

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES

Issue: *Critical Contributions of the Orbitofrontal Cortex to Behavior***Contributions of the ventromedial prefrontal cortex to goal-directed action selection**

John P. O'Doherty

Computation and Neural Systems, California Institute of Technology, Pasadena, California

Address for correspondence: John P. O'Doherty, Computation and Neural Systems, California Institute of Technology, 1200 E. California Blvd, Pasadena, CA 91125. jdoherly@caltech.edu

In this article, it will be argued that one of the key contributions of the ventromedial prefrontal cortex (vmPFC) to goal-directed action selection lies both in retrieving the value of goals that are the putative outcomes of the decision process and in establishing a relative preference ranking for these goals by taking into account the value of each of the different goals under consideration in a given decision-making scenario. These goal-value signals are then suggested to be used as an input into the on-line computation of action values mediated by brain regions outside of the vmPFC, such as parts of the parietal cortex, supplementary motor cortex, and dorsal striatum. Collectively, these areas can be considered to be constituent elements of a multistage decision process whereby the values of different goals must first be represented and ranked before the value of different courses of action available for the pursuit of those goals can be computed.

Keywords: decision making; reward; human

Introduction

Imagine you are faced with the challenge of obtaining lunch and that, in your town, there are three restaurants: one offers Chinese food, another offers Italian food, and the third offers French food. Solving this problem requires several sequential computations. The first of these involves encoding a value for each of the possible goal states (in this case, the prospect of eating lunch in each of these restaurants) and of computing a relative preference ranking for these outcomes. Once determined, the next step in the process is to work out what actions to take to attain each of the different goals, taking into account the probability that those actions will be successful in attaining the goal, and the amount of effort required to complete them: for example, should you walk or drive? how far is it to each restaurant, and what are the chances of being stuck in traffic? Finally, one must compare and contrast between the different available decision options integrating across the value of each goal and the action contingencies and costs required to obtain each goal. In this article, it will be argued that the ventromedial prefrontal cor-

tex (vmPFC) is part of a network of brain areas that in concert with other brain regions enables a solution to this type of decision problem. The vmPFC as defined here refers to the medial orbitofrontal cortex (OFC; area 11 and 14) and adjacent ventral medial cortex (part of area 10); it does not include the central (area 13) or lateral orbitofrontal cortices (areas 12/47). It will be proposed that one of the main functions for the vmPFC in this decision-making process is the encoding and retrieval of the value of available goal states, taking into account the current motivational state of the organism, and in computing the relative or comparative value of these goals within a given decision context. Such goal values can be retrieved by associative links between particular actions and associated outcomes or can be retrieved by discriminative stimuli that signal the availability of individual outcomes. The contribution of vmPFC to the decision process can be contrasted with that of other areas such as the anterior dorsal striatum, supplementary motor cortex, and lateral intraparietal (LIP) sulcus, which instead contribute toward the process of assigning values to actions available to pursue particular goal states.

Together, these brain areas can be viewed as participating nodes in a multicomponent decision process for goal-directed behavior, enabling relative preferences to be calculated over available goals in one part of the system and for an action or set of actions to be selected to pursue those goals in another part.

The role of vmPFC in representing goal values

There is evidence to suggest that vmPFC is involved in representing the value of goal states at several stages in the decision process. First, this area is involved in encoding the value of goals as they are actually attained and experienced. Second, there is some emerging evidence to suggest that this region may be involved in encoding the relative values of different available goals, which could yield a preference ranking over goal states. Finally, this region is involved in encoding the value of chosen options. We will now describe each of these signals in turn.

Value signals for experienced goals (outcome values)

There is now considerable evidence to suggest that vmPFC (particularly the medial orbital surface) is involved in encoding the value of outcomes as they are attained and experienced by an animal. Such signals were first reported in neurophysiological recording studies from the OFC of nonhuman primates, where OFC neurons were found to respond to the taste or odor of foods presented to the monkey with the additional property that such neurons exhibited decreased responding when the animal was satiated compared to when it was hungry, suggesting that those neurons were modulated as a function of the motivational significance of the stimuli to the animal.^{1,2} Similar findings have been obtained in humans for odor and gustatory stimuli. O'Doherty *et al.*³ scanned hungry human participants while they were presented with two food-related odors. Subjects were subsequently removed from the scanner, fed one of the corresponding foods to satiety so that the pleasantness of that food and its associated odor decreased relative to the pleasantness of the food not eaten, and then placed back into the scanner. Activity was found in the OFC in response to both food odors, and moreover, responses to the odor of the food eaten appeared to show a selective decrease after feeding relative to the odor of the food

not eaten. These results, therefore, suggest a role for human OFC in encoding the reward value of the odor stimuli. Other studies have reported correlations between subjective reports of odor pleasantness for a variety of different odors and activity in human OFC.^{4,5} Small *et al.*⁶ also found differential activity in the OFC in response to gustatory stimuli of different subjective value while⁷ reporting modulation in this region as a function of changes in pleasantness for the taste of water in thirsty subjects who were subsequently satiated. Similar differential responses in the OFC have been reported for somatosensory stimuli.⁸ Blood *et al.*⁹ presented musical sequences while varying the degree of harmonization between the tones within a sequence, thereby eliciting increased levels of perceived pleasantness as harmonization was increased. Activity in the medial OFC in response to the musical tones was found to increase the more pleasant the sequence from what it was reported to be, implicating this region in encoding the hedonic value of stimuli in the auditory domain. This region has also been found to contribute to hedonic evaluation for visual stimuli: O'Doherty *et al.*¹⁰ scanned subjects with functional magnetic resonance imaging (fMRI) while presenting face stimuli that varied in physical attractiveness. A significant correlation was found between the degree of rated attractiveness and the level of activity in the medial OFC (see also Refs. 11 and 12). Similar results have been found in OFC for visual stimuli other than attractive faces, such as pictorial scenes.¹³ Further evidence of a role for this area in outcome value coding has come from a number of studies reporting activity in the OFC to abstract rewards not tied to a particular sensory modality such as receiving monetary gains^{14,15} or social praise or feedback.¹⁶ Collectively, these findings implicate the OFC, particularly its medial aspects, in representing the positive hedonic value of outcomes. It is important to emphasize that outcome value representations in the OFC are not invariant. Not only can such representations be modulated as a function of changes in internal motivation state as mentioned earlier, but value-related activity in this region can also be influenced by cognitive factors such as the provision of price information¹⁷ or merely the use of semantic information or even brand labels.^{18,19} Thus, the on-line computation of outcome value in the OFC is highly flexible and can be directly influenced by a variety of internal and external factors.

Evidence for prechoice goal-value signals

We will next consider signals that are present at the time of decision making pertaining to the value of available goals. Returning to our food example, if making a decision between Chinese, Italian, or French cuisine, the value of each of those foods will need to be separately encoded so that they can be entered into the decision process. Thus, to find evidence for prechoice signals representing goal values, we would need to find evidence for unique and separable signals pertaining to the value of each available goal option in a given decision-making context. In a single-unit recording study by Camillo Pado-Schioppa and Assad,²⁰ at the time of decision making, separate yet spatially intermingled populations of neurons in the central OFC were found to be correlated with the value of each of the particular decision options or “goods” available to the monkey on that trial. Given these findings and the likelihood that in humans neurons encoding individual goals are likely also to be spatially intermingled, it is especially challenging to obtain evidence supporting the existence of such signals with the spatial resolution available in human neuroimaging studies. Nevertheless, there is some evidence suggestive of the existence of such signals within vmPFC. If distinct populations of neurons in a particular brain area are involved in encoding the value of individual available goals, then at the level of the BOLD (blood oxygen level-dependent) fMRI signal, such a region would elicit a signal approximating the sum of those individual goal values under the assumption that these signals aggregate in an approximately linear fashion. Thus, a candidate area for encoding individual prechoice goal values would be an area correlating with the sum of the values of available outcomes. Wunderlich *et al.*²¹ reported just such an area in an anterior part of the vmPFC above the orbital surface; this area correlated with the sum of the values for the decision options presented on a given trial as denoted by discriminative stimuli. However, it should be noted that finding such an area representing the sum of values is not an unequivocal demonstration of individual goal-value signals, as naturally this area could simply encode the sum of values *per se* instead of the overlapping values of individual goals. A better approach to address this question in future would be to use multivariate analysis techniques to detect distributed patterns of voxel activity unique to individual goal states or al-

ternatively to temporally separate the presentation of goal options at a level discriminable by the BOLD signal. Furthermore, if goal values are encoded on a relative scale (see the next section), then such signals would not be detectable as variation in the average across trials because by definition the average activity would always be the same on each trial (because of scaling).

Other evidence consistent with the existence of goal values in vmPFC was provided in a series of studies using a “willingness to pay” paradigm by Plassmann *et al.*^{22,23} In this paradigm, hungry human subjects are scanned while being presented with pictures denoting a variety of foodstuffs while indicating their “willingness-to-pay” (WTP) for each of the food items, out of an initial endowment of four dollars available for each item. After the experiment was over, one of the trials is selected at random and if the reported WTP exceeds a random draw from a lottery then subjects are provided with the good and invited to consume it (and their endowment is drawn on); otherwise, they keep the endowment and do not receive the good. This procedure is designed to ensure that the subjects give their true underlying valuation for each of the items. In several studies, activity in a region of vmPFC was found to be correlated with trial-by-trial variations in WTP, suggesting a role for this region in encoding the goal-value of the potential outcomes. A follow-up study used a similar procedure to measure goal-value responses to a variety of different classes of goods, including monetary gambles, trinkets, and food items. An overlapping region of vmPFC (just above the orbital surface) was found to be correlated with the value of all three classes of items, suggesting that goal-value coding for many different categories of goods may all converge within the same region of vmPFC. Such a region would, therefore, be an excellent candidate for mediating coding of the utilities assigned to diverse types of goal stimuli.²⁴

vmPFC computes the relative value of available goals

There is also now emerging evidence that this region may contribute to generating a preference ranking over available goals. In a neurophysiology study by Tremblay and Schultz,²⁵ recording from a part of central OFC, neurons were found to exhibit a relative value code, responding more for a stimulus indicating the subsequent delivery of a particular

juice reward if that juice was the most preferred juice over the set of available juice stimuli in a block of trials, whereas responding less to the same stimulus predicting the same juice when presented in blocks of trials where that juice was the least preferred outcome (but see also Ref. 26 that reported a population of OFC neurons that did not exhibit such relative preference coding). In an fMRI study,²⁷ stimuli were presented that predicted the delivery of a subsequent monetary reward. Neural responses in vmPFC to a stimulus that predicted a specific monetary amount were different depending on whether that stimulus was presented in a context where the monetary amount was the maximum available reward over a block of trials when compared to a context where the same monetary amount was the smallest amount available over a block of trials. A recent study by Lim *et al.*²⁸ also showed evidence for a relative value code in this region but, in addition, suggested that visual attention might modulate such a signal. Participants viewed two food stimuli over which they had to render a choice. However, on each trial they were asked to attend specifically to one of the two food-items by means of a visual fixation. Activity in vmPFC was found to correlate with the subjective value of the attended item relative to the value of the unattended item, providing further evidence that this region is involved in computing a relative value code. Moreover, these findings suggest that visual attention might be a device by which certain goals are highlighted over others within this structure during the evolving temporal dynamics of the decision process.

Value of chosen options

Another signal present in the vmPFC is the chosen value, which corresponds to the expected value of the decision option that is ultimately selected. Chosen values are in common with outcome values, a postdecision selection signal, in that they are a consequence as opposed to a precursor or input into the decision process. A number of studies have revealed these signals to be present while human subjects perform simple choice tasks.^{21,29–32} Furthermore, such signals have also been reported in the activity of single neurons within the monkey OFC.²⁰

Given that the chosen-value signals are postchoice signals and are thus not likely to play a causal role in the decision process itself, what could their function be? One possibility is that these signals could be used

alongside the value of the experienced outcomes to generate prediction error signals by comparing the value of the experienced outcome with the expected value of the chosen option. Such prediction errors have been found to be carried by the phasic activity of dopamine neurons and they have been found in a number of areas in human fMRI studies, most prominently in the striatum. The most widely hypothesized function for such signals is to underpin learning and updating of stimulus- and action-value signals in the striatum and perhaps elsewhere. However, the extent to which such reward prediction error signals would be required within the goal-directed system (as opposed to systems involved in habitual and Pavlovian learning; see the section on multiple systems given later) is controversial, particularly within the context of computational theories of goal-directed learning that use “model-based” forward planning.³³ Within such a framework, value computations for actions are computed on-line and thus do not require a reward prediction error for trial and error learning (although other kinds of prediction errors may contribute to learning other features such as the state-space; see Ref. 34). However, a signal reporting the degree of discrepancy between one’s predictions and actual experience could also serve other functions even within a model-based system, such as by acting as a “model-based” critic indicating how well one’s goal-directed action system is performing—large prediction errors within this system might serve an alerting function, in essence acting as a wake up call suggesting the need to urgently reevaluate the system’s predictions. In addition, such a signal might directly influence behavior: a large positive prediction error would indicate that the system should focus on selecting the immediately preceding action, whereas a large negative prediction error would indicate the need to try something else.

Value signals in vmPFC: retrieved by actions or stimuli?

The finding of goal-value signals in the vmPFC raises the question as to what associative mechanism underlies the retrieval of such signals. Under situations where different actions are associated with different possible goal states, one way in which the value of the goal state could be retrieved is through learned action–outcome associations: consideration of the action would therefore result in the associative

retrieval of the value of the attendant goal outcome. Yet another mechanism by which goal values could be retrieved is through associations formed between discriminative stimuli that indicate the availability of particular decision options or goals and the identity of the goal outcomes. In a typical decision-making task, both possible associative mechanisms are available in which participants usually need to select from several physical actions to make a choice; furthermore, these decision options are usually signaled by means of discriminative stimuli (e.g., different fractal patterns, different slot machines, or different colors). To discriminate between these possibilities, Glascher *et al.*²⁹ compared and contrasted the neural representation of goal-value signals in vmPFC in one of the two decision-making tasks. In one stimulus-based task, the available decision options were signaled by means of two discriminative stimuli (fractals), which in turn were associated with performance of particular actions to obtain those outcomes. In contrast, in an action-based task, no unique discriminative stimulus was presented to indicate specific actions; instead participants had to render choices over two unique actions (rotating a tracker ball vs. pressing a button). In both cases, an overlapping region of vmPFC was found to be correlated with the value of the decision option that was ultimately chosen. These findings suggest that the value of the underlying outcomes could be retrieved by associations learned through pairings with different physical actions, and that such goal-value signals, therefore, are not necessarily elicited exclusively by means of discriminative stimuli.

Further evidence that chosen-value representations can be elicited by specific actions came from Wunderlich *et al.*³² This study also involved an action-based choice task, except in this case on each trial subjects needed to choose between two different physical action modalities: making an eye movement (a saccade from a central fixation point to a target in the right of the visual field) versus making a hand movement (using a right-handed button press). The probability of winning on each action was varied over time using a random walk, and subjects had to keep sampling the two actions to work out which movement was generating the greatest probability of reward so that this movement could be exploited until such time as the contingencies had changed again. Once again, Wunderlich *et al.*, found signals correlating with the value of the chosen ac-

tion in vmPFC, again consistent with the suggestion that these signals could be driven by action representations as opposed to purely being stimulus-bound. However, even more critically, Wunderlich *et al.* observed a topographical arrangement of choice values within vmPFC with regard to the specific action modality that was chosen: whereas an anterior region of vmPFC was correlated with choice values irrespective of whether the action chosen was an eye or a hand movement, a midregion of vmPFC was correlated with choice values only when a hand movement was selected, and a more posterior region again correlated with choice values only when an eye movement was selected. Such an action-dependent topography could not easily be explained by means of an associative retrieval of goal values depending solely on discriminative stimuli.

These findings raise the question of whether goal-value representations in the vmPFC are necessarily elicited only by action-dependent associations or whether such representations can be elicited purely by discriminative stimuli. To address this question, Wunderlich *et al.*²¹ used a task in which the presentation of stimuli depicting the available decision options was temporally separated from the time at which specific actions could be selected to make a choice. At the beginning of a trial, the subject was presented with a choice between two of three possible stimuli, each of which was associated with distinct drifting probabilities of obtaining reward. After a delay, subjects were subsequently presented with additional symbols informing them whether a given stimulus could be selected by performing an eye movement or else a hand movement. Critically, subjects were not informed at the time when the choice stimuli were presented what particular action was needed to select a particular stimulus; the stimulus–action mappings were only made available subsequently. Thus, subjects could potentially make a choice over stimuli before they could select particular actions to implement that choice. Once again, Wunderlich *et al.* used a computational model to derive trial-by-trial predictions for the chosen-value signal and tested for the presence of this signal at two different time points in the trial: first, when the choice stimuli were initially presented, and second, once the actions required to implement the choice were made available immediately before an action was performed. Remarkably, at the time of stimulus presentation, a robust chosen-value signal

was observed, whereas at the time that the actions were made available, no such signal was found. These findings first of all demonstrate that at least, in some circumstances, choices can be made over stimuli even before the specific actions needed to implement that choice are made available; however, the findings also show that chosen-value signals can be computed even in the absence of specific actions, suggesting that chosen-value signals in vmPFC need not necessarily be elicited by actions but can be elicited by discriminative stimuli as well.

Action values outside of the vmPFC

An action-value signal codes for the value of the expected outcome when taking that particular action in a particular context. For instance, in the restaurant example, if the goal being considered is to obtain Chinese food and the Chinese restaurant is located a few blocks to the south, then taking the action of walking south will have higher value than taking the action of walking north. Unlike chosen values, which as we have seen can also be retrieved by specific actions but are postchoice signals, the action values considered here are prechoice signals. In a decision-making scenario whereby several possible actions can be selected to attain a particular goal, if each possible action is assigned an action value, it is possible to compare these action values to determine which action to select to attain that goal. What evidence exists for the presence of action values in the brain? Within the vmPFC, although some neurophysiology studies have reported neuronal activity related to particular actions³⁵ and, as we have described earlier, there is evidence that chosen-value signals in this region can be elicited by specific actions and/or action modalities, there is arguably no direct evidence as of yet for the existence of pure action-value signals in this area. However, in other parts of the brain there is now emerging evidence suggesting the presence of such signals. Samejima *et al.*,³⁶ recorded from the striatum while monkeys performed a simple choice task in which they could choose between one of four possible actions associated with different amounts of reinforcement in different stimulus conditions. By correlating neural signals against action values generated from a reinforcement-learning (RL) model, they reported that some neurons appeared to uniquely code for the value of individual actions (as opposed to the value of the action ultimately chosen). Lau and

Glimcher³⁷ also reported action-value signals in the striatum using a similar monkey choice paradigm. Many other neurophysiology studies have reported action-dependent value signals in striatum and in areas of the cortex such as LIP sulcus or supplementary motor cortex.^{38–40} However, with respect to LIP, these signals appear not to be pure action-value signals in that they are modulated as a function of what action is ultimately chosen. In humans, the Wunderlich *et al.*'s³² study referred to earlier attempted to identify the existence of action values in the brain above and beyond chosen values. Recall that the paradigm involved required subjects to make a choice between an eye movement and a hand movement, whereby the probability of obtaining a reward after choice of these actions changed over the course of the experiment. The reason in that experiment for requiring choices to be made between effector modalities as opposed to between individual motor movements within a modality was that whereas it might be expected that neurons coding for action values within a modality would be located within the same overlapping area of cortex and/or striatum and thus be difficult to separate with fMRI, action values for between modality effectors might be expected to be represented in spatially distinct brain areas thus enabling their identification with fMRI. To identify action values for eye movements, it was possible to look for areas correlating with the RL-derived value of the eye movement, with the proviso that such value signals are invariant irrespective of whether the eye movement is chosen. Similarly, to identify action values for hand movements, it was possible to look for areas showing correlations with the value of hand movement irrespective of whether that movement was chosen on a given trial. Activity in a region of the presupplementary eye fields in the medial frontal cortex was found to be correlated with action values for eye movements, whereas a nearby region of supplementary motor cortex was found to be correlated with action values for hand movements. These findings suggest that it is indeed possible to detect prechoice action-value signals in the human brain and that these signals can be found in supplementary motor areas. Given the single-unit neurophysiology findings described earlier, it is also likely that such signals are present elsewhere in the human brain such as the dorsal striatum, although in those other areas, if such signals are intermixed at the neuronal population level, conventional

univariate fMRI analyses may not have permitted their identification.

The computation of goal-directed action values: goal values, action contingency, and action costs

Let's consider precisely, in brief, how an action value is constructed. One component of an action value is the signal we've considered already: the relative value of the goal outcome. However, alongside goal values, another critical signal necessary for the computation of action values is the probability with which one can expect to obtain the outcome when a particular action is performed. Tanaka *et al.*⁴¹ conducted a study to investigate neural systems involved in encoding action probabilities. These authors focused specifically on the probability of an outcome given an action and found that activity within the vmPFC and anterior dorsal striatum correlated with the level of correlation between actions and outcomes, consistent with other evidence these regions are involved in encoding the expected value of chosen actions as reviewed earlier. However, to establish the value of an action it is necessary not only to compute the probability of obtaining an outcome given that action was performed, but also to take into account the probability of obtaining that outcome if the action is not performed. Contingency is the term used in animal learning to describe the combination of these two variables: it is the probability of obtaining a particular outcome given that an action is performed relative to the probability of obtaining that outcome given that the action is not performed. Thus, an action can be associated with a high probability of an outcome occurring if that action is initiated (so that it is highly contiguous), yet this action would be very low in contingency if the probability of obtaining the same outcome when the action is not performed is equally high. In other words, if you get the outcome irrespective of whether you perform the action or not, contingency would be very low, whereas if you get the reward only if you perform the action, the action contingency would be high. Liljeholm *et al.*⁴² examined the neural correlates of action contingency. Although vmPFC was still found to be correlated with the probability of reward given an action was performed, this region did not show evidence for encoding of the probability of obtaining an outcome if that action is not performed. Instead, integration of these different

action probabilities was found elsewhere within the inferior parietal lobule and medial frontal gyrus. It is of interest that similar regions to that found to encode contingency have recently been implicated in encoding a "state" prediction error, which could underpin learning of a model of the state-space of a decision problem, in particular of the transition probabilities between states contingent on the performance of particular actions.³⁴ The computation of action contingency could be mediated by a tree search over state-action pairs within such a model-based state-space representation, in which the probability of obtaining a particular outcome given the performance of a sequence of actions is calculated.

In addition to action contingency, another variable of relevance when computing an action value is the cost or effort associated with that action. Naturally, some actions are more effortful than others, and the degree of effort involved in performing a particular action needs to be integrated along with contingency to compute an overall action value and, therefore, to establish whether that particular action should be favored over alternatives. Although little is currently understood about where action effort or costs are encoded, preliminary evidence suggests a role for the anterior cingulate and insular cortex.⁴³ Taken together, this evidence is consistent with the suggestion that the integration of effort costs and contingency to compute action values takes place outside of the vmPFC.

The relationship of the goal-directed system to other systems for behavioral control

In this article, we have focused on the contributions of the vmPFC to goal-directed behavior, by which we mean the selection of actions (A) in a given context (S) to obtain a goal outcome (O). It is, of course, well established that there are at least two other systems involved in controlling behavior: a habit system and a Pavlovian system.^{44–47} In contrast to the goal-directed system, the habit system is concerned with selecting actions on the basis of learned stimulus-response (S–R) associations without any consideration of the incentive value of an outcome or goal state. On the other hand, the Pavlovian system is concerned with eliciting conditioned responses (essentially reflexive skeletomotor behaviors as well as autonomic and other physiological changes) in

response to learned associations between stimuli and behaviorally significant outcomes such as rewards or punishers (see Ref. 45 for a detailed discussion of these different systems). Because each of these different learning systems will often be engaged at the same time during a particular decision-making task, it can be very challenging to work out which of these associative learning systems are controlling behavior during task performance, and to ascertain which of the underlying associations account for observed neural activity.

How can these contributions be differentiated? The classical behavioral assay for discriminating goal-directed from habitual behavior is to devalue the outcome with which a particular action is associated (by, e.g., feeding the subject to satiety on that action). If behavior is controlled by the goal-directed system, responding will decrease on that action, whereas conversely if behavior is driven by stimulus–response associations, responding will remain unchanged after devaluation.⁴⁸ Similarly in the brain, reward-related neural activity pertaining to S–R habits can be differentiated from goal-directed processing by testing for changes in activity at the time of action selection: if activity decreases after devaluation, this implicates the region in goal-directed control; if not, then this could implicate the region in S–R habits. With respect to the vmPFC, precisely this manipulation has been done by Valentin *et al.*, and it has been found that this region follows the response profile consistent with a role for this region in goal-directed but not habitual processing.⁴⁹

However, this still leaves us with a competing Pavlovian account for the vmPFC activity. Evidence inconsistent with a Pavlovian explanation for goal-value signals in the vmPFC itself comes from one of the WTP paradigms described previously.²³ In that study, in addition to the free-choice condition during which participants had to actively select a WTP for each item, participants also took part in a “forced-choice” condition during which the computer program selected a WTP that they were then forced to select. In this condition, no significant activity was observed in the vmPFC correlating with the goal value of the item and this region was significantly less active in the forced-choice compared to the free-choice condition. This result is incompatible with a Pavlovian account, because if activity in response to the goal outcomes did reflect a purely

Pavlovian value, one would expect to find this signal to be present in the forced condition as much as in the free-choice condition (as Pavlovian effects will be present irrespective of choice). Further evidence incompatible with the Pavlovian explanation comes from the study by Glascher *et al.*²⁹ described earlier, in which strong value-related activity was found in the vmPFC during performance of the choice task, even when those actions were selected in the absence of discriminative stimuli (such that a Pavlovian association is unlikely to be elicited). Finally, De Witt *et al.*⁵⁰ performed an fMRI study in which they conducted an instrumental outcome devaluation procedure similar to that described in Valentin *et al.*, but where the devaluation was performed in the absence of discriminative stimuli that could guide performance. As in the Valentin *et al.* study, the vmPFC was found to track the value of the associated outcome. Taken together, therefore, these results suggest that the vmPFC is most likely to be involved in goal-directed component of action selection and not in either habitual or Pavlovian control.

Note that ruling out a role for the vmPFC in Pavlovian mechanisms does not exclude a contribution for this region in using discriminative stimuli to elicit goal-state representations as part of the machinery of the goal-directed system. Unlike the Pavlovian scenario where stimuli come through learning to elicit conditioned reflexes, in this case, discriminative stimuli are used to retrieve information pertaining to the goal state, which is then fed into the system for computing goal-directed action values. As described earlier, there is indeed evidence to suggest that the vmPFC contributes to just such a (non-Pavlovian) stimulus-driven process, but equally as we just considered, the vmPFC can also retrieve goal-value information in the absence of discriminative stimuli. Thus, discriminative stimuli are just one of several mechanisms by which goal-value information can be retrieved.

Multistage process of goal-directed action selection and the role of the vmPFC

Putting all of this together, we can start to build a picture of what a neural system for mediating goal-directed action selection might look like (Fig. 1). First of all, the system will require a detailed model of the decision space, which would consist of representations of the relevant stimuli and actions,

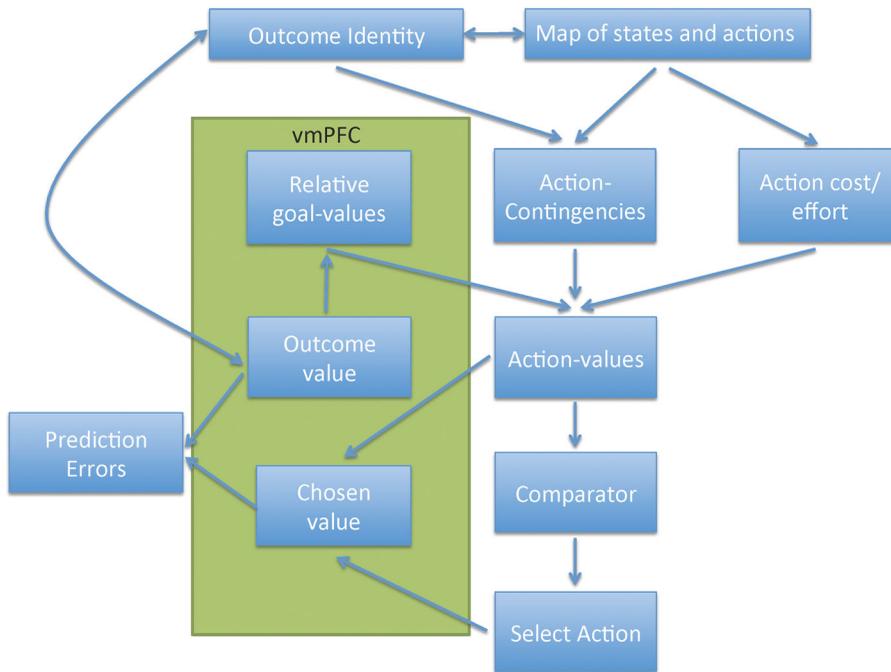


Figure 1. Illustration of some of the key component processes involved in goal-directed decision making within the framework of model-based reinforcement learning. An associative map of stimuli and actions bound together by transition probabilities, together with a representation of the outcome stimuli, forms a representation of the “model.” Within this model representation through learned associations, stimuli or actions can elicit a representation of the associated outcome that in turn retrieves a representation of the incentive value for that outcome (goal). This is done over all known available outcomes, enabling a relative goal-value signal to be computed, taking into account the range of possible outcomes. Action contingencies are then calculated for a given outcome using the forward model machinery, which, when combined with the relative goal-value signal and effort cost, constitute an action value. This signal is in turn fed into a decision comparator that enables selection of a particular action, which is then fed to the motor output system as well as enabling the chosen value to be elicited. The vmPFC is suggested to be involved in encoding a number of these signals (illustrated by the green area), particularly those pertaining to encoding the value of goal outcomes. Yet, the process of computing action values and ultimately generating a choice is suggested to occur outside of this structure.

the transition probabilities between different states contingent on particular actions, and the identity of particular outcomes or goals. At the time of decision making, it would then be possible to retrieve the identity of different goal outcomes, which in turn can elicit through incentive learning,⁴⁴ the hedonic value of the goal outcome in different motivational states. We have argued here for a role of the vmPFC in the retrieval of such goal values at the time of decision making. Second, we have argued for a role for this area in comparing and contrasting between the values of different available goals so that a relative preference ranking is computed over available goals. Next, the normalized goal values need to be combined with action contingencies and action/effort costs to compute overall goal-directed action values. It is argued here that these action value computations may be mediated outside of the vmPFC

in areas such as parts of the parietal cortex, anterior dorsal striatum, and supplementary motor cortex. These action-value signals then must be fed into a decision comparator to ultimately enable action selection. We have not considered in this article where and how this type of action-value comparison might take place—this issue is discussed in detail elsewhere,^{51,52} but in brief there is emerging evidence for the existence of comparator signals in a number of areas in the brain including the lateral intraparietal cortex and the dorsomedial frontal cortex.^{31,53,54} Once the decision is rendered, it has been shown that the vmPFC is once again prominently involved in representing the expected value of the chosen option and the value of the experienced outcome. The value of the chosen outcome can, when combined with the experienced outcome signal, facilitate the computation of reward prediction

errors, which in turn might serve a number of functions even within a model-based RL framework (where reward predictions are not required for value learning *per se*).

It is important to note that the framework proposed here does have strong similarities to other proposals regarding differential functions of these areas,⁵⁵ in which it has been suggested that parts of the dorsal cortex, including the dorso-medial prefrontal cortex, are involved in encoding action-dependent values and decisions over actions, whereas the vmPFC is instead involved in mediating choices over stimuli.⁵⁶ The difference between this proposal and our proposal is that we suggest that rather than being part of a separate stimulus-based choice system, the vmPFC might be better characterized as a contributing node of a goal-directed system in which this area encodes goal values that are used to compute overall values for actions that in turn form an input into the decision process. Although we do suggest that the vmPFC is involved in a form of comparison over goal values, we suggest that this comparison is part of the process of ranking available goals (i.e., to compute a relative value scale) rather than constituting a decision about which goal to select *per se*. On many occasions, the goal ranked highest will turn out to be the one selected, but under situations where, for example, the most desirable goal is also the one requiring the most effort, or where that goal is associated with a low probability of attainment given the available action contingencies, then the most preferred goal will not be synonymous with the action–outcome pair ultimately selected. The proposal advanced here has the advantage of parsimony (having two independent goal-directed choice systems would appear to introduce considerable redundancy), but also takes into account the fact that eventually all choices must be reflected in the selection of a particular action that ultimately has to involve cortical areas involved in the control of movement.

Another alternative perspective is the proposal that all economic decision making is necessarily done in “goods” and not action space, and then once the decision is rendered over goods, the necessary actions are subsequently selected as a secondary stage.^{57,58} Yet it is unclear why a decision should be rendered exclusively over goods without considering the specific individual actions and corresponding unique costs and contingencies at the same time, if

such information is available at the time of decision making. If I make a choice exclusively in the domain of goods, I would still need to compute an overall integrated decision utility whereas at the same time incorporating an estimate of the contingency and costs associated with every possible action available for selecting that good. For instance, in our lunch example, there’s no point in selecting the Chinese option, even although it is my most preferred food, if the only actions available to select that goal involve a treacherous 5-mile walk with a substantial risk of falling into a pothole on the way. Once I have selected a “good” in the abstract, I still need to go back over the contingencies and costs of the available actions to make a subsequent second decision over which particular action I’m going to select to obtain that good. I, thus, would need to perform computations over actions twice: first when generating my overall decision utility for the “good” and in choosing a good, and then subsequently when choosing an action to select that good. A far more parsimonious mechanism would be to make a single choice over action–goal pairs where each action–goal pair is linked to an action value integrating over all costs, benefits, and contingencies associated with that action–goal pair.

It should be noted that, as discussed earlier in the context of the Wunderlich *et al.*²¹ study, clearly it is sometimes possible for decisions to be made about goals without knowing the specific actions available and hence not knowing about specific action costs or contingencies. In that case, it is possible that “proxy” actions are used—for example, “select the red stimulus” to assign values to different options as an input into the decision process. However, once specific actions are known about, then the effort and contingencies associated with those actions need to be incorporated into the decision utility computation and the decision to be rendered is necessarily one to be made over action–goal pairs instead of just over goals.

Conclusions

Here, we reviewed evidence on the functions of the vmPFC in goal-directed decision making. It has been argued that this region may contribute in particular to encoding the value of goals and that the contribution of this region in goal-value processing can be contrasted with that of other regions, such as parts of the parietal cortex, dorsal striatum, and

supplementary motor cortex, which contribute instead to computing the values of particular actions. Thus, although the vmPFC can be seen as a key brain structure involved in encoding the value signals necessary for goal-directed decision making in the mammalian brain, it is only through interactions with other brain systems that values assigned to goal states can be translated into decisions over actions and ultimately into behavior.

Acknowledgments

The author thanks Mimi Liljeholm and Seung-Lark Lim for helpful comments on the manuscript. This work is supported by grants to J.O.D. from the U.S. National Science Foundation, the National Institute of Drug Abuse, and the Gordon and Betty Moore Foundation.

Conflicts of interest

The author declares no conflict of interest.

References

1. Thorpe, S.J., E.T. Rolls & S. Maddison. 1983. The orbitofrontal cortex: neuronal activity in the behaving monkey. *Exp. Brain Res.* **49**: 93–115.
2. Critchley, H.D. & E.T. Rolls. 1996. Hunger and satiety modify the responses of olfactory and visual neurons in the primate orbitofrontal cortex. *J. Neurophysiol.* **75**: 1673–1686.
3. O'Doherty, J. *et al.* 2000. Sensory-specific satiety-related olfactory activation of the human orbitofrontal cortex. *Neuroreport* **11**: 893–897.
4. Rolls, E.T., M.L. Kringelbach & I.E. de Araujo. 2003. Differential representations of pleasant and unpleasant odours in the human brain. *Eur. J. Neurosci.* **18**: 695–703.
5. Anderson, A.K. *et al.* 2003. Dissociated neural representations of intensity and valence in human olfaction. *Nat. Neurosci.* **6**: 196–202.
6. Small, D.M. *et al.* 2003. Dissociation of neural representation of intensity and affective valuation in human gustation. *Neuron* **39**: 701–711.
7. de Araujo, I.E. *et al.* 2003. Human cortical responses to water in the mouth, and the effects of thirst. *J. Neurophysiol.* **90**: 1865–1876.
8. Rolls, E.T. *et al.* 2003. Representations of pleasant and painful touch in the human orbitofrontal and cingulate cortices. *Cereb. Cortex* **13**: 308–317.
9. Blood, A.J. *et al.* 1999. Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. *Nat. Neurosci.* **2**: 382–387.
10. O'Doherty, J. *et al.* 2003. Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia* **41**: 147–155.
11. Cloutier, J. *et al.* 2008. Are attractive people rewarding? Sex differences in the neural substrates of facial attractiveness. *J. Cogn. Neurosci.* **20**: 941–951.
12. Kranz, F. & A. Ishai. 2006. Face perception is modulated by sexual preference. *Curr. Biol.* **16**: 63–68.
13. Kirk, U. *et al.* 2009. Modulation of aesthetic value by semantic context: an fMRI study. *Neuroimage* **44**: 1125–1132.
14. O'Doherty, J. *et al.* 2001. Abstract reward and punishment representations in the human orbitofrontal cortex. *Nat. Neurosci.* **4**: 95–102.
15. Knutson, B. *et al.* 2001. Dissociation of reward anticipation and outcome with event-related fMRI. *Neuroreport* **12**: 3683–3687.
16. Elliott, R., C.D. Frith & R.J. Dolan. 1997. Differential neural response to positive and negative feedback in planning and guessing tasks. *Neuropsychologia* **35**: 1395–1404.
17. Plassmann, H. *et al.* 2008. Marketing actions can modulate neural representations of experienced pleasantness. *Proc. Natl. Acad. Sci. USA* **105**: 1050–1054.
18. de Araujo, I.E. *et al.* 2005. Cognitive modulation of olfactory processing. *Neuron* **46**: 671–679.
19. McClure, S.M. *et al.* 2004. Neural correlates of behavioral preference for culturally familiar drinks. *Neuron* **44**: 379–387.
20. Padoa-Schioppa, C. & J.A. Assad. 2006. Neurons in the orbitofrontal cortex encode economic value. *Nature* **441**: 223–226.
21. Wunderlich, K., A. Rangel & J.P. O'Doherty. 2010. Economic choices can be made using only stimulus values. *Proc. Natl. Acad. Sci. USA* **107**: 15005–15010.
22. Plassmann, H., J.P. O'Doherty & A. Rangel. 2010. Appetitive and aversive goal values are encoded in the medial orbitofrontal cortex at the time of decision making. *J. Neurosci.* **30**: 10799–10808.
23. Plassmann, H., J. O'Doherty & A. Rangel. 2007. Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *J. Neurosci.* **27**: 9984–9988.
24. Montague, P.R. & G.S. Berns. 2002. Neural economics and the biological substrates of valuation. *Neuron* **36**: 265–284.
25. Tremblay, L. & W. Schultz. 1999. Relative reward preference in primate orbitofrontal cortex. *Nature* **398**: 704–708.
26. Padoa-Schioppa, C. & J.A. Assad. 2008. The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. *Nat. Neurosci.* **11**: 95–102.
27. Elliott, R., Z. Agnew & J.F. Deakin. 2008. Medial orbitofrontal cortex codes relative rather than absolute value of financial rewards in humans. *Eur. J. Neurosci.* **27**: 2213–2218.
28. Lim, S.L., J.P. O'Doherty & A. Rangel. 2011. The decision value computations in the vmPFC and striatum use a relative value code that is guided by visual attention. *J. Neurosci.* **31**: 13214–13223.
29. Gläscher, J., A.N. Hampton & J.P. O'Doherty. 2009. Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. *Cereb. Cortex* **19**: 483–495.
30. Hampton, A.N. *et al.* 2007. Contributions of the amygdala to reward expectancy and choice signals in human prefrontal cortex. *Neuron* **55**: 545–555.
31. Hampton, A.N., P. Bossaerts & J.P. O'Doherty. 2006. The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *J. Neurosci.* **26**: 8360–8367.

32. Wunderlich, K., A. Rangel & J.P. O'Doherty. 2009. Neural computations underlying action-based decision making in the human brain. *Proc. Natl. Acad. Sci. USA* **106**: 17199–17204.
33. Daw, N.D., Y. Niv & P. Dayan. 2005. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* **8**: 1704–1711.
34. Glascher, J. *et al.* 2010. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* **66**: 585–595.
35. Feierstein, C.E. *et al.* 2006. Representation of spatial goals in rat orbitofrontal cortex. *Neuron* **51**: 495–507.
36. Samejima, K. *et al.* 2005. Representation of action-specific reward values in the striatum. *Science* **310**: 1337–1340.
37. Lau, B. & P.W. Glimcher. 2007. Action and outcome encoding in the primate caudate nucleus. *J. Neurosci.* **27**: 14502–14514.
38. Sugrue, L.P., G.S. Corrado & W.T. Newsome. 2004. Matching behavior and the representation of value in the parietal cortex. *Science* **304**: 1782–1787.
39. Platt, M.L. & P.W. Glimcher. 1999. Neural correlates of decision variables in parietal cortex. *Nature* **400**: 233–238.
40. Sohn, J.W. & D. Lee. 2007. Order-dependent modulation of directional signals in the supplementary and presupplementary motor areas. *J. Neurosci.* **27**: 13655–13666.
41. Tanaka, S.C., B.W. Balleine & J.P. O'Doherty. 2008. Calculating consequences: brain systems that encode the causal effects of actions. *J. Neurosci.* **28**: 6750–6755.
42. Liljeholm, M. *et al.* 2011. Neural correlates of instrumental contingency learning: differential effects of action-reward conjunction and disjunction. *J. Neurosci.* **31**: 2474–2480.
43. Prevoist, C. *et al.* 2010. Separate valuation subsystems for delay and effort decision costs. *J. Neurosci.* **30**: 14080–14090.
44. Balleine, B.W. & A. Dickinson. 1998. Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* **37**: 407–419.
45. Balleine, B.W., N.D. Daw & J.P. O'Doherty. 2008. Multiple forms of value learning and the function of dopamine. In *Neuroeconomics: Decision Making and the Brain*. P.W. Glimcher *et al.*, Eds.: 367–388. Academic Press. New York.
46. Balleine, B.W. & J.P. O'Doherty. 2010. Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology* **35**: 48–69.
47. Dayan, P. & Y. Niv. 2008. Reinforcement learning: the good, the bad and the ugly. *Curr. Opin. Neurobiol.* **18**: 185–196.
48. Adams, C.D. 1981. Variations in the sensitivity of instrumental responding to reinforcer devaluation. *Q. J. Exp. Psychol.* **34B**: 77–98.
49. Valentin, V.V., A. Dickinson & J.P. O'Doherty. 2007. Determining the neural substrates of goal-directed learning in the human brain. *J. Neurosci.* **27**: 4019–4026.
50. de Wit, S. *et al.* 2009. Differential engagement of the ventromedial prefrontal cortex by goal-directed and habitual behavior toward food pictures in humans. *J. Neuro.* **29**: 11330–11338.
51. Rangel, A. & T. Hare. 2010. Neural computations associated with goal-directed choice. *Curr. Opin. Neurobiol.* **20**: 262–270.
52. Gold, J.I. & M.N. Shadlen. 2007. The neural basis of decision making. *Annu. Rev. Neurosci.* **30**: 535–574.
53. Gold, J.I. & M.N. Shadlen. 2000. Representation of a perceptual decision in developing oculomotor commands. *Nature* **404**: 390–394.
54. Basten, U. *et al.* 2010. How the brain integrates costs and benefits during decision making. *Proc. Natl. Acad. Sci. USA*
55. Rushworth, M.F. *et al.* 2007. Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour. *Trends Cogn. Sci.* **11**: 168–176.
56. Walton, M.E. *et al.* 2010. Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. *Neuron* **65**: 927–939.
57. Padoa-Schioppa, C. 2007. Orbitofrontal cortex and the computation of economic value. *Ann. N. Y. Acad. Sci.* **1121**: 232–253.
58. Grabenhorst, F. & E.T. Rolls. 2011. Value, pleasure and choice in the ventral prefrontal cortex. *Trends Cogn. Sci.* **15**: 56–67.