Overview

Model-based approaches to neuroimaging: combining reinforcement learning theory with fMRI data

Jan P. Gläscher[1] and John P. O'Doherty[1,2]*

The combination of functional magnetic resonance imaging (fMRI) with computational models for a given cognitive process provides a powerful framework for testing hypotheses about the neural computations underlying such processes in the brain. Here, we outline the steps involved in implementing this approach with reference to the application of reinforcement learning (RL) models that can account for human choice behavior during value-based decision making. The model generates internal variables which can be used to construct fMRI predictor variables and regressed against individual subjects' fMRI data. The resulting regression coefficients reflect the strength of the correlation with blood oxygenation level dependent (BOLD) activity and the relevant internal variables from the model. In the second part of this review, we describe human neuroimaging studies that have employed this analysis strategy to identify brain regions involved in the computations mediating reward-related decision making. © 2010 John Wiley & Sons, Ltd. *WIREs Cogn Sci* 2010 1 501–510

In this review, we will describe a recent development in the analysis for functional magnetic resonance imaging (fMRI) data that is aimed at combining such data with quantitative computational models of cognitive function. These mathematical models specify the computational processes required for solving a cognitive task and they define internal variables that instantiate these computations. The goal of model-based fMRI is to establish the neurophysiological validity of these models by correlating neural activity in particular areas of the brain with the model's internal variables. Model-based fMRI, therefore, allows one to begin to address the question of *how* a particular brain region might carry out a particular cognitive operation as opposed to merely identifying where in the brain such a putative operation is carried out, as is often the case in more conventional trial-based fMRI analyses. In the first part of the review, we will provide a description of a typical model-based fMRI data analysis using reinforcement learning (RL) as the computational framework and in the second part we will provide examples of how model-based neuroimaging methods have been applied to the topic of human instrumental reward-learning and decision making.

Here, we will use the computational framework of RL[1] to exemplify the principles of model-based fMRI. However, the analysis strategy outlined below is by no means limited to only the different variants of RL. In fact, any kind of computational model that defines specific internal variables on a trial-by-trial basis (or even at subtrial temporal resolution) can be employed. For instance, another class of learning models involves Bayesian updating which is conceptually very different from RL, in that it is predicated on an ideal observer.[2,3]

The approach described in this article is also distinctly different from connectivity models, which constitute a different class of model-based analyses of fMRI data. The power of effective connective models lies in the characterization of the flow of information across different brain regions and is thus suited to identify brain *networks* involved in solving a task. The computational approach that we describe in this review specifies the putative computations carried out

*Correspondence to: odoherjp@tcd.ie

[1]California Institute of Technology, Division of the Humanities and Social Sciences, Pasadena, CA 91125, USA

[2]Trinity College, Institute of Neuroscience, Dublin, Ireland

DOI: 10.1002/wcs.57

 © 2010 John Wiley & Sons, Ltd.

within a single region. An exciting future prospect lies in the combination of these two approaches[4] to derive a more comprehensive characterization of how the brain solves a particular cognitive task.

MODEL-BASED fMRI

Model-based approaches to fMRI involve at least three crucial steps: (1) the definition or selection of a quantitative computational model for a given cognitive process, (2) the determination of free model parameters e.g., by fitting of this model to the behavior, and (3) the use of the model's internal variables as regressors in the analyses of fMRI data to detect those brain regions exhibiting significant correlations with those signals. We will illustrate each of these steps with an example (assembled in Figures 1 and 2).

Defining a Computational Model

Following David Marr's[5] taxonomy, computational models of cognitive processes can be described at the computational, algorithmic, and implementational levels. Whereas the computational level specifies the goal of a computation, the algorithmic level focuses on how the computational theory can be implemented, i.e., how the input and output are specified and what mathematical operations can be used to transform the former into the latter. Finally, the implementational level specifies how the algorithm can be physically implemented in the underlying neural circuitry. Due to the indirect measurement of neuronal activity via a hemodynamic response and the resulting coarseness of the data, most of the computational models that are used in the fMRI can be characterized at the computational and/or algorithmic level; i.e., these models are concerned with a representation of the mathematical operations used in a cognitive process without paying too much attention as to how these computations are physically implemented and whether that implementation closely matches the actual neuronal architecture.

The approach of defining computational models for various cognitive operations has long been a cornerstone of experimental psychology, particularly in the cognitive tradition.[6] However, it is difficult if not impossible to discriminate between certain classes of models that can potentially provide an equally good characterization of behavioral output. In many cases arguably, behavioral observations alone often do not provide sufficient constraints to allow one to discriminate between different models with different internal variables. By providing a window into ongoing neuronal activity, modern functional brain imaging techniques including fMRI can be used to test for the presence of internal representations corresponding to different models. Thus, it is possible to use the neural data as a means to determine which model provides a better account of the specified cognitive operation. We will now illustrate the procedure of model-based fMRI with reference to a class of models with particular relevance to the area of reward-learning and decision making.

Reinforcement Learning

Reinforcement learning[1] describes a class of models commonly used in learning and decision making. The framework provides computational models of different flavors, but central to all of them is an expected value signal V associated with each stimulus that reflects the current estimate of the future expected rewards. This value signal is updated on each trial by a temporal difference (TD) error δ, originally proposed by Sutton and Barto[7] as a real-time expansion of the prediction error (PE) formulated in the Rescorla–Wagner (RW) model[8] (Figure 1). This TD error is computed as the sum of the rewards obtained at the current time point and the difference between the value signal of the next and the current points in time: $\delta = R_t + \gamma V_{t+1} - V_t$ (γ is the discount factor for future rewards). In essence, as in the RW model, this TD error represents the difference between the actual and expected outcomes at a particular point in time. The value signal is then updated with this TD error, weighted by a learning rate α, which controls the influence of the TD error on the change in value signal ($V_{t+1} = V_t + \alpha\delta$). The learning rate is a free model parameter that can be determined using optimization routines. In order to arrive at a decision between different options in a decision-making task, the value signals of all decision cues are normalized using a sigmoid function (softmax action selection), whose slope is controlled by the softmax temperature τ, thus yielding the choice probabilities by the model for each option at the current point in time. The softmax temperature is the second free model parameter, which regulates the stochasticity of action selection: with a small τ (small slope of the sigmoid), even large value differences between two decision options yield similar choice probabilities, thus making action selection a more or less stochastic process; conversely, with a large τ even small value differences are amplified. Thus, this parameter in essence regulates the sensitivity of the agent's choices to differences in reward-value between the options. The softmax temperature τ can also be fitted using optimization procedures.

 © 2010 John Wiley & Sons, Ltd.

19395086, 2010, 4, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcs.57 by Bcu Lausanne, Wiley Online Library on [09/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

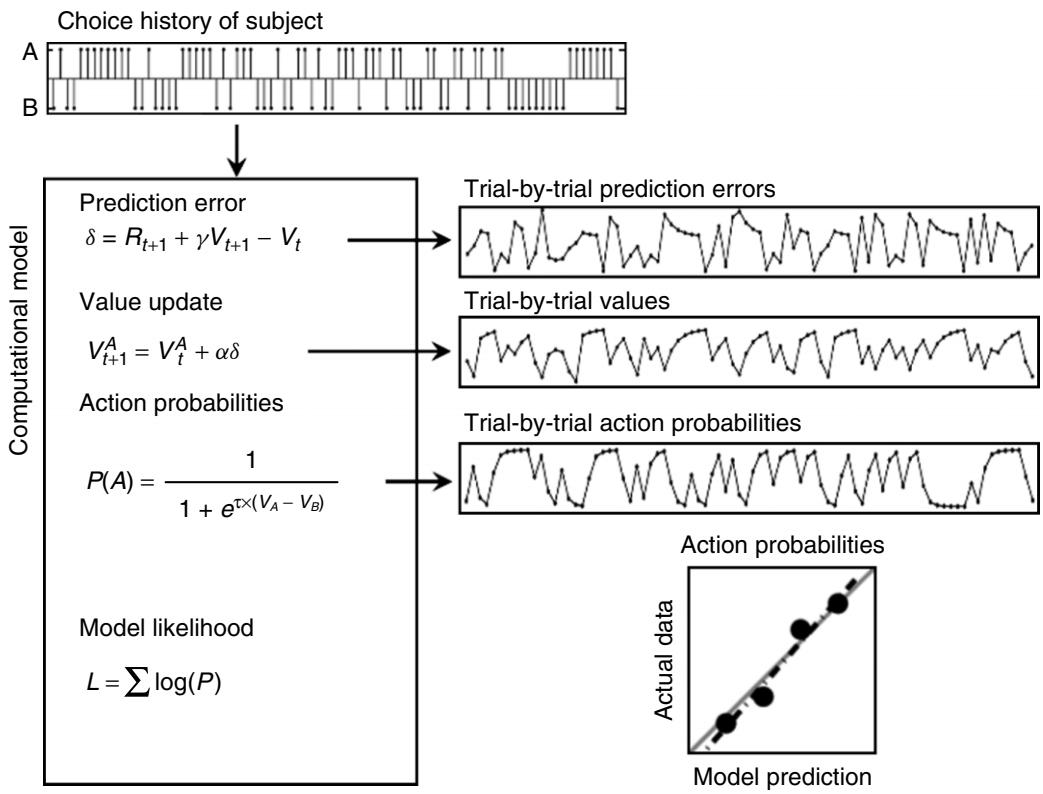

FIGURE 1 | An example of a computational model which can be used in combination with functional magnetic resonance imaging data: reinforcement learning (RL). The goal of this model is to learn about the expected reward attributable to a set of actions in the world (e.g., A and B), and to guide action selection so that the action associated with the highest expected reward is favored. This particular RL model instantiation uses a temporal difference learning rule to learn the value predictions and a softmax rule for action selection. The index variable t denotes within-trial time. The model has five internal variables: the prediction error (PE) δ, and the estimated value predictions for the two actions V_A and V_B, along with the softmax transformed action probabilities P_A and P_B. These variables are plotted in a trial-by-trial resolution but are modeled at different time points within a trial when converted to a predictor in a general linear model (Figure 2). The PE δ (weighted by the learning rate α) regulates the size of the value update on each trial. Softmax action selection is realized by filtering the value difference through a sigmoid function, whose slope is controlled by the inverse temperature τ. This operation converts the values to action probabilities. This parameter represents the stochasticity of the choices, or conversely, the reward sensitivity: if τ is small, even large value differences will result in very similar action probabilities and the model's choices are virtually random. In contrast, if τ is large, even small value differences in the medium value range can be exaggerated, thus leading to different choices. The model likelihood is used as a cost function in an optimization procedure to determine the model parameters α and τ so that model's fit with the individual choice history is maximal. As an initial visual quality check, the model's binned action probabilities for one particular action (e.g., A) can be plotted against the actual choice probabilities (determined, e.g., as percentage of choices for option A) and the increase across these different bins can be examined (lower right panel). Deviations of this linear increase from the $y = x$ line can indicate whether the model is severely over- or underpredicting the actual choices of a subject.

Determining Free Model Parameters

After a computational model of a cognitive process has been selected, the free model parameters have to be determined, a step which is crucial for the subsequent interpretation of the fMRI findings. In principle, there are several ways for choosing concrete values for model parameters.

Firstly, these parameter values can be chosen, such that the predictions of the model provide the best fit to the observable behavioral data. This provides an important link to the fMRI analysis because it ensures that the activation pattern in a particular brain region is instantiating a computation that is behaviorally relevant, which therefore confers psychological validity. In the example of Figure 1, the degree to which the model with a specific set of parameters explains the behavioral data is computed by summing across all trials over the logarithm of action probabilities derived from the model for the action chosen on that trial. During the optimization procedure, the free model parameters α (learning rate) and τ (softmax temperature) are iteratively adjusted to minimize the negative model likelihood, which serves as a cost function. In the RL case, this is equivalent to minimizing

 © 2010 John Wiley & Sons, Ltd.

19395086, 2010, 4, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcs.57 by Bcu Lausanne, Wiley Online Library on [09/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

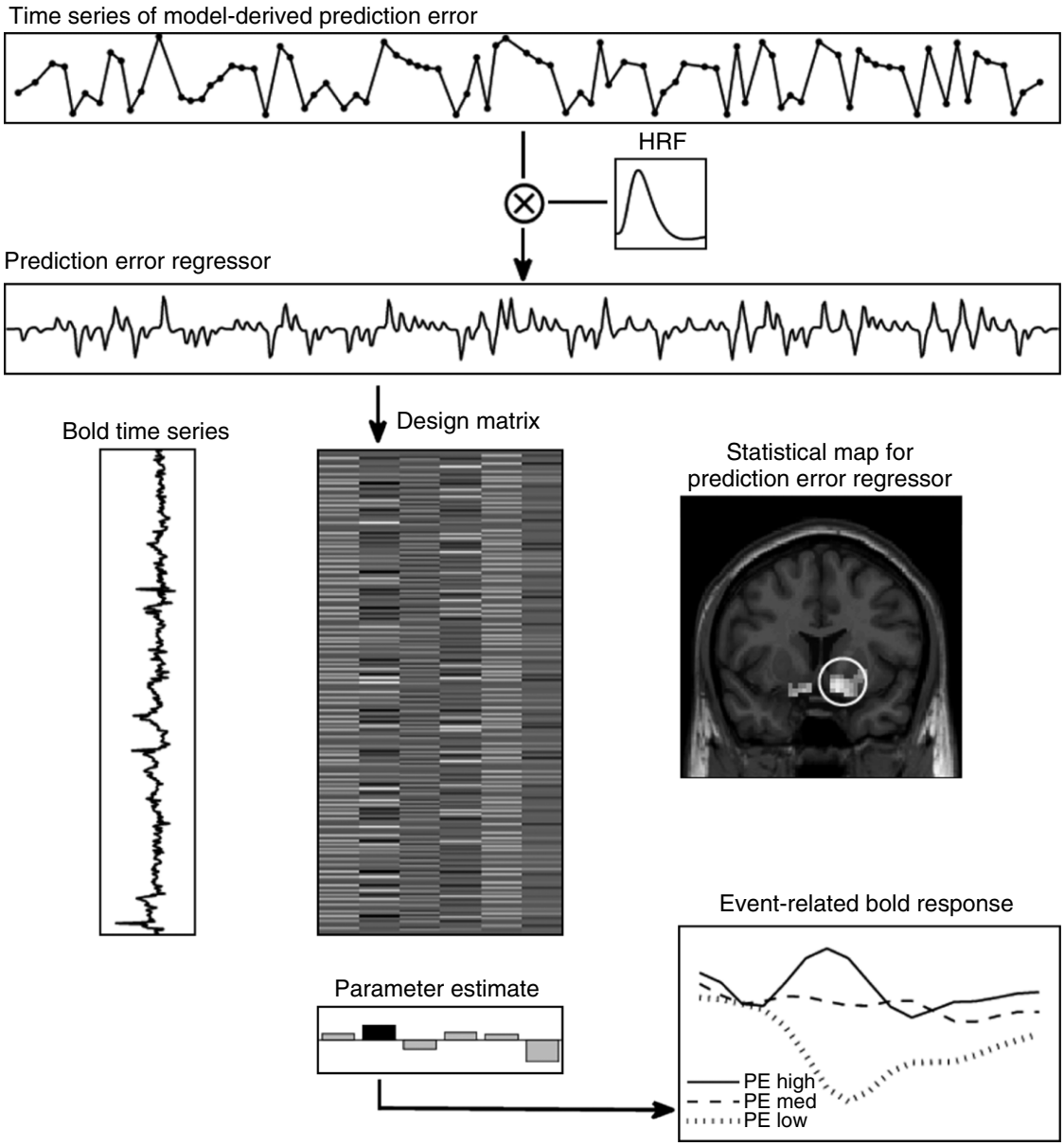

FIGURE 2 | Application of the computational model to functional magnetic resonance imaging (fMRI) data. Internal variables derived from the model [e.g., prediction errors (PEs), modeled at the time of the outcome presentation of each trial] are converted into a time series and convolved with a hemodynamic response function thus yielding a regressor in a single-subject fMRI design matrix. This general linear model is fitted at each voxel in the brain. Subsequent statistical contrasts for the parameter estimates of the newly created regressor yield a statistical map describing the degree of correlation between activity in a particular BOLD time series voxel and the internal variable of interest (in this case the PE). Finally, the goodness of fit of the model-based variable with the time series in a particular brain region can be visualized by plotting the event-related averaged time series for a given trial or event, separated into bins, which capture different levels of the internal variable (here: low, medium, and high PEs).

the difference between the subject's actual choice history and the model-predicted choices embodied in the model's action probabilities. This approach has been successfully employed in a variety of studies.[9–11] A key disadvantage of such an approach is that the limited repertoire of behavioral data available for a given experiment may be insufficient to meaningfully constrain the model fit, particularly for models with a large number of free parameters. Another issue is that if it is the case that there are multiple controllers in the brain which must interact together in order to produce behavior,[12,13] then neural activity related to a single controller would not necessarily be expected to map directly onto observed behavior.

19395086, 2010, 4, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcs.57 by Bcu Lausanne, Wiley Online Library on [09/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

In the absence of meaningful behavioral data, it is tempting to revert to the previously published literature to determine free model parameters. This approach, however, is problematic, because experimental designs are usually not identical and model parameters like learning rates tend to vary across studies.[4,9–11,14–16] Thus, it is highly questionable whether model parameters extracted from previous studies will produce appropriate model fits. Consequently, the psychological validity of the model and the interpretation of the parameters are called into question.

An alternative to using behavioral data is to instead use the goodness of fit between a model-based time series and the BOLD signal in a particular brain region or group of regions for parameter estimation. Although fitting to the BOLD data is potentially much more powerful than behavioral fitting due to the vast increase in the dimensionality of the data source, it does have a number of drawbacks, such as the potential circularity of using the same data for optimizing parameters as is being used for inference, as well as the danger of diminished psychological validity should the resulting model fail to provide a good account of subjects' actual behavior.

It is worth noting that a strategy of parameter estimation which combined across both behavioral and BOLD measures, perhaps also including other data sources such as physiological measures, might potentially overcome many of the disadvantages of each approach alone.

Furthermore, it is possible to use the final model likelihood as a metric of the quality of the model fit in order to compare the performance of different competing models. In the event that these models differ in the number of free parameters, it is necessary to take this into account since a model with more free parameters will necessarily have a better fit. This can be done by means of a Bayesian Information Criterion (BIC) or Akaike Information Criterion (AIC), which adjusts the model likelihood by the number of free parameters (BIC, AIC) and number of trials (BIC), thus controlling model complexity by penalizing an excessive number of free parameters. However, both of these criteria can fail if the model parameters are correlated (i.e., dependent upon one another). In this case, use of the negative free energy[17] is more appropriate as it penalizes the effective rather than absolute number of free parameters in a model. If competing models are nested within another, it is also possible to compare the model fits between them directly with a likelihood ratio test.

It is important to note that merely identifying the best fitting model among a set of competing alternatives (which could mean the same model but with different parameter values) does not necessarily guarantee that the model is capturing all aspects of the data. It is possible that the optimization procedure can capitalize on fitting only certain aspects of the data, while ignoring others. Therefore, although one model might fit better than others, nonetheless the fit between the model output (e.g., choice probabilities) and the actual data for that model (e.g., choice history of the subject) (Figure 1) might still be rather poor. Thus, it is important to check the overall quality of the fit across the full range of states for the model.

Once the optimal model and its parameters have been identified, it is then possible to extract the model's internal variables (e.g., in RL: expected value and PE signals) and then use these signals as predictors in an fMRI analysis.

Using the Model's Internal Variables as fMRI Predictors

The first step in applying the computational model to the fMRI data involves the creation of a regressor which is assigned numerical values generated by a particular internal variable in the model. This regressor has to be associated with particular time points in the experiment, thus creating a model-derived time series. It is important to choose the correct time points at which the internal variable is expected to occur. For instance, in the example described above, a value signal is expected to arise during the presentation of the different decision options, whereas in its simplest form (as in the trial-based RW model), a PE signal might occur specifically at the time of the outcome within the trial. Furthermore, to account for the delay induced by the hemodynamic response, this time series is usually convolved with a canonical hemodynamic response function (Figure 2).

This newly generated regressor can be then included as a predictor variable in a single-subject fMRI design matrix (Figure 2), which is then estimated at each voxel in the brain using standard multiple linear regression techniques. A statistical contrast on the parameter estimate for the model-derived regressors yields a statistical map of those brain regions in which the BOLD response exhibits a significant correlation with the model's internal variables.

MODEL-BASED fMRI IN VALUE-BASED LEARNING AND DECISION MAKING

We will now review some of the recent studies that have employed model-based analyses of functional

 © 2010 John Wiley & Sons, Ltd.

19395086, 2010, 4, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcs.57 by Bcu Lausanne, Wiley Online Library on [09/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

neuroimaging data on the topic of value-based decision making in humans. We will first summarize the findings supporting the instantiation of the core components of RL in the brain. Next, we will discuss additional evidence in support of the possibility that additional mechanisms beyond simple RL are also in operation during certain classes of decision and learning problems.

Neural Correlates of Reinforcement Learning

The model outlined in Figure 1 is an example of a simple RL model, which defines two core variables crucial for decision making: the expected reward V and the PE δ. Building on the observation that dopamine neurons exhibit properties of a temporal PE from single unit recordings in monkeys,[18] a model-based fMRI study was conducted using a simple classical conditioning paradigm with sweet taste rewards, affectively neutral outcome, or no outcome as unconditioned stimuli.[19] Sometimes learned expectations were violated (e.g., delivery of unexpected rewards and omissions of expected rewards) thus incurring a PE. A full TD learning algorithm was fit to subjects' trial history which generated a trial-by-trial TD PE signal that was then subsequently correlated against the fMRI data. Activity in the ventral striatum and some other brain regions such as orbitofrontal cortex was found to correlate with this signal, a finding that has also been shown using more conventional trial-based analysis approaches.[20,21]

The other core component of the RL framework is the expected reward, V. A number of studies have now provided evidence that the medial orbitofrontal cortex (mOFC) is involved in the representation of this signal. For instance, in studies using instrumental choice tasks, it was observed that mOFC was tracking the expected value of the chosen option.[9,11,22]

Within these simple RL models, an important free parameter is the learning rate which sets the rate at which the PE is used to incrementally update the expected reward signal. During instrumental tasks that evoke behavioral choices from the subjects the learning rate can be determined by fitting the model to the choices data. However, in passive Pavlovian tasks, these data are not available and other nonvolitional behavioral measures such as psychophysiological responses (e.g. skin conductance responses) or reaction times are usually noisy and therefore pose a challenge for fitting a computational model. Alternatively, one strategy is to choose biologically meaningful learning rates (e.g., a lower and an upper bound) and compare the correlations of both model variants with the BOLD signals. For instance, O'Doherty et al.[19] chose learning rates of $\alpha = 0.2$ and $\alpha = 0.7$ and found similar PE-related responses in ventral striatum and orbitofrontal cortex (OFC), although the smaller learning rate yielded a better model fit. In a systematic comparison of learning rate across the entire possible value spectrum, Glascher and Buchel[15] found that different brain regions [amygdala and the fusiform face area (FFA)] exhibited different preferential learning rates: whereas the amygdala showed the greatest effect size at small learning rates of $\alpha = 0.05$, the preferential learning rate for the FFA was at $\alpha = 0.95$. In this context, the learning rate regulates how much the PE (difference between the actual outcome and the *currently* expected value) dominates the value update on each trial: for very small learning rates, this influence is dramatically reduced and the current belief about the reward contingencies (summarized in the currently expected value) is largely retained. Stated differently, the information (PE) gathered on the current trial is mostly ignored. Large learning rates on the other hand indicate that the expectancy violation of the current trial dominates the value update and that a cumulative representation of the expected reward across several trials is never built. Consistent with this interpretation, the small learning rates in the amygdala may be related to the implicit nature of their task, a differential Pavlovian conditioning paradigm with gradually changing reinforcement contingencies, which subjects did not recognize explicitly. Similarly, small learning rates have also been reported in other implicit associative learning tasks.[4] The high learning rates found in the FFA suggest that it serves as a change detector instead of building a consistent representation across several trials and that learning (in the sense of an integration of recent experience across several trials) does not occur. This is to be expected from higher order perceptual association areas, which are merely concerned with building a perceptual representation of the sensory world.

Finally, in a decision-making paradigm that involved phases of both stable and unstable reward contingencies, Behrens et al.[3] found that the learning rate was influenced by the volatility of the environment suggesting that during times of unstable contingencies the learning is enhanced to accelerate the acquisition of the new expected values. Although the behavioral data of this experiment were analyzed using an RL model, the analysis of the imaging data relied on an ideal Bayesian observer. This model represents probability distributions of model parameters instead of the trial-based scalars for expected values and

 © 2010 John Wiley & Sons, Ltd.

19395086, 2010, 4, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcs.57 by Bcu Lausanne, Wiley Online Library on [09/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

PEs in RL. The parameters in this Bayesian model, which estimate different aspects of the environment, were organized in a hierarchical fashion: the observed outcome is influenced by the reward probability, which in turn is governed by the volatility of reward contingencies. Finally, the volatility is governed by a control parameter that estimates the change rate in the volatility and adjusts the learning rate. This latter parameter was correlated with the BOLD signal in the anterior cingulate cortex.

In conclusion, an accumulating number of neuroimaging studies which have employed simple RL for characterizing hemodynamic responses during learning and decision making provide evidence that distinct brain regions (ventral striatum, OFC, and amygdala) exhibit response patterns consistent with expected reward and PE signals derived from these models. Next, we consider situations under which additional processes beyond that captured by simple RL are likely to be engaged.

Extensions of Reinforcement Learning

In simple RL, only the value of the chosen option is updated on every trial, whereas the values of the other options remain stationary until they are chosen again. While the PEs in different RL variants capitalize on the different possible action values, recent neurophysiological evidence suggests that dopaminergic neurons may compute a PE by comparing the actual outcome against the value of the chosen option.[23,24] However, it is known that decision-making behavior in humans can also be influenced by contextual information such as the amount of reward available on average in a given scenario, or counterfactual information about what could have been won had a different action been chosen. For instance, if a subject wins $10 on a particular trial, but is also aware that (s)he could have won $100 if a different option had been chosen, then the experienced utility of that $10 is different than had the $100 alternative not been available.[25,26] Evidence that neural signals of reward reflect the processing of counterfactual information has come from a study by Coricelli et al.[27] wherein subjects had to choose between two risky gambles. Following the choice, subjects were shown the outcome of the chosen option gamble, but also, in some cases, the outcome that would have been attained had they chosen the alternative. In cases when the counterfactual outcome exceeds the actual outcome, a state defined as 'regret' ensues, which was found to be correlated with activity in mOFC. Underlying the learning of such 'regret' signals is the notion of a fictitious PE signal that compares the magnitude of the unchosen outcome against the actual outcome. In a study that allowed subjects to bid on the outcome of virtual financial markets, Lohrenz et al.[28] found that both the actual and the fictitious PE signals correlated with the BOLD signal in the ventral striatum, but that the fictitious error also extended into dorsal striatum, suggesting that this fictitious error signal may be at least partly neurally distinct from the reward PE signal exhibited by simple RL.

Another situation, under which simple RL models can fail to adequately capture the mechanisms used by the brain to guide decisions, is when a decision problem contains hidden structure. An example of such structure is the anticorrelation between the reward probabilities assigned to decision options in tasks such as probabilistic reversal learning. On such a task, at any one time one particular option out of a choice pair is highly rewarding while the other is not, and the contingencies assigned to these options periodically reverse. A simple RL model that fails to take into account such structure would perform suboptimally, as both options would be treated as being independent and would be learnt about separately. On the other hand, a model which exploits this known task structure would incorporate the knowledge that the contingencies are anticorrelated and are subject to reversal and use this knowledge when computing expected rewards. Hampton et al.[10] compared a Hidden Markov model, which incorporated the structure of the reversal task against a simple RL model that did not incorporate such structure. Neural responses in ventromedial prefrontal cortex (vmPFC) were better explained by the model that incorporated the task structure than by a model which did not do so.

Another scenario under which additional computational mechanisms other than RL are likely to be involved is when humans must make decisions in competitive social interactions, whereby in order to choose optimally it is necessary to take into account the behavior of an adaptive opponent, a form of thinking sometimes referred to as 'strategizing.' Hampton et al.[29] used an economic game called the work/shirk task which is 'zero-sum' in that on any trial only one out of the two opponents can win, while the other must lose depending on the choices each makes. Thus in order to perform well on the task, it is beneficial for a player to be able to predict what action her opponent will take next, so that she can choose the option most likely to thwart that opponent. Hampton et al. scanned a group of subjects playing this task in real-time against opponents outside the scanner. Using again the likelihood as a metric to determine

 © 2010 John Wiley & Sons, Ltd.

19395086, 2010, 4, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcs.57 by Bcu Lausanne, Wiley Online Library on [09/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the best fitting model, the authors compared three different computational algorithms for their capacity to account for subjects' behavioral choices and the pattern of neural activity during performance on the task. The first model was simple RL, in which the model learns to take the action that gave the most reward in the past. This is not necessarily a good strategy for success on such a game, because this policy could be easily be exploited by an adaptive opponent. A slightly more sophisticated strategy is to keep track of the actions of the opponent and use an estimate of the action favored most often by the opponent in the recent past to inform choice of the action to be taken on the current trial. Finally, an even more sophisticated strategy in this case called the 'influence model' is to not only keep track of the opponent's actions but also to estimate the opponent's likely predictions about your own actions in order to pick the best response to their best response. In essence, this latter strategy could be summarized as 'thinking about you thinking about me,' and could be considered to be a computational analog for an elementary form of 'mentalizing.'[30] Not only was the influence model a better predictor of subjects' actual behavior on the task than the other models, but also activity in vmPFC and more dorsal segments of medial prefrontal cortex was better accounted for by the influence model than by the other models. Furthermore, a form of PE signal used by the influence model to update the predictions of the opponent's response to the player's strategy was found to correlate with activity in superior temporal sulcus (STS), an area previously identified as being involved in theory of mind or mentalizing, suggesting that this elementary computational model, which can be viewed as an extension of RL, is capturing some of the underlying computations likely being implemented in mentalizing brain regions. In a study also looking at learning in the social domain, Behrens et al.[2] explored the neural mechanisms when learning reward predictions in a social context whereby a subject can obtain advice from a confederate on a trial-by-trial basis. While a reward PE was found in the ventral striatum, a 'social' PE related to determining the fidelity of the opponent was found in the temporoparietal junction and STS as well as in the medial frontal cortex. Once again, this study points to the presence of distinct types of computational signals in the brain during decision making, some of which are accounted for by additional computational mechanism other than simple RL.

CONCLUSIONS AND PERSPECTIVE

In this paper, we have outlined the approach of model-based fMRI in which one aims to characterize BOLD responses in particular brain regions in terms of the internal variables present in particular computational models of neural function. Using RL as an example, we have outlined the typical procedure used to conduct a model-based analysis of fMRI data. We have shown with reference to the field of value-based learning and decision making how it is possible to start with a basic model of a particular cognitive function such as simple RL, and then study situations under which these models either succeed or fail to capture behavior and neural activity in specific brain regions. It has been shown that simple extensions of this RL model such as the introduction of additional updating signals can capture a wide range of complex decision making phenomena including decisions made under situations with counterfactual information and/or hidden structure, and decisions made in competitive social contexts.

It should be noted that an important limitation of the model-based approach described here is that this technique is regionally specific, in that each voxel in the brain is separately analyzed with respect to the same model-based time series. However, no brain region contributes to a given computational process in isolation, but rather such computations likely emerge as a result of dynamic interactions and information flow between regions. An important development in fMRI analysis methods over the past decade has been the emergence of sophisticated analysis tools for capturing dynamic interactions between brain regions in support of particular cognitive processes, such as dynamic causal modeling.[31] Recently, such connectivity methods have been integrated with the computational model-based approach, such that the internal variable from a computational model can be used to modulate connection strengths between brain regions in a dynamic causal model.[4,32] This is a promising new avenue that will allow the incorporation of theoretical models into the characterization of information flow between multiple brain areas.

Overall, although the approach of model-based imaging is still in its infancy, this approach offers considerable promise as a means of advancing our understanding of the neural computations underlying a diverse array of cognitive phenomena ranging from Pavlovian conditioning all the way to theory of mind.

 © 2010 John Wiley & Sons, Ltd.

19395086, 2010, 4, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcs.57 by Bcu Lausanne, Wiley Online Library on [09/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

REFERENCES

1. Sutton RS, Barto AG. *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press; 1998.
2. Behrens TE, Hunt LT, Woolrich MW, Rushworth MF. Associative learning of social value. *Nature* 2008, 456:245–249.
3. Behrens TE, Woolrich MW, Walton ME, Rushworth MF. Learning the value of information in an uncertain world. *Nat Neurosci* 2007, 10:1214–1221.
4. den Ouden HE, Friston KJ, Daw ND, McIntosh AR, Stephan KE. A dual role for prediction error in associative learning. *Cereb Cortex* 2009, 19:1175–1185.
5. Marr D. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York: Freeman; 1982.
6. Simon HA. *Models of Thought*. New Haven: Yale University Press; 1979.
7. Sutton RS, Barto AG. Time-derivative models of Pavlovian conditioning. In: Gabriel M, Moore J, eds. *Learning and Computational Neuroscience: Foundations of Adaptive Networks*. Cambridge, Mass: MIT Press; 1990.
8. Rescorla RA, Wagner AR. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and non reinforcement. In: Black AH, Prokasy WF, eds. *Classical Conditioning II*. New York: Appleton-Century-Croft; 1972, 64–99.
9. Kim H, Shimojo S, O'Doherty JP. Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. *PLoS Biol* 2006, 4:e233.
10. Hampton AN, Bossaerts P, O'Doherty JP. The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *J Neurosci* 2006, 26:8360–8367.
11. Glascher J, Hampton AN, O'Doherty JP. Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. *Cereb Cortex* 2009, 19:483–495.
12. Daw ND, Niv Y, Dayan P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat Neurosci* 2005, 8:1704–1711.
13. Balleine BW, Dickinson A. Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* 1998, 37:407–419.
14. Wittmann BC, Daw ND, Seymour B, Dolan RJ. Striatal activity underlies novelty-based choice in humans. *Neuron* 2008, 58:967–973.
15. Glascher J, Buchel C. Formal learning theory dissociates brain regions with different temporal integration. *Neuron* 2005, 47:295–306.
16. Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ. Cortical substrates for exploratory decisions in humans. *Nature* 2006, 441:876–879.
17. Friston KJ, Stephan KE. Free-energy and the brain. *Synthese* 2007, 159:417–458.
18. Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. *Science* 1997, 275:1593–1599.
19. O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ. Temporal difference models and reward-related learning in the human brain. *Neuron* 2003, 38:329–337.
20. Pagnoni G, Zink CF, Montague PR, Berns GS. Activity in human ventral striatum locked to errors of reward prediction. *Nat Neurosci* 2002, 5:97–98.
21. McClure SM, Berns GS, Montague PR. Temporal prediction errors in a passive learning task activate human striatum. *Neuron* 2003, 38:339–346.
22. O'Doherty J, Critchley H, Deichmann R, Dolan RJ. Dissociating valence of outcome from behavioral control in human orbital and ventral prefrontal cortices. *J Neurosci* 2003, 23:7931–7939.
23. Niv Y, Daw ND, Dayan P. Choice values. *Nat Neurosci* 2006, 9:987–988.
24. Morris G, Nevet A, Arkadir D, Vaadia E, Bergman H. Midbrain dopamine neurons encode decisions for future action. *Nat Neurosci* 2006, 9:1057–1063.
25. Loomes G, Sugden R. Regret theory: an alternative theory of rational choice under uncertainty. *Econ J* 1982, 92:805–824.
26. Bell DE. Regret in decision making under uncertainty. *Oper Res* 1982, 30:961–981.
27. Coricelli G, Critchley HD, Joffily M, O'Doherty JP, Sirigu A, et al. Regret and its avoidance: a neuroimaging study of choice behavior. *Nat Neurosci* 2005, 8:1255–1262.
28. Lohrenz T, McCabe K, Camerer CF, Montague PR. Neural signature of fictive learning signals in a sequential investment task. *Proc Natl Acad Sci U S A* 2007, 104:9493–9498.
29. Hampton AN, Bossaerts P, O'Doherty JP. Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proc Natl Acad Sci U S A* 2008, 105:6741–6746.
30. Amodio DM, Frith CD. Meeting of minds: the medial frontal cortex and social cognition. *Nat Rev Neurosci* 2006, 7:268–277.
31. Friston KJ, Harrison L, Penny W. Dynamic causal modelling. *Neuroimage* 2003, 19:1273–1302.
32. Stephan KE, Kaspar L, Harrison LM, Daunizeau J, den Ouden HE, et al. Nonlinear dynamic causal models for fMRI. *Neuroimage* 2008, 42:649–662.

 © 2010 John Wiley & Sons, Ltd.

19395086, 2010, 4, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcs.57 by Bcu Lausanne, Wiley Online Library on [09/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

FURTHER READING

O'Doherty JP, Hampton AN, Kim H. Model-based fMRI and its application to reward and decision-making. *Ann N Y Acad Sci* 2006, 1104: 35–53.

Daw N. Trial-based data analysis using computational models. In: Phelps EA, Robbins TW, Delgado M, eds. *Affect, Learning and Decision Making, Attention and Performance XXII*. Oxford University Press. In press.

 © 2010 John Wiley & Sons, Ltd.

19395086, 2010, 4, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcs.57 by Bcu Lausanne, Wiley Online Library on [09/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License