

---

# Is model fitting necessary for model-based fMRI?

---

**Robert C. Wilson**  
Princeton Neuroscience Institute  
Princeton University  
Princeton, NJ 08540  
rcw2@princeton.edu

**Yael Niv**  
Princeton Neuroscience Institute  
Princeton University  
Princeton, NJ 08540  
yael@princeton.edu

## Abstract

Model-based analysis of functional magnetic resonance imaging (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 latent variables that are key to specific algorithms, such as prediction errors in temporal difference learning. A key step in this type of analysis is model fitting. Most commonly, a model is first fit to behavioral data to establish ‘good’ parameters. These are then used to generate model-based regressors of the quantity of interest, for regressing against brain activations acquired using fMRI. While such model fitting may intuitively seem like good practice, in this work we ask whether it is really necessary. We focus on the classic reinforcement learning regressors for value and prediction error and examine their sensitivity to perturbations of the learning rate parameter both in theory and in a previously published dataset. Surprisingly, in many cases, we find that fitting the learning rate is not necessary to generate good regressors and in some situations, even use of the worst possible parameter settings affects the model-based analysis only marginally. Our results suggest that precise model fitting is not necessary for model-based fMRI, thereby freeing experimental design from the constraint of allowing precise fits. They also highlight the limited use of fMRI data for arbitrating between different (correlated) models or model parameters.

**Keywords:** model-based fMRI, reinforcement learning, prediction errors

## Acknowledgements

This research was funded by NIH grant R01MH098861 and a T32 grant to the Princeton Neuroscience Institute

# 1 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 methods have been developed to overcome this limitation by using computational models of behavior to shed light on latent variables (such as prediction errors) [1]. 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. [2, 3]).

In a typical model-based fMRI analysis, one begins by fitting a model (e.g. Q-learning or SARSA) to the observed behavior, most commonly at the single subject level. This fitting procedure produces a set of best-fitting parameters (e.g. learning rate for each subject) which are then used in the model to generate trial-by-trial measures of various latent variables in the model (e.g. values and prediction errors). These measures can be used to construct regressors for the variables of interest and, by the magic of GLM analysis, find the areas encoding these variables in the brain.

One potential bottleneck in this approach is the model-fitting step. Because of the relatively small number of behavioral trials in fMRI experiments, the accuracy with which the parameters can be estimated is often poor, nor is there any guarantee that the parameter values expressed in behavior will match the parameter values used in the brain. For instance, observed behavior may result from a combination of learning modules, each of which uses a learning rate different from the observed composite rate [4]. Thus a key question is: to what extent are model-based fMRI analyses sensitive to errors in parameter estimation? Do we really have to work so hard to obtain the best possible parameter fits? Here we address this question for the case of simple reinforcement learning tasks. Amazingly we find that the value and prediction error regressors are relatively insensitive to settings of the learning rate in many experimental designs. This theoretical prediction is borne out by analysis of real fMRI data.

## 2 Methods

Our goal is to determine the extent to which a model-based fMRI analysis of a simple reinforcement learning task will be influenced by the choice of learning rate parameter. We focus on the value and prediction error terms and how the regression weights for these components change as a function of the learning rate. In particular, we assume a ‘ground truth’ learning rate  $\alpha_1$  and ask how the regression coefficient is predicted to change as a function of the ‘fit’ learning rate  $\alpha_2$ .

### 2.1 Regression weights as a function of model parameters

Model-based fMRI analysis uses a linear model with (parametrically modulated) regressors for the variables of interest. For example, if we are interested in the neural substrates of prediction errors we construct a model-based regressor for brain activity,  $\mathbf{x}(\alpha_1)$ , whose height at the time of each (hypothetical, model-defined) prediction error is proportional to the size of that error assuming that the learner was using a learning rate (or step size)  $\alpha_1$ . If  $\alpha_1$  is the ‘ground truth’ value of the learning rate, then the generative model for the fMRI data of interest  $\mathbf{Y}$  is

$$\mathbf{Y} = \beta(\alpha_1)\mathbf{x}(\alpha_1) + \epsilon \quad (1)$$

where  $\beta(\alpha_1)$  is the regression coefficient for the prediction error regressor and  $\epsilon$  indicates zero mean noise. Our goal is to estimate the magnitude of the regression coefficient,  $\hat{\beta}(\alpha_2)$ , as a function of an incorrect *fit* learning rate  $\alpha_2$  (which is in general not equal to  $\alpha_1$ ), as a function of the ‘true’ regression coefficient,  $\beta(\alpha_1)$ . If the regression coefficients are estimated using ordinary least squares regression then, on average (for zero mean noise) we have

$$\begin{aligned} \hat{\beta}(\alpha_2) &= (\mathbf{x}(\alpha_2)^T \mathbf{x}(\alpha_2))^{-1} \mathbf{x}(\alpha_2)^T \mathbf{Y} \\ &= (\mathbf{x}(\alpha_2)^T \mathbf{x}(\alpha_2))^{-1} (\mathbf{x}(\alpha_2)^T \beta(\alpha_1) \mathbf{x}(\alpha_1)) \end{aligned} \quad (2)$$

The above equation implies that

$$\frac{\hat{\beta}(\alpha_2)}{\beta(\alpha_1)} \approx \rho(\mathbf{x}(\alpha_1), \mathbf{x}(\alpha_2)) \frac{\sigma(\mathbf{x}(\alpha_1))}{\sigma(\mathbf{x}(\alpha_2))} \quad (3)$$

where  $\rho(\mathbf{x}(\alpha_1), \mathbf{x}(\alpha_2))$  is the correlation between  $\mathbf{x}(\alpha_1)$  and  $\mathbf{x}(\alpha_2)$  and  $\sigma(\mathbf{x}(\alpha_1))$  is the standard deviation of  $\mathbf{x}(\alpha_1)$ . Thus, if the regressors are normalized to have unit variance, the size of the regression coefficients is entirely determined by the correlation coefficient between the regressors at different learning rates.

### 2.2 Computing the correlations

For simplicity we consider the Rescorla-Wagner learning rule [5]. In this case, the learner is assumed to update its estimates of the value of a stimulus,  $V$ , in proportion to the prediction error,  $\delta = r_t - V$ , between this expected value and

the actual reward outcome,  $r_t$ . We denote the value learned with learning rate  $\alpha_1$  as  $V_1$ , and  $V_2$  are the values when the learning rate is  $\alpha_2$ . The update equations for the model with learning rate  $\alpha_i$  are thus

$$\delta_i = r_t - V_i \quad \text{and} \quad V_i \leftarrow V_i + \alpha_i \delta_i \quad (4)$$

Our goal is to compute the correlation coefficients  $\rho(V_1, V_2)$  and  $\rho(\delta_1, \delta_2)$ . Recall that the correlation coefficient is

$$\rho(x, y) = \frac{\text{cov}(x, y)}{\sigma(x)\sigma(y)} = \frac{\langle xy \rangle - \langle x \rangle \langle y \rangle}{\sqrt{(\langle x^2 \rangle - \langle x \rangle^2)(\langle y^2 \rangle - \langle y \rangle^2)}} \quad (5)$$

where  $\langle \cdot \rangle$  denotes taking the average, in our case, over all possible reward sequences  $\{r_1, r_2, \dots, r_t\}$ . Thus, to compute the standard deviation and correlations we need to compute the averages:  $\langle V_i \rangle$ ,  $\langle V_i^2 \rangle$ ,  $\langle V_1 V_2 \rangle$ ,  $\langle \delta_i \rangle$ ,  $\langle \delta_i^2 \rangle$  and  $\langle \delta_1 \delta_2 \rangle$ .

To get a handle on these averages we first write the value at time  $t$  as a sum over all previous rewards

$$V_i = \sum_{a=1}^t \alpha_i (1 - \alpha_i)^{t-a} r_a = \sum_{a=1}^t \kappa_{ia} r_a \quad (6)$$

where we have defined the kernel  $\kappa_{ia} = \alpha_i (1 - \alpha_i)^{t-a}$ . Because this kernel is independent of the reward, we get

$$\langle V_i \rangle = \sum_{a=1}^t \kappa_{ia} \langle r_a \rangle \quad \text{and} \quad \langle V_i^2 \rangle = \sum_{a=1}^t \sum_{b=1}^t \kappa_{ia} \kappa_{ib} \langle r_a r_b \rangle \quad \text{and} \quad \langle V_1 V_2 \rangle = \sum_{a=1}^t \sum_{b=1}^t \kappa_{1a} \kappa_{2b} \langle r_a r_b \rangle \quad (7)$$

where  $\langle r_a \rangle$  is the average reward at time  $a$  and  $\langle r_a r_b \rangle$  is the average of the product of the rewards at times  $a$  and  $b$ . Similarly, since  $\delta_i = r_t - V_i$ , for the prediction errors we can write

$$\begin{aligned} \langle \delta_i \rangle &= \langle r_t \rangle - \langle V_i \rangle = \langle r_t \rangle - \sum_{a=1}^t \kappa_{ia} \langle r_a \rangle \\ \langle \delta_i^2 \rangle &= \langle r_t^2 \rangle - 2 \langle r_t V_i \rangle + \langle V_i^2 \rangle = \langle r_t^2 \rangle - 2 \sum_{a=1}^t \kappa_{ia} \langle r_a r_t \rangle + \sum_{a=1}^t \sum_{b=1}^t \kappa_{ia} \kappa_{ib} \langle r_a r_b \rangle \\ \langle \delta_1 \delta_2 \rangle &= \langle r_t^2 \rangle - \langle r_t V_1 \rangle - \langle r_t V_2 \rangle + \langle V_1 V_2 \rangle = \langle r_t^2 \rangle - \sum_{a=1}^t \kappa_{1a} \langle r_a r_t \rangle - \sum_{b=1}^t \kappa_{2b} \langle r_b r_t \rangle + \sum_{a=1}^t \sum_{b=1}^t \kappa_{1a} \kappa_{2b} \langle r_a r_b \rangle. \end{aligned} \quad (8)$$

Equations 7 and 8 imply that to compute the required correlations we only need the averages over the reward distribution:  $\langle r_a \rangle$  and  $\langle r_a r_b \rangle$ . The exact form of these averages depends crucially on the dynamics of the reward generating-process in the experiment. We now consider the most commonly used cases in which the rewards are sampled from probability distributions that are either constant, drifting or piecewise constant over time.

### 3 Results

In this section we investigate the extent to which changes in the learning rate affect the fMRI signal in three different cases characterized by the temporal dynamics of the reward generating distribution. In particular we consider the cases in which the reward distribution is constant, gradually changing or abruptly changing. We discuss the constant case at length, presenting the analytic results in full as well as comparisons with real experimental data. In the interests of space, the gradual and abruptly changing cases are covered in less detail, although they are similarly amenable to analysis.

#### 3.1 Reward probability is constant

In the constant reward distribution case we have

$$\langle r_a r_b \rangle = \begin{cases} m^2 + \sigma_n^2 & a = b \\ m^2 & a \neq b \end{cases} \quad (9)$$

where  $m$  and  $\sigma_n$  are the mean and standard deviation of the reward distribution. This form allows us to compute the sums in equations 7 and 8 exactly (as sums of geometric series), and leads to the following expressions for the correlations between the values and the prediction errors for different learning rates:

$$\rho(V_1, V_2) = \frac{\sqrt{\alpha_1 \alpha_2 (2 - \alpha_1)(2 - \alpha_2)}}{\alpha_1 + \alpha_2 - \alpha_1 \alpha_2} \quad \text{and} \quad \rho(\delta_1, \delta_2) = \frac{(\alpha_1 + \alpha_2) \sqrt{(2 - \alpha_1)(2 - \alpha_2)}}{2(\alpha_1 + \alpha_2 - \alpha_1 \alpha_2)} \quad (10)$$

Contour plots of these function are shown in figure 1A and B. Strikingly, the correlations for both value and prediction error are relatively insensitive to the mismatch in learning rates. Indeed, for prediction errors, the minimum possible value of the correlation (at  $\alpha_1 = 0$  and  $\alpha_2 = 1$  or vice versa) is  $1/\sqrt{2} \approx 0.7$ . This implies that even with the worst possible parameter settings the prediction error regressor would still be quite good!

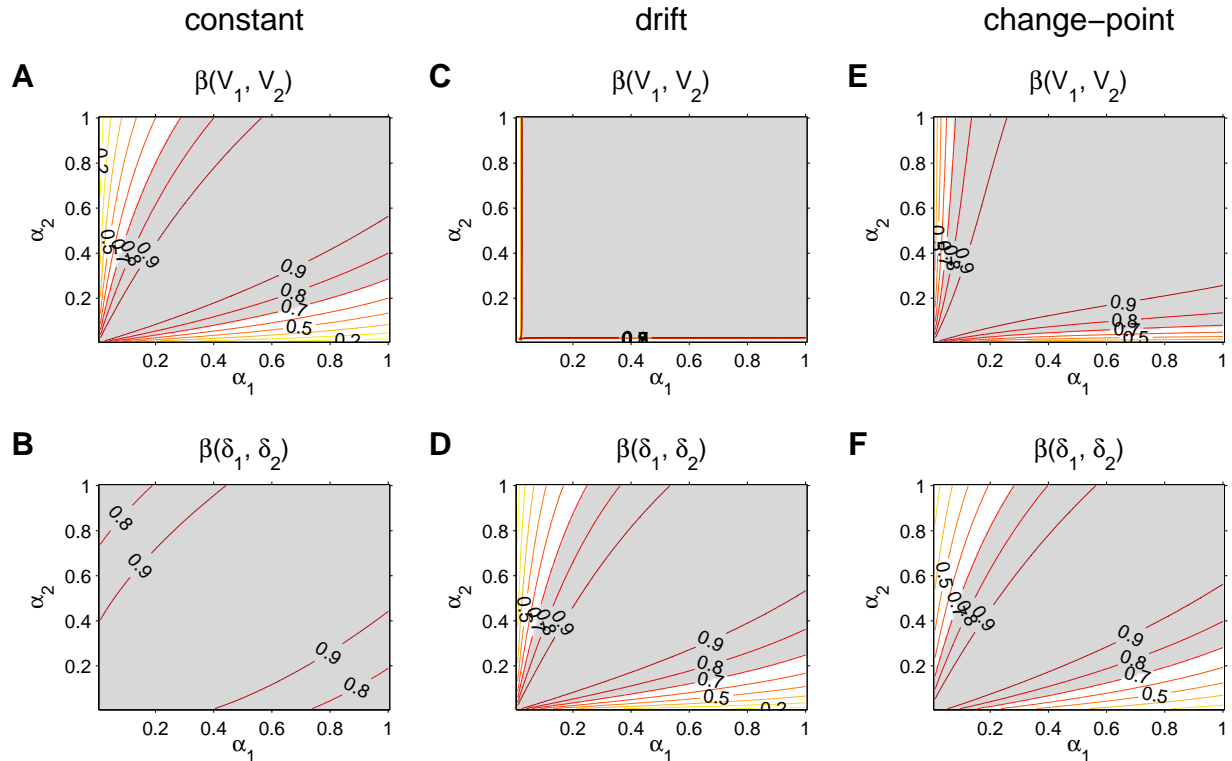


Figure 1: Contour plots showing the correlations for the constant (A, B), drifting (C, D) and change-point cases (E, F) as a function of the two learning rates,  $\alpha_1$  and  $\alpha_2$ . Top row: correlations between values; Bottom row: correlations between prediction errors. The shaded region indicates correlations in excess of 0.7. Note that in all cases there is the large range of parameter space where the variables are strongly correlated.

### 3.2 Analysis of fMRI data for the constant reward probability case

To test this prediction in actual brain data, we used data from [6]. 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: one paid out 20¢ with 100% probability (henceforth ‘sure 20’), one paid out 40¢ (sure 40), two options paid out 0¢ (sure 0) and one paid out either 0 or 40¢ with 50% probability (risky 0/40). 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 sure 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.

According to reinforcement learning theory there are two prediction errors on each trial: one at the time of stimulus onset/choice, equal to the value of the chosen stimulus  $V^{\text{chosen}}$ , and one at the time of reward, given by the difference between reward outcome and the value of the chosen option,  $r - V^{\text{chosen}}$ . Because subjects were trained on the task prior to the scan, we assumed that they knew the values for the constantly rewarding stimuli, and would only continue to update their estimate of value for the 0/40 stimulus for which rewards were probabilistic.

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 (e.g. [6] and references within) putatively due to the strong dopaminergic afferents to that area. We extracted the average BOLD signal from anatomically defined regions of interest in the left and right NAc and regressed this signal vector,  $\bar{Y}$ , 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. We repeated this analysis using different values of the learning rate.

As figure 2 shows, consistent with the theory, at both stimulus onset and reward onset the regression coefficient depends relatively weakly on the learning rate used to generate the regressor. This is true for single subjects (figure 2A,C) and at the group level (figure 2B,D). Also in line with theory, the size of the prediction error regressor varied less (in terms of range) than that of the value regressor, as a function of learning rate (one-sided t-test,  $t(15) = 2.0, p < 0.05$ ).

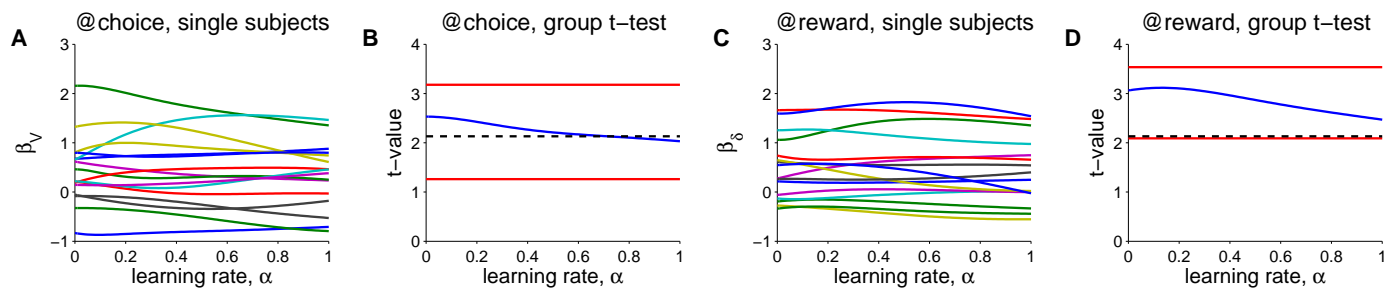


Figure 2: The effect of learning rate on the fMRI signal in the NAc. (A) Regression weights for value of the chosen option at the time of stimulus onset, as a function of learning rate. Each line represents a single subject. (B) Group analysis at the time of stimulus showing the group t-statistic as a function of 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  $\beta$  for each subject. Dashed black line denotes the  $p = 0.05$  threshold. (C) Single subject regression weights for prediction error at the time of reward. (D) Group analysis at the time of reward. In all cases, as predicted, the effect of the learning rate parameter is small.

### 3.3 Dynamic cases: drifting probability of reward and change-points

Our approach can also be applied to cases in which the reward distribution changes gradually (e.g. [3]) or abruptly over time (e.g. [7]). Because the temporal dynamics are more complex, so are the expressions for the correlations, and we omit them here in the interest of space. We note however, that these expressions, in addition to being dependent on the two learning rates, also depend on the parameters of the reward generating distribution such as the noise variance, drift variance and hazard rate at which change-points occur.

Figure 1C–F, shows the correlations in these cases for a set of ‘typical’ parameter values. As with the constant case, there are large regions of parameter space in which both the values and prediction errors corresponding to different learning rates are strongly correlated. In contrast to the constant reward probability case, where the prediction error varied less than the value, in the dynamic cases the values tend to be more correlated than the prediction errors. This is because in the dynamic cases the true value of the stimulus varies more widely and thus there are more prediction errors.

## 4 Discussion

In this paper we considered the extent to which errors in the estimation of model parameters impact model-based fMRI. Surprisingly, we found that for the case of learning rate in a simple reinforcement learning algorithm, the answer was: not very much. Indeed in the constant case, when analyzing prediction error signals the estimated learning rate had close to no effect on the final result either in theory or in the experimental data. Similar results also held for the dynamic cases of drifting and abruptly changing reward distributions. These results are good news for model-based fMRI as they suggest that many analyses will be valid even if the parameters cannot be well estimated from behavior. They also tell a cautionary tale regarding the futility of trying to estimate learning rates from fMRI BOLD data, or distinguishing between the learning rates used by different brain areas. Whether this same result holds for other reinforcement learning parameters of interest, such as discount rates, is left to future work.

## References

- [1] John P O’Doherty, Alan Hampton, and Hackjin Kim. Model-based fMRI and its application to reward learning and decision making. *Ann N Y Acad Sci*, 1104:35–53, May 2007.
- [2] John O’Doherty, Peter Dayan, Johannes Schultz, Ralf Deichmann, Karl Friston, and Raymond J Dolan. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, 304(5669):452–4, Apr 2004.
- [3] Nathaniel D Daw, John P O’Doherty, Peter Dayan, Ben Seymour, and Raymond J Dolan. Cortical substrates for exploratory decisions in humans. *Nature*, 441(7095):876–9, Jun 2006.
- [4] Robert C. Wilson, Matthew R. Nassar, and Joshua I. Gold. A mixture of Delta-rules approximation to Bayesian inference in change-point problems. *PLoS Computational Biology*, accepted.
- [5] R. A. Rescorla and A. R. Wagner. *A theory of Pavlovian conditioning variations in the effectiveness of reinforcement and nonreinforcement*. 1972.
- [6] Yael Niv, Jeffrey A. Edlund, Peter Dayan, and John P. O’Doherty. Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. *The Journal of Neuroscience*, 32(2):551–562, 2012.
- [7] Timothy E. J. Behrens, Mark W. Woolrich, Mark E. Walton, and Matthew F. S. Rushworth. Learning the value of information in an uncertain world. *Nature Neuroscience*, 10(9):1214–1221, 2007.