wilson RC, Nassar MR, Gold JI (2013) A mixture of Delta-rules approximation to Bayesian inference in change-point problems. *PLoS Computational Biology* 9: e1003150. doi: [10.1371/journal.pcbi.1003150](https://doi.org/10.1371/journal.pcbi.1003150) PMID: [23935472](https://pubmed.ncbi.nlm.nih.gov/23935472/)
40. Redish AD, Kurth-Nelson Z (2010) Neural models of temporal discounting. In: *Impulsivity: Theory, Science, and Neuroscience of Discounting*, APA books, chapter 5. pp. 123–158.