



# HHS Public Access

Author manuscript

*Nat Neurosci.* Author manuscript; available in PMC 2017 August 03.

Published in final edited form as:

*Nat Neurosci.* 2015 May ; 18(5): 620–627. doi:10.1038/nn.3982.

## What the orbitofrontal cortex does not do

Thomas A Stalnaker, Nisha K Cooch, and Geoffrey Schoenbaum

National Institute on Drug Abuse Intramural Research Program, US National Institutes of Health, Baltimore, Maryland, USA

### Abstract

The number of papers about the orbitofrontal cortex (OFC) has grown from 1 per month in 1987 to a current rate of over 50 per month. This publication stream has implicated the OFC in nearly every function known to cognitive neuroscience and in most neuropsychiatric diseases. However, new ideas about OFC function are typically based on limited data sets and often ignore or minimize competing ideas or contradictory findings. Yet true progress in our understanding of an area's function comes as much from invalidating existing ideas as proposing new ones. Here we consider the proposed roles for OFC, critically examining the level of support for these claims and highlighting the data that call them into question.

---

Invalidation is fundamental to the advancement of scientific ideas<sup>1</sup>. When evaluating new findings, it is important to assess not only which hypotheses are supported, but also which are called into question. Although exciting positive results that establish the viability of a new idea attract the most attention, it is the negative results, particularly across experiments, which help to constrain viable explanations for a particular brain region's function. In this spirit, we offer this review of what we think the OFC does not—and probably does not—do. We focus primarily on what is now termed lateral OFC, encompassing lateral orbital and agranular regions in rats and the limbic or lateral parts of areas 11, 13 and parts of 12 in monkeys (Fig. 1). For reviews focusing more specifically on the functional distinction between medial and lateral OFC regions, see refs. 2–4.

### Response inhibition

Response inhibition is one of the first and perhaps still most influential ideas put forth as the function of prefrontal areas<sup>5</sup>. A general inhibitory function was popularized by the tale of Phineas Gage<sup>6</sup>, who became disinhibited after a traumatic injury affecting his prefrontal and particularly orbital regions. This idea was first operationalized using reversal learning, during which a subject is first taught to associate reward with one cue or response and not another, after which the opposite associations must be learned. OFC damage typically causes deficits in the rapid switching of behavior after reversal of the associations. This is not an isolated finding; deficits have been reported in rodent species, monkeys and humans, in both

---

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

Correspondence should be addressed to G.S. ([geoffrey.schoenbaum@nih.gov](mailto:geoffrey.schoenbaum@nih.gov)).

#### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

go, no-go and choice tasks, and using a variety of learning materials and training procedures<sup>7-18</sup>. Even today, such deficits are often described as reflecting an inability to inhibit responding<sup>7,19,20</sup>.

But is response inhibition a core function of OFC? Although OFC damage typically causes subjects to be slower than normal at acquiring reversals, they are typically unimpaired at acquiring the initial discriminations, despite the fact that these often require inhibiting some default response strategy. This is true even in go, no-go tasks in which the go response has been heavily pre-trained during shaping<sup>21,22</sup>.

These data suggest there is something special about response inhibition after a reversal, and yet, even here, OFC damage does not always lead to deficits in response inhibition<sup>8,18</sup> and may not lead to deficits at all<sup>23-27</sup>. For example, a reversal task involving not two, but three, options, which makes it possible to identify whether reversal errors truly reflect an inability to withhold responding to the previously rewarded option, revealed that monkeys with lateral OFC damage exhibited suboptimal choices, but were in fact more likely to switch responding after reversal<sup>8</sup>.

Previous work has also shown that the OFC is not necessary in a variety of settings that require response inhibition. For example, if monkeys observe a large and a small reward being hidden under two objects, they will subsequently select the one under which the large reward was placed. However, they can be trained to pick the small reward object to get the large reward. OFC lesions do not affect the rate at which the prepotent strategy is reversed<sup>27</sup>. Similarly, OFC damage causes rats to exhibit impulsive or more rapid switching to a small, immediate reward from a larger, increasingly delayed reward<sup>26</sup>.

Furthermore, limited aspiration or even relatively large neurotoxic lesions of the OFC may leave reversal learning completely unaffected<sup>23,24</sup>, even when the lesions impair performance on reinforcer devaluation, another OFC-dependent task<sup>24</sup>. In this case, reversal deficits were reinstated by aspirating a small cortical strip at the caudal end of OFC containing passing fibers from temporal lobe to other prefrontal areas, consistent with ideas that response inhibition is primarily mediated by nearby prefrontal areas rather than OFC<sup>28,29</sup>.

In short, the growing number of reports showing that OFC is not necessary for reversal learning violates a key prediction of the response inhibition hypothesis, which proposes that the OFC should always be necessary for inhibiting responses. As we shall see, there are also numerous studies showing effects of OFC damage on tasks that do not seem to require response inhibition at all. Together, these works provide strong evidence that response inhibition is not a core function of the OFC<sup>28,29</sup> (Fig. 2).

## Flexible representation of stimulus-outcome associations

Another highly influential idea suggests that OFC operates as an especially flexible associative look-up table<sup>30</sup>. This proposal, that OFC is required for the flexible representation of stimulus-outcome associations, arose intuitively from single-unit recordings in which activity apparently tracked the initial and reversed associations during

reversal learning<sup>31–34</sup>. Under this theory, the ability to track associative information would be redundant with that of other regions, with OFC distinguished by the special flexibility of its associative encoding<sup>35</sup>.

However, as reviewed above, the OFC is often not necessary for reversal learning. Just as this evidence contradicts the response inhibition hypothesis, it also contradicts predictions of the flexible encoding idea. This evidence is bolstered by numerous reports in other settings showing that OFC is not necessary for altering established associative behavior, including Pavlovian reversal and extinction, conditioned taste aversion, set shifting, and even changes in instrumental responding after reinforcer devaluation<sup>36–44</sup>. Such evidence at least places serious qualifications on when the OFC's flexible encoding function is brought to bear (Fig. 2).

Moreover, neural representations in many other areas are far more flexible than those in the OFC. For example, amygdalar activity in rats<sup>45</sup> and monkeys<sup>46,47</sup> can encode the associative significance of cues more rapidly and with greater fidelity, even across reversals, than OFC neurons. Similar results have been obtained in recordings from other areas during reversal tasks<sup>48–50</sup>. Encoding in OFC often lags behind changes in other areas, and, in some settings, the flexibility seems to be inversely correlated with the speed of reversal learning<sup>51</sup>. In retrospect, encoding in the OFC seems as remarkable for its inflexibility or specificity as it is for its flexibility and ability to 'reverse'. These observations raise the question of precisely which aspect of associative information is being encoded in OFC<sup>52</sup>, and which task features define when OFC-dependent information is required for flexible behavior.

## Emotions or somatic markers

A third influential proposal, which begins to address the question regarding the content of associative representations in OFC, is that the OFC has a central role in signaling emotions. This idea harkens back to the idea that we experience emotions as a result of peripheral feedback about bodily states<sup>53</sup>. In this implementation, the OFC guides behavior via its modulation of these bodily states or somatic markers.

The central evidence supporting this proposal comes from patients with OFC damage, particularly in the ventromedial part, who are impaired on the Iowa gambling task<sup>11</sup>. In the original version of this task, patients had to choose from four decks of cards: two 'bad' and two 'good'. Bad decks were associated with large gains on each trial, but also often led to large losses, whereas good decks led to relatively small gains on each trial, but had correspondingly small and less frequent losses. Although both normal and brain-damaged subjects began by choosing mostly from decks that yielded large rewards, normal subjects rapidly switched to choosing the small reward decks. This switching was associated with the development of elevated skin conductance, a proxy for arousal and anxiety, during impending choices of the bad deck. Patients with ventromedial OFC damage failed to switch their choices, continuing to choose the bad decks long after controls had stopped and also failing to manifest the skin conductance responses. These impairments, which occurred even though patients could verbally identify the bad decks, were interpreted as reflecting a dissociation between the rational and emotional control of behavior, leading to the idea that

the OFC triggers emotional states about impending events to help guide advantageous choices<sup>54</sup>.

This proposal has had an enormous effect, particularly in reawakening interest in the OFC, yet, although the basic result has been replicated many times in humans and animals<sup>55–62</sup>, there is relatively little evidence that the deficit reflects a fundamental inability to trigger emotions. Patients with OFC damage do not exhibit flat affect or lack of emotion, nor are they unable to engage emotions in decision-making. They simply do so in a way that is unlike normal subjects.

In addition, the OFC-related deficit in the Iowa gambling task turns out to be dependent on the precise arrangement of the contingencies. If the penalties are delivered when subjects first choose the ‘bad’ decks, then OFC-damaged patients show little or no impairment<sup>63</sup>. This suggests that the original formulation of the task, which is highly sensitive to OFC damage, is essentially reproducing the reversal deficit seen after OFC damage in some other settings<sup>64</sup>. As such, it is open to all of the issues with reversal learning impairments discussed above.

Of course, the idea that normal decision-making reflects more than rational, explicit evaluation is certainly a good one. Computational neuroscience and learning theory have hypothesized two general mechanisms of behavioral control: one involving rational simulations of future consequences and the other relying on pre-computed values or policies<sup>65</sup>. Although this is not directly analogous to the distinction between rational decision-making and decision-making guided by somatic markers, it is somewhat similar in that one form of control involves a multi-step, explicit evaluation of options, whereas the other uses a kind of heuristic shortcut to call up a pre-computed value<sup>66</sup>. Recent accounts have even suggested how these two forms of control may interact, with some branches of a decision-tree being pruned out of consideration by the assignment of a pre-computed value to the entire branch<sup>67</sup>. Interestingly, in these somewhat related frameworks, the OFC appears to be much more important for behaviors supported by the process of mental simulation than it is for behaviors relying on pre-computed values or policies<sup>68</sup>—the opposite dissociation proposed by the somatic marker hypothesis. These accounts may be unified by postulating that the OFC is critical for applying emotional information to decisions, but only when such emotions are derived from a process involving mental simulation or inference. In this framework, somatic markers might be either synthesized on-line or pre-computed, with OFC involved in the former, but not the latter. This might be consistent with more recent versions of the somatic marker idea in which the OFC-related impairment has been described as “myopia for the future”<sup>69</sup>. However, such a characterization of the deficit places less emphasis on OFC triggering emotions and more on its involvement in imagining or inferring future outcomes. In this regard, the idea is more similar to several proposals described below than it is to the original conception of somatic markers (Fig. 2).

### **(Economic) value**

A more recent proposal suggests that the OFC is critical for signaling value. Although firmly grounded in the historical understanding of OFC’s role in associative encoding and emotion,

this idea is most strongly associated with the emergence of economic theory as a framework for neuroscience research<sup>70,71</sup>. Here the OFC has been proposed to serve as a final accountant of value, converting information about available outcomes—their probability, magnitude, time to receipt, current desirability, costs, etc.—into a common neural currency on which to base choices, particularly between different goods<sup>72–74</sup>.

The best support for the strong version of this hypothesis came from work showing that, during choices between different juices, the firing rates of some single units in the OFC were correlated with the chosen juice's subjective value<sup>75</sup>. These neural correlates were indifferent to the identity of the juice chosen, the absolute quantity, the direction of the required response, the features of the predictive cues, and even to some extent the period of the trial. They fired simply on the basis of the chosen juice's subjective value, such that a linear function relating value to firing rate could be derived for each juice. Across sessions, the ratio between the slopes of the value functions of the paired juices was predictive of idiosyncratic shifts in preference between them. This relationship between firing ratio and relative preference across sessions rules out any simple explanation that neuronal firing is based on ingredient or another static feature not influenced by or linked to the subjective value the monkey places on the juices.

Subsequent studies have replicated this result and shown further that these neural responses obey key principles for economic value including transitivity and menu invariance<sup>76,77</sup>, and similar results have been reported in humans using functional magnetic resonance imaging (fMRI)<sup>78–81</sup>. For example, BOLD responses in medial orbital areas are correlated with a subject's 'willingness to pay' for items during a decision-making task<sup>78</sup>. In addition, BOLD correlates of value in nearby medial OFC that are specifically invariant to the identity of the expected outcomes have been reported<sup>80</sup> (see refs. 2–4 for more extensive reviews of medial versus lateral subdivisions in primates and humans).

Of course, these are correlative measures. Although no single approach is without drawbacks, this issue is particularly problematic for fMRI, as the analysis inherently aggregates single units and may therefore average out their unique functions, such as coding of information about identity<sup>82</sup>. In unit recording work<sup>75,83,84</sup>, such specific correlates are plentiful, and in at least one instance, the strength of such value-neutral outcome representations was closely related to choice behavior<sup>82</sup>. Although one might argue that the aggregate signal is likely to be the function of the area overall, it seems equally likely that individual units or small ensembles could send their unique product downstream. BOLD signal is also sensitive to input and even subthreshold events. Thus, it may reflect as much what is happening upstream of an area as it does the unique output product of that region, as is the case for prediction error signals reported in ventral striatum commonly thought to reflect dopaminergic input<sup>85</sup>.

Similarly, although the demonstration that individual single units fire in a way that reflects pure value is more definitive, there are issues with using this as a theoretical foundation without additional causal evidence. Value is heavily confounded with arousal<sup>86</sup> and salience<sup>87,88</sup>, and there is some evidence that OFC neurons signal salience (or risk or

decision confidence, which are concepts associated with salience)<sup>89–91</sup>. Disentangling value and related constructs has been problematic in studies of other brain regions<sup>92</sup>.

Although solving all these issues here may not be realistic, they illustrate the limitations of making arguments solely on the basis of neural correlates. What one wants of course is convergent, causal evidence that the OFC is required for the function reflected in the correlates. With the economic value hypothesis, the supporting causal data come from studies of preference<sup>93–95</sup> reporting that OFC-damaged subjects exhibit choices that violate transitivity<sup>96</sup>, which requires that choices reflecting economic value should reveal a consistent rank ordering across a group of items. Simply put, if you prefer A to B and B to C, then you should also prefer A to C.

If OFC signals economic value, then the decisions of OFC-damaged patients should violate transitivity. One study<sup>94</sup> has tested this by asking subjects to choose between pairs of goods (food, people, color). Subjects chose between all possible pairs in a category once. Subsequently the items in each category were arranged in the order that minimized the number of irrational or intransitive choices. 6% of the choices of patients with damage to the ventromedial OFC still violated transitivity, compared with only ~3% of choices of controls or patients with prefrontal damage that excluded the OFC. This result, confirmed in both humans and monkeys<sup>93,95</sup>, provides evidence that consistency in subjective preferences can require the orbital area.

However, consistency of choices in this setting reflects both the ability to signal economic or pure value and the ability to infer structure or relationships among items from a set (for example, as in transitive inference<sup>97</sup>). When preferences are tested in the absence of required inference—that is, preferences among familiar rewards—OFC damage leaves them unaffected<sup>12,98</sup>. This may result from two ways of making preference judgments: an OFC-independent method that relies on a person's historical 'preference history' and a second OFC-dependent method that reflects a 'dynamic assessment of relative value'<sup>94</sup>.

Also problematic to this account, value signals, broadly defined, are generally ubiquitous in the brain, and there is growing evidence that even the very specific economic value correlate is not unique to OFC. For example, similar signals have been found in parietal cortex, anterior cingulate and other prefrontal regions<sup>48–50,99–102</sup>. Recordings from across several prefrontal areas in monkeys choosing between cues predictive of rewards differing in probability, payoff or effort required found that the anterior cingulate cortex had the simplest pure value correlates, as neurons there were more likely than those in OFC or lateral prefrontal cortex to code value monotonically across all three value dimensions<sup>3,50</sup>. In fact a growing number of reports have failed to find integration of different value dimensions in OFC single units when the dimensions are not properties of the actual outcome; thus, although reward identity and magnitude may be integrated, similar studies of the integration of reward magnitude with effort, delay, risk and social value have found largely independent representations<sup>90,102–105</sup>. Of course this doesn't mean the pure value correlates in the OFC are not important on a more restricted scale, nor does it preclude weaker versions of this idea on the basis of multi-unit ensemble coding schemes<sup>71</sup>, but these results do question whether

the OFC's role is all-encompassing or even unique, particularly given that OFC manipulations cannot be said to generally disrupt value-guided behavior<sup>106</sup>.

### **(Inferred) value**

OFC is not typically necessary for value-guided choices<sup>106</sup>. For example, OFC damage produces no deficits in Pavlovian or instrumental learning when subjects are required to discriminate between cues or actions predicting different-sized rewards<sup>12,38,42,107</sup>. Similarly, both blocking and unblocking, when they can be accounted for by value, do not require the OFC<sup>98,108</sup>. On the other hand, OFC is necessary for superficially similar behaviors when they require knowledge of specific outcome features to recognize errors or to infer a value<sup>12,38,42,107–109</sup>.

These observations lead to a weaker or more nuanced form of the value hypothesis, in which the OFC is necessary only when the value driving behavior or learning is derived from mental simulation or model-based processing. Thus, the OFC is not necessary for Pavlovian conditioning, which can be driven by prior experience, but it is necessary for modifying that response if the predicted outcome is devalued by pairing it with illness<sup>38</sup>. The feature of this design that may require OFC is not response inhibition, as sometimes assumed given that lesioned rats extinguish responses normally during the probe test (in which food is omitted), but rather the need to link two independently acquired pieces of associative information, the cue-outcome association and the outcome-illness association. Notably, similar results have been obtained in monkeys after OFC lesions (even fiber sparing)<sup>12,24,110</sup>, and OFC BOLD signal and single-unit responses to predictive cues change selectively after outcome devaluation<sup>32,107,111,112</sup> or preference changes<sup>75</sup>. These deficits are observed even if lesions are made after initial learning and devaluation or if OFC is transiently inactivated only during the critical probe test<sup>113,114</sup>.

This modified version of the value hypothesis explains why OFC is or is not necessary across a wide variety of value-based behaviors. When the behavior is based on previous experience, which would allow a relevant value or policy to be pre-computed without simulating or imagining the future and without integrating new information, then OFC is not necessary. However, if normal behavior requires a novel value to be computed on the fly using new information or predictions that have been acquired since the original learning, then OFC is required. Other higher order processes, such as computing confidence<sup>91,115</sup>, experiencing hypothetical or imagined outcomes<sup>116</sup>, and generating regret<sup>117,118</sup>, may also be products of this sort of representative structure inasmuch as they rely on the ability to mentally simulate information about predicted outcomes. This account would even explain why OFC is necessary for preferences to satisfy transitivity, as in this context, transitive choices require a comparison between items that typically have not been directly and repeatedly experienced together. In other words, for new choices to be transitive, they cannot easily reflect pre-computed or cached valuations (for example, preference history<sup>94</sup>) and instead must rely on inference.

If economic value were conceptualized as only reflecting such computed-on-the-fly or inferred values, then the strong and weak versions of this hypothesis would converge<sup>73</sup>.

Although this assertion seems reasonable, the experimental procedures employed to examine neural correlates of economic value have not explicitly controlled for the associative basis of the underlying behavior in the way that formal learning theory requires (Fig. 2).

Yet even if the OFC is critical to value-based behavior only when it requires a model-based value computation, this leaves open questions regarding the OFC's role in that framework. Does it simply signal the inferred value to allow other areas to compare options or is it essential for some process through which those values are derived? Furthermore, why is the OFC necessary for behavior in situations that do not require value at all when behavior or learning is driven by information about the specific identity of predicted outcomes<sup>42,108</sup>? And finally, how does any explanation of the OFC's role in guiding behavior account for its role in learning<sup>37,108,109,119</sup>? We will address these questions in the next two sections.

## Prediction errors

Why is the OFC sometimes important for learning? One powerful proposal is that the OFC signals prediction errors. Prediction errors have long been hypothesized as key to associative learning<sup>120–122</sup>, and their neural signature has been reported most notably in midbrain dopamine neurons<sup>123–130</sup>. What if the OFC also signaled prediction errors to other areas? Although it would not explain its role in reinforcer devaluation or sensory preconditioning<sup>113,114,119</sup>, as in these cases the OFC is necessary at the time new information is used, direct signaling of errors would provide a powerful mechanism whereby OFC might facilitate learning.

Error signals have indeed been reported in OFC<sup>31,131–134</sup>. Although many such reports involve BOLD signal, which might reflect input from other areas, some report single-unit activity. For example, during reversal of a visual discrimination, some OFC neurons fire on receipt of saline when it is unexpected<sup>31</sup>. The examples reported (4 neurons, 1.3% of the population) did not fire to saline when it was given expectedly, and in one case the neuron also fired on trials when reward was expected, but the pump had been disconnected. More recently, the activity of a much larger proportion of OFC neurons was recorded in rats performing a spatial two-armed bandit task correlated with reward prediction errors<sup>134</sup>. Notably, their T-maze task allowed the rats to have full knowledge of the choices available on each upcoming trial. Thus, when the rats experienced a reward that was better than expected, a change in firing could reflect the reward prediction error; however, it could also reflect the updating of the value of that choice for the next trial, essentially a prediction about the next trial (see ref. 135 for fuller discussion).

Contradicting these findings are a number of negative reports in which OFC single units recorded under conditions that should have elicited prediction errors failed to show any change in firing, even during OFC-dependent learning<sup>37,50,83,136</sup>. In one study, dopamine neurons served as a positive control<sup>37</sup>. This study used an odor-guided choice task that was conceptually similar to the T-maze employed in ref. 134, except that reward delivery did not inform the rats about the available choices on the subsequent trial. With this confound removed, dopamine neurons signaled prediction errors at the time of reward, but OFC neurons did not.



Overall, these data fail to support direct error signaling as a viable explanation for OFC-dependent learning deficits. However, OFC may still participate in learning through its influence on error signals elsewhere. Consistent with this idea, when OFC and midbrain data were juxtaposed, anticipatory activity in the OFC was inversely related to dopaminergic error signaling downstream<sup>37</sup>. This suggests that the error signals in other brain areas might depend partly on OFC input for properly calculating the errors<sup>137</sup> (Fig. 2). This idea has been partially confirmed for error signals in midbrain dopamine neurons, at least in rats<sup>138</sup>. Notably, an indirect role would explain why reinforcement-learning deficits after OFC damage are often (though not always) most evident in response to reward omission or ‘negative feedback’<sup>139</sup>, as these errors require a prediction of reward. Consistent with this idea, the OFC is thought to be particularly important in rats for facilitating reversal learning when contingencies have been stable, which would emphasize the contribution of these predictive signals<sup>18</sup>.

### Credit assignment

Another proposal to explain why the OFC is important for learning is that it is necessary for appropriate credit assignment<sup>2,8,140</sup>. Credit assignment refers to the proper attribution of prediction errors to specific causes<sup>141</sup>. Without proper credit assignment, one might have intact error signaling mechanisms, but lose the ability to learn appropriately or as rapidly as normal, particularly when multiple possible antecedents may be related to the error or when the recent choice history is variable. Under these conditions, it becomes critical to assign credit to the most recent choice and ignore previous alternative selections. This idea was explored using a reversal task with three cues predicting different probabilities of reward<sup>8</sup>. Monkeys with lateral OFC lesions performed normally on the initial discrimination and could track the best option as well as controls when reward probabilities were varied without reversal; however, when the low- and high-probability cues were switched, lesioned monkeys showed the classic OFC-dependent reversal deficit. As noted earlier, by including a third option, the authors showed that the response pattern after reversal did not reflect perseveration, which would have manifested as responding for the previously best option, but instead appeared to reflect somewhat random responding. Closer analysis showed that this occurred because the credit for reward (or non-reward) on the current trial was spread abnormally back to cues selected on preceding trials in lesioned monkeys.

The inability to assign credit appropriately provides an elegant explanation for reversal deficits after OFC damage, as this ability would be useful after reversal when choice patterns become unstable, but not after changes in probability without reversal, when choice patterns remain stable. Only with an unstable pattern of prior choices would the spread of effect lead to markedly impaired learning.

However, taken at face value, this does not provide a straightforward explanation of why the OFC is required in other settings, such as the probe phase of devaluation studies or sensory preconditioning, in which OFC inactivation leads to deficits when learning is not necessary<sup>113,114,119</sup> (Fig. 2). This dichotomy raises the question of whether it is necessary to have separate mechanisms to explain why the OFC is involved in learning versus guiding

behavior or whether there might be deeper underlying functions necessary for both credit assignment and the behavioral control mediated by OFC.

## A cognitive map?

Perhaps the most recent proposal, which may resolve some of the above issues, suggests that the OFC may support the formation of a so-called cognitive map<sup>142</sup> that defines the current task space<sup>143</sup>. Such an associative structure is at the heart of ideas about the implementation of behavioral control. It is necessary for what has been termed goal-directed behavior by learning theorists and model-based behavior by computational neuroscientists. Behavior guided by inference or mental simulation of consequences not directly experienced previously, such as changes in learned behaviors after reinforcer devaluation, would be iconic examples of this. A cognitive map would also be required to generate specific predictions about impending events, such as their identity or features, and for using contextual or temporal structure in the environment to allow old rules to be disregarded so that new ones can be rapidly acquired, as is sometimes the case after reversal. It might also help maintain information about what specific events had just transpired, so that in particularly complex tasks, one could appropriately assign credit when errors are detected.

Of course, constructing and using this associative structure would not depend on any single brain area, but would reflect the operation of a circuit, perhaps spanning much of the brain. This then raises the obvious question of OFC's precise contribution in this circuit. The OFC may represent the underlying structure (for example, the cognitive map) once it has been acquired<sup>143</sup>, perhaps with an emphasis on structures related to biologically relevant outcomes. The OFC receives input from hippocampal areas<sup>144</sup>, which may be important for organizing complex associative representations, and the OFC interacts with broad areas of dorsal and ventral striatum<sup>145,146</sup>. These connections may allow the OFC to acquire and maintain associative representations and to utilize them to influence how simpler associative information is accessed to guide behavior and learning. This would be consistent with the OFC's involvement during both the learning and utilization phases of tasks that require cognitive maps of the task's space<sup>119</sup>. Alternatively, the OFC might only represent the individual parts that comprise the task space, such as the states used to define the space and various events. This would still make the OFC critical to the circuit's operation, but would no longer make this area essential for explicitly signaling value or directly driving learning or response inhibition, all of which appear to characterize some, but not all, of the deficits caused by OFC damage.

The possibility that the OFC provides state information, which is then used by other areas, could resolve the contradictions between the evidence and the various theorized functions discussed above (Fig. 2). Thus, the OFC is not necessary for inhibiting responses, calculating value, signaling errors or even credit assignment *per se* (for example, credit is still assigned in the absence of OFC, but it is not assigned as discretely as normal); however, it is necessary for each of these functions when the function utilizes, or is constrained by, the cognitive map that the OFC provides. For example, reversal learning would be OFC-dependent only when the normal learning rate improves via the formation of a new state space after learning; if the normal learning rate does not require this function, then OFC

lesions (fiber sparing or not) would have no apparent effect. This prediction might be tested with a reversal task in which normal rats showed very low levels of spontaneous recovery or renewal of the original associations. Low levels of recovery would be consistent with unlearning or over-writing of the original learning as a result of a failure to use a new state for post-reversal learning. Under these conditions, reversal learning should not depend on OFC if the OFC is facilitating state creation.

The idea of state creation would explain the OFC's involvement in value-guided behaviors<sup>38,98,119</sup> and would also be broadly consistent with its role in well-constrained learning tasks<sup>108,119</sup>. This view also aligns with unit-recording and fMRI studies that emphasize the complex nature of representations in lateral OFC, where units and BOLD responses are driven not only by reward value, but also by reward identity, cues that precede rewards and even structure in the trial sequences<sup>147,148</sup>. Such rich representations would allow the derivation of values in new situations, would facilitate the appropriate assignment of credit, and so on.

## Conclusions

Invalidation is fundamental to the advancement of scientific ideas. Here we have briefly reviewed a number of ideas concerning OFC function from this perspective. This exercise is important because it helps to narrow down the potential directions for future work. We believe, based on an overview of existing data, that response inhibition, flexible associative encoding, emotion or value alone do not provide adequate, freestanding explanations of OFC function; at best, they explain limited data sets (Fig. 2). Although the verdict is less clear on more recent proposals, such as signaling economic or derived value, credit assignment, and cognitive mapping, we think that there is already some data that at least forces modification of key predictions of many of these accounts. To refine these ideas, it will be important to design experiments that challenge their key predictions and to pay attention to the accumulation of evidence that calls them into question.

## Acknowledgments

The authors would like to thank C. Padoa-Schioppa, J. Wallis and P. Rudebeck for critical readings of earlier versions. This work was supported by grants to G.S. from the National Institute on Drug Abuse, the National Institute of Mental Health and the National Institute on Aging while G.S. was employed at the University of Maryland, Baltimore, and by funding from the National Institute on Drug Abuse at the Intramural Research Program. The opinions expressed in this article are the authors' own and do not reflect the view of the US National Institutes of Health, the Department of Health and Human Services, or the United States government.

## References

1. Popper, K. *The Logic of Scientific Discovery*. Hutchinson & Co; 1959.
2. Noonan MP, Kolling N, Walton ME, Rushworth MF. Re-evaluating the role of the orbitofrontal cortex in reward and reinforcement. *Eur J Neurosci*. 2012; 35:997–1010. [PubMed: 22487031]
3. Wallis JD. Cross-species studies of orbitofrontal cortex and value-based decision-making. *Nat Neurosci*. 2012; 15:13–19.
4. Rudebeck PH, Murray EA. Balkanizing the primate orbitofrontal cortex: distinct subregions for comparing and contrasting values. *Ann NY Acad Sci*. 2011; 1239:1–13. [PubMed: 22145870]
5. Ferrier, D. *The Functions of the Brain*. GP Putnam's Sons; New York: 1876.

6. Harlow JM. Recovery after passage of an iron bar through the head. *Pub Mass Med Soc.* 1868; 2:329–346.
7. Jones B, Mishkin M. Limbic lesions and the problem of stimulus-reinforcement associations. *Exp Neurol.* 1972; 36:362–377. [PubMed: 4626489]
8. Walton ME, Behrens TEJ, Buckley MJ, Rudebeck PH, Rushworth MFS. Separable learning systems in the macaque brain and the role of the orbitofrontal cortex in contingent learning. *Neuron.* 2010; 65:927–939. [PubMed: 20346766]
9. Rolls ET, Hornak J, Wade D, McGrath J. Emotion-related learning in patients with social and emotional changes associated with frontal lobe damage. *J Neurol Neurosurg Psychiatry.* 1994; 57:1518–1524. [PubMed: 7798983]
10. Meunier M, Bachevalier J, Mishkin M. Effects of orbital frontal and anterior cingulate lesions on object and spatial memory in rhesus monkeys. *Neuropsychologia.* 1997; 35:999–1015. [PubMed: 9226661]
11. Bechara A, Damasio H, Tranel D, Damasio AR. Deciding advantageously before knowing the advantageous strategy. *Science.* 1997; 275:1293–1295. [PubMed: 9036851]
12. Izquierdo A, Suda RK, Murray EA. Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. *J Neurosci.* 2004; 24:7540–7548. [PubMed: 15329401]
13. Chudasama Y, Robbins TW. Dissociable contributions of the orbitofrontal and infralimbic cortex to pavlovian autoshaping and discrimination reversal learning: further evidence for the functional heterogeneity of the rodent frontal cortex. *J Neurosci.* 2003; 23:8771–8780. [PubMed: 14507977]
14. Hornak J, et al. Reward-related reversal learning after surgical excisions in orbito-frontal or dorsolateral prefrontal cortex in humans. *J Cogn Neurosci.* 2004; 16:463–478. [PubMed: 15072681]
15. Kim J, Ragozzino KE. The involvement of the orbitofrontal cortex in learning under changing task contingencies. *Neurobiol Learn Mem.* 2005; 83:125–133. [PubMed: 15721796]
16. Fellows LK, Farah MJ. Ventromedial frontal cortex mediates affective shifting in humans: evidence from a reversal learning paradigm. *Brain.* 2003; 126:1830–1837. [PubMed: 12821528]
17. Tsuchida A, Doll BB, Fellows LK. Beyond reversal: a critical role for human orbitofrontal cortex in flexible learning from probabilistic feedback. *J Neurosci.* 2010; 30:16868–16875. [PubMed: 21159958]
18. Riceberg JS, Shapiro ML. Reward stability determines the contribution of orbitofrontal cortex to adaptive behavior. *J Neurosci.* 2012; 32:16402–16409. [PubMed: 23152622]
19. Izquierdo A, Jentsch JD. Reversal learning as a measure of impulsive and compulsive behavior in addictions. *Psychopharmacology (Berl).* 2012; 219:607–620. [PubMed: 22134477]
20. Kringelbach ML, Rolls ET. The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. *Prog Neurobiol.* 2004; 72:341–372. [PubMed: 15157726]
21. Schoenbaum G, Nugent S, Saddoris MP, Setlow B. Orbitofrontal lesions in rats impair reversal but not acquisition of go, no-go odor discriminations. *Neuroreport.* 2002; 13:885–890. [PubMed: 11997707]
22. Schoenbaum G, Setlow B, Nugent SL, Saddoris MP, Gallagher M. Lesions of orbitofrontal cortex and basolateral amygdala complex disrupt acquisition of odor-guided discriminations and reversals. *Learn Mem.* 2003; 10:129–140. [PubMed: 12663751]
23. Kazama A, Bachevalier J. Selective aspiration or neurotoxic lesions of orbital frontal areas 11 and 13 spared monkeys' performance on the object discrimination reversal task. *J Neurosci.* 2009; 29:2794–2804. [PubMed: 19261875]
24. Rudebeck PH, Saunders RC, Prescott AT, Chau LS, Murray EA. Prefrontal mechanisms of behavioral flexibility, emotion regulation and value updating. *Nat Neurosci.* 2013; 16:1140–1145. [PubMed: 23792944]
25. Rudebeck PH, Murray EA. Amygdala and orbitofrontal cortex lesions differentially influence choices during object reversal learning. *J Neurosci.* 2008; 28:8338–8343. [PubMed: 18701696]

26. Winstanley CA, Theobald DEH, Cardinal RN, Robbins TW. Contrasting roles of basolateral amygdala and orbitofrontal cortex in impulsive choice. *J Neurosci*. 2004; 24:4718–4722. [PubMed: 15152031]
27. Chudasama Y, Kralik JD, Murray EA. Rhesus monkeys with orbital prefrontal cortex lesions can learn to inhibit prepotent responses in the reversed reward contingency task. *Cereb Cortex*. 2007; 17:1154–1159. [PubMed: 16774961]
28. Wallis JD. Orbitofrontal cortex and its contribution to decision-making. *Annu Rev Neurosci*. 2007; 30:31–56. [PubMed: 17417936]
29. Aron AR, Robbins TW, Poldrack RA. Inhibition and the right inferior frontal cortex: one decade on. *Trends Cogn Sci*. 2014; 18:177–185. [PubMed: 24440116]
30. Rolls ET. The orbitofrontal cortex. *Phil Trans R Soc Lond B*. 1996; 351:1433–1443. [PubMed: 8941955]
31. Thorpe SJ, Rolls ET, Maddison S. The orbitofrontal cortex: neuronal activity in the behaving monkey. *Exp Brain Res*. 1983; 49:93–115. [PubMed: 6861938]
32. Critchley HD, Rolls ET. Hunger and satiety modify the responses of olfactory and visual neurons in the primate orbitofrontal cortex. *J Neurophysiol*. 1996; 75:1673–1686. [PubMed: 8727405]
33. Critchley HD, Rolls ET. Olfactory neuronal responses in the primate orbitofrontal cortex: analysis in an olfactory discrimination task. *J Neurophysiol*. 1996; 75:1659–1672. [PubMed: 8727404]
34. Rolls ET, Critchley HD, Mason R, Wakeman EA. Orbitofrontal cortex neurons: role in olfactory and visual association learning. *J Neurophysiol*. 1996; 75:1970–1981. [PubMed: 8734596]
35. Cohen JY, Haesler S, Vong L, Lowell BB, Uchida N. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature*. 2012 (in the press).
36. Burke KA, Takahashi YK, Correll J, Brown PL, Schoenbaum G. Orbitofrontal inactivation impairs reversal of Pavlovian learning by interfering with ‘disinhibition’ of responding for previously unrewarded cues. *Eur J Neurosci*. 2009; 30:1941–1946. [PubMed: 19912335]
37. Takahashi YK, et al. The orbitofrontal cortex and ventral tegmental area are necessary for learning from unexpected outcomes. *Neuron*. 2009; 62:269–280. [PubMed: 19409271]
38. Gallagher M, McMahan RW, Schoenbaum G. Orbitofrontal cortex and representation of incentive value in associative learning. *J Neurosci*. 1999; 19:6610–6614. [PubMed: 10414988]
39. Dias R, Robbins TW, Roberts AC. Dissociation in prefrontal cortex of affective and attentional shifts. *Nature*. 1996; 380:69–72. [PubMed: 8598908]
40. McAlonan K, Brown VJ. Orbital prefrontal cortex mediates reversal learning and not attentional set shifting in the rat. *Behav Brain Res*. 2003; 146:97–103. [PubMed: 14643463]
41. Bissonette GB, et al. Double dissociation of the effects of medial and orbital prefrontal cortical lesions on attentional and affective shifts in mice. *J Neurosci*. 2008; 28:11124–11130. [PubMed: 18971455]
42. Ostlund SB, Balleine BW. Orbitofrontal cortex mediates outcome encoding in Pavlovian but not instrumental learning. *J Neurosci*. 2007; 27:4819–4825. [PubMed: 17475789]
43. Izquierdo AD, Murray EA. Bilateral orbital prefrontal cortex lesions disrupt reinforcer devaluation effects in rhesus monkeys. *Soc Neurosci Abstr*. 2000; 26:978.
44. West EA, Forcelli PA, McCue DL, Malkova L. Differential effects of serotonin-specific and excitotoxic lesions of OFC on conditioned reinforcer devaluation and extinction in rats. *Behav Brain Res*. 2013; 246:10–14. [PubMed: 23458741]
45. Schoenbaum G, Chiba AA, Gallagher M. Neural encoding in orbitofrontal cortex and basolateral amygdala during olfactory discrimination learning. *J Neurosci*. 1999; 19:1876–1884. [PubMed: 10024371]
46. Paton JJ, Belova MA, Morrison SE, Salzman CD. The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature*. 2006; 439:865–870. [PubMed: 16482160]
47. Morrison SE, Saez A, Lau B, Salzman CD. Different time courses for learning-related changes in amygdala and orbitofrontal cortex. *Neuron*. 2011; 71:1127–1140. [PubMed: 21943608]
48. Kennerley SW, Dahmubed AF, Lara AH, Wallis JD. Neurons in the frontal lobe encode the value of multiple decision variables. *J Cogn Neurosci*. 2009; 21:1162–1178. [PubMed: 18752411]

49. Kennerley SW, Wallis JD. Evaluating choices by single neurons in the frontal lobe: outcome value encoded across multiple decision variables. *Eur J Neurosci.* 2009; 29:2061–2073. [PubMed: 19453638]
50. Kennerley SW, Behrens TE, Wallis JD. Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nat Neurosci.* 2011; 14:181–1589. [PubMed: 21186356]
51. Stalnaker TA, Roesch MR, Franz TM, Burke KA, Schoenbaum G. Abnormal associative encoding in orbitofrontal neurons in cocaine-experienced rats during decision-making. *Eur J Neurosci.* 2006; 24:2643–2653. [PubMed: 17100852]
52. Rescorla RA. Pavlovian conditioning: it's not what you think it is. *Am Psychol.* 1988; 43:151–160. [PubMed: 3364852]
53. Lang PJ. The varieties of emotional experience: a meditation on James-Lange theory. *Psychol Rev.* 1994; 101:211–221. [PubMed: 8022956]
54. Maia TV, McClelland JL. A reexamination of the evidence for the somatic marker hypothesis: what participants really know in the Iowa gambling task. *Proc Natl Acad Sci USA.* 2004; 101:16075–16080. [PubMed: 15501919]
55. Bechara A, Damasio H, Damasio AR, Lee GP. Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. *J Neurosci.* 1999; 19:5473–5481. [PubMed: 10377356]
56. Bechara A, Tranel D, Damasio H. Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions. *Brain.* 2000; 123:2189–2202. [PubMed: 11050020]
57. Bechara A. Neurobiology of decision-making: risk and reward. *Semin Clin Neuropsychiatry.* 2001; 6:205–216. [PubMed: 11447572]
58. Bechara A, et al. Decision-making deficits, linked to a dysfunctional ventromedial prefrontal cortex, revealed in alcohol and stimulant abusers. *Neuropsychologia.* 2001; 39:376–389. [PubMed: 11164876]
59. Bechara A, Damasio H. Decision-making and addiction (part I): impaired activation of somatic states in substance dependent individuals when pondering decisions with negative future consequences. *Neuropsychologia.* 2002; 40:1675–1689. [PubMed: 11992656]
60. Pais-Vieira M, Lima D, Galhardo V. Orbitofrontal cortex lesions disrupt risk assessment in a novel serial decision-making task for rats. *Neuroscience.* 2007; 145:225–231. [PubMed: 17204373]
61. Zeeb FD, Winstanley CA. Lesions of the basolateral amygdala and the orbitofrontal cortex differentially affect acquisition and performance of a rodent gambling task. *J Neurosci.* 2011; 31:2197–2204. [PubMed: 21307256]
62. Zeeb FD, Winstanley CA. Functional disconnection of the orbitofrontal cortex and basolateral amygdala impairs acquisition of a rat gambling task and disrupts animals' ability to alter decision-making behavior after reinforcer devaluation. *J Neurosci.* 2013; 33:6434–6443. [PubMed: 23575841]
63. Fellows LK, Farah MJ. Different underlying impairments in decision-making following ventromedial and dorsolateral frontal lobe damage in humans. *Cereb Cortex.* 2005; 15:58–63. [PubMed: 15217900]
64. Fellows LK. The role of orbitofrontal cortex in decision making: a component process account. *Ann NY Acad Sci.* 2007; 1121:421–430. [PubMed: 17846161]
65. Daw ND, Niv Y, Dayan P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat Neurosci.* 2005; 8:1704–1711. [PubMed: 16286932]
66. Dayan P, Niv Y, Seymour P, Daw ND. The misbehavior of value and the discipline of the will. *Neural Netw.* 2006; 19:1153–1160. [PubMed: 16938432]
67. Huys QJ, et al. Bonsai trees in your head: how the Pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS Comput Biol.* 2012; 8:e1002410. [PubMed: 22412360]
68. McDannald MA, et al. Model-based learning and the contribution of the orbitofrontal cortex to the model-free world. *Eur J Neurosci.* 2012; 35:991–996. [PubMed: 22487030]
69. Bechara A, Dolan S, Hindes A. Decision-making and addiction (part II): myopia for the future or hypersensitivity to reward? *Neuropsychologia.* 2002; 40:1690–1705. [PubMed: 11992657]
70. O'Doherty JP. The problem with value. *Neurosci Biobehav Rev.* 2014; 43:259–268. [PubMed: 24726573]

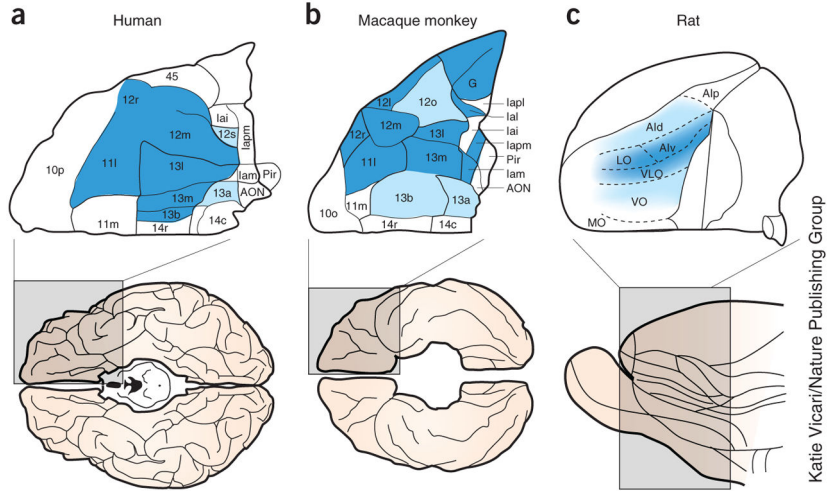
71. Pearson JM, Watson KK, Platt ML. Decision making: the neuroethological turn. *Neuron*. 2014; 82:950–965. [PubMed: 24908481]
72. Montague PR, Berns GS. Neural economics and the biological substrates of valuation. *Neuron*. 2002; 36:265–284. [PubMed: 12383781]
73. Padoa-Schioppa C. Neurobiology of economic choice: a goods-based model. *Annu Rev Neurosci*. 2011; 34:333–359. [PubMed: 21456961]
74. Levy DJ, Glimcher PW. The root of all value: a neural common currency for choice. *Curr Opin Neurobiol*. 2012; 22:1027–1038. [PubMed: 22766486]
75. Padoa-Schioppa C, Assad JA. Neurons in orbitofrontal cortex encode economic value. *Nature*. 2006; 441:223–226. [PubMed: 16633341]
76. Padoa-Schioppa C, Assad JA. The representation of economic value in the orbitofrontal cortex is invariant for changes in menu. *Nat Neurosci*. 2008; 11:95–102. [PubMed: 18066060]
77. Padoa-Schioppa C. Range-adapting representation of economic value in the orbitofrontal cortex. *J Neurosci*. 2009; 29:14004–14014. [PubMed: 19890010]
78. Plassmann H, O’Doherty J, Rangel A. Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *J Neurosci*. 2007; 27:9984–9988. [PubMed: 17855612]
79. McNamee D, Rangel A, O’Doherty JP. Category-dependent and category-independent goal-value codes in human ventromedial prefrontal cortex. *Nat Neurosci*. 2013; 16:479–485. [PubMed: 23416449]
80. Levy DJ, Glimcher PW. Comparing apples and oranges: using reward-specific and reward-general subjective value representation in the brain. *J Neurosci*. 2011; 31:14693–14707. [PubMed: 21994386]
81. Lebreton M, Jorge S, Michel V, Thirion B, Pessiglione M. An automatic valuation system in the human brain: evidence from functional neuroimaging. *Neuron*. 2009; 64:431–439. [PubMed: 19914190]
82. Padoa-Schioppa C. Neuronal origins of choice variability in economic decisions. *Neuron*. 2013; 80:1322–1336. [PubMed: 24314733]
83. McDannald MA, et al. Orbitofrontal neurons acquire responses to ‘valueless’ Pavlovian cues during unblocking. *Elife*. 2014; 3:e02653. [PubMed: 25037263]
84. Stalnaker TA, et al. Orbitofrontal neurons infer the value and identity of predicted outcomes. *Nat Commun*. 2014; 5:3926. [PubMed: 24894805]
85. Knutson B, Gibbs SEB. Linking nucleus accumbens dopamine and blood oxygenation. *Psychopharmacology (Berl)*. 2007; 191:813–822. [PubMed: 17279377]
86. Maunsell JH. Neuronal representations of cognitive state: reward or attention? *Trends Cogn Sci*. 2004; 8:261–265. [PubMed: 15165551]
87. Pearce, JM., Kaye, H., Hall, G. Predictive accuracy and stimulus associability: development of a model for Pavlovian learning. In: Commons, ML, Herrnstein, RJ., Wagner, AR., editors. *Quantitative Analyses of Behavior*. Ballinger; 1982. p. 241–255.
88. Esber GR, Haselgrove M. Reconciling the influence of predictiveness and uncertainty on stimulus salience: a model of attention in associative learning. *Proc Biol Sci*. 2011; 278:2553–2561. [PubMed: 21653585]
89. Ogawa M, et al. Risk-responsive orbitofrontal neurons track acquired salience. *Neuron*. 2013; 77:251–258. [PubMed: 23352162]
90. O’Neill M, Schultz W. Coding of reward risk by orbitofrontal neurons is mostly distinct from coding of reward value. *Neuron*. 2010; 68:789–800. [PubMed: 21092866]
91. Kepecs A, Uchida N, Zariwala HA, Mainen ZF. Neural correlates, computation and behavioural impact of decision confidence. *Nature*. 2008; 455:227–231. [PubMed: 18690210]
92. Leathers ML, Olson CR. In monkeys making value-based decisions, LIP neurons encode cue salience and not action value. *Science*. 2012; 338:132–135. [PubMed: 23042897]
93. Camille N, Griffiths CA, Vo K, Fellows LK, Kable JW. Ventromedial frontal lobe damage disrupts value maximization in humans. *J Neurosci*. 2011; 31:7527–7532. [PubMed: 21593337]
94. Fellows LK, Farah MJ. The role of ventromedial prefrontal cortex in decision making: judgment under uncertainty or judgment per se? *Cereb Cortex*. 2007; 17:2669–2674. [PubMed: 17259643]

95. Rudebeck PH, Murray EA. Dissociable effects of subtotal lesions within the macaque orbital prefrontal cortex on reward-guided behavior. *J Neurosci*. 2011; 31:10569–10578. [PubMed: 21775601]
96. von Neumann, J., Morgenstern, O. *Theory of Games and Economic Behavior*. Princeton University Press; 1947.
97. Bunsey M, Eichenbaum E. Conservation of hippocampal memory function in rats and humans. *Nature*. 1996; 379:255–257. [PubMed: 8538790]
98. Burke KA, Franz TM, Miller DN, Schoenbaum G. The role of the orbitofrontal cortex in the pursuit of happiness and more specific rewards. *Nature*. 2008; 454:340–344. [PubMed: 18563088]
99. Platt ML, Glimcher PW. Neural correlates of decision variables in parietal cortex. *Nature*. 1999; 400:233–238. [PubMed: 10421364]
100. Louie K, Grattan LE, Glimcher PW. Reward value-based gain control: divisive normalization in parietal cortex. *J Neurosci*. 2011; 31:10627–10639. [PubMed: 21775606]
101. Cai X, Padoa-Schioppa C. Neuronal encoding of subjective value in dorsal and ventral anterior cingulate cortex. *J Neurosci*. 2012; 32:3791–3808. [PubMed: 22423100]
102. Hosokawa T, Kennerley SW, Sloan J, Wallis JD. Single-neuron mechanisms underlying cost-benefit analysis in frontal cortex. *J Neurosci*. 2013; 33:17385–17397. [PubMed: 24174671]
103. Watson KK, Platt ML. Social signals in primate orbitofrontal cortex. *Curr Biol*. 2012; 22:2268–2273. [PubMed: 23122847]
104. Roesch MR, Taylor AR, Schoenbaum G. Encoding of time-discounted rewards in orbitofrontal cortex is independent of value representation. *Neuron*. 2006; 51:509–520. [PubMed: 16908415]
105. Blanchard TC, Hayden BY, Bromberg-Martin ES. Orbitofrontal cortex uses distinct codes for different choice attributes in decisions motivated by curiosity. *Neuron*. (in the press).
106. Schoenbaum G, Takahashi YK, Liu TL, McDannald M. Does the orbitofrontal cortex signal value? *Ann NY Acad Sci*. 2011; 1239:87–99. [PubMed: 22145878]
107. Gremel CM, Costa RM. Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. *Nat Commun*. 2013; 4:2264. [PubMed: 23921250]
108. McDannald MA, Lucantonio F, Burke KA, Niv Y, Schoenbaum G. Ventral striatum and orbitofrontal cortex are both required for model-based, but not model-free, reinforcement learning. *J Neurosci*. 2011; 31:2700–2705. [PubMed: 21325538]
109. McDannald MA, Saddoris MP, Gallagher M, Holland PC. Lesions of orbitofrontal cortex impair rats' differential outcome expectancy learning but not conditioned stimulus-potentiated feeding. *J Neurosci*. 2005; 25:4626–4632. [PubMed: 15872110]
110. Machado CJ, Bachevalier J. The effects of selective amygdala, orbital frontal cortex or hippocampal formation lesions on reward assessment in nonhuman primates. *Eur J Neurosci*. 2007; 25:2885–2904. [PubMed: 17561849]
111. Gottfried JA, O'Doherty J, Dolan RJ. Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science*. 2003; 301:1104–1107. [PubMed: 12934011]
112. O'Doherty J, et al. Sensory-specific satiety-related olfactory activation of the human orbitofrontal cortex. *Neuroreport*. 2000; 11:893–897. [PubMed: 10757540]
113. Pickens CL, Saddoris MP, Gallagher M, Holland PC. Orbitofrontal lesions impair use of cue-outcome associations in a devaluation task. *Behav Neurosci*. 2005; 119:317–322. [PubMed: 15727536]
114. West EA, DesJardin JT, Gale K, Malkova L. Transient inactivation of orbitofrontal cortex blocks reinforcer devaluation in macaques. *J Neurosci*. 2011; 31:15128–15135. [PubMed: 22016546]
115. Lak A, et al. Orbitofrontal cortex is required for optimal waiting based on decision confidence. *Neuron*. 2014; 84:190–201. [PubMed: 25242219]
116. Abe H, Lee D. Distributed coding of actual and hypothetical outcomes in the orbital and dorsolateral prefrontal cortex. *Neuron*. 2011; 70:731–741. [PubMed: 21609828]
117. Steiner AP, Redish AD. Behavioral and neurophysiological correlates of regret in rat decision-making on a neuroeconomic task. *Nat Neurosci*. 2014; 17:995–1002. [PubMed: 24908102]
118. Camille N, et al. The involvement of the orbitofrontal cortex in the experience of regret. *Science*. 2004; 304:1167–1170. [PubMed: 15155951]



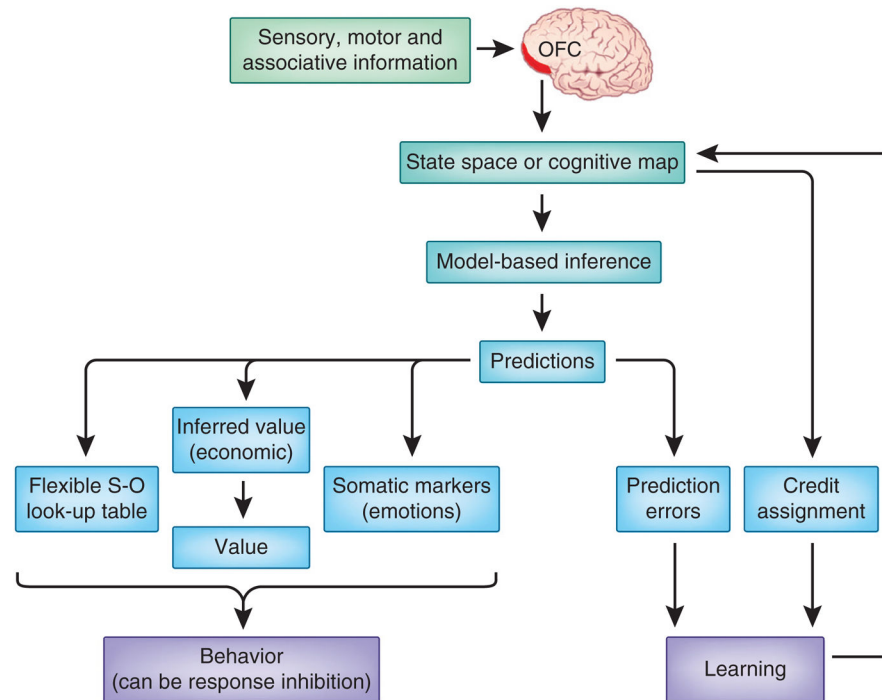
119. Jones JL, et al. Orbitofrontal cortex supports behavior and learning using inferred but not cached values. *Science*. 2012; 338:953–956. [PubMed: 23162000]
120. Rescorla, RA., Wagner, AR. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Black, AH., Prokasy, WF., editors. *Classical Conditioning II: Current Research and Theory*. Appleton Century Crofts; 1972. p. 64-99.
121. Pearce JM, Hall G. A model for Pavlovian learning: variations in the effectiveness of conditioned, but not of unconditioned, stimuli. *Psychol Rev*. 1980; 87:532–552. [PubMed: 7443916]
122. Sutton RS. Learning to predict by the method of temporal difference. *Mach Learn*. 1988; 3:9–44.
123. Mirenowicz J, Schultz W. Importance of unpredictability for reward responses in primate dopamine neurons. *J Neurophysiol*. 1994; 72:1024–1027. [PubMed: 7983508]
124. Schultz W, Dayan P, Montague PR. A neural substrate for prediction and reward. *Science*. 1997; 275:1593–1599. [PubMed: 9054347]
125. Waelti P, Dickinson A, Schultz W. Dopamine responses comply with basic assumptions of formal learning theory. *Nature*. 2001; 412:43–48. [PubMed: 11452299]
126. Pan WX, Schmidt R, Wickens JR, Hyland BI. Dopamine cells respond to predicted events during classical conditioning: evidence for eligibility traces in the reward-learning network. *J Neurosci*. 2005; 25:6235–6242. [PubMed: 15987953]
127. Roesch MR, Calu DJ, Schoenbaum G. Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nat Neurosci*. 2007; 10:1615–1624. [PubMed: 18026098]
128. Bayer HM, Glimcher P. Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*. 2005; 47:129–141. [PubMed: 15996553]
129. Morris G, Nevet A, Arkadir D, Vaadia E, Bergman H. Midbrain dopamine neurons encode decisions for future action. *Nat Neurosci*. 2006; 9:1057–1063. [PubMed: 16862149]
130. D’Ardenne K, McClure SM, Nystrom LE, Cohen JD. BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. *Science*. 2008; 319:1264–1267. [PubMed: 18309087]
131. Schultz W, Dickinson A. Neuronal coding of prediction errors. *Annu Rev Neurosci*. 2000; 23:473–500. [PubMed: 10845072]
132. Tobler PN, O’Doherty J, Dolan RJ, Schultz W. Human neural learning depends on reward prediction errors in the blocking paradigm. *J Neurophysiol*. 2006; 95:301–310. [PubMed: 16192329]
133. Nobre AC, Coull JT, Frith CD, Mesulam MM. Orbitofrontal cortex is activated during breaches of expectation in tasks of visual attention. *Nat Neurosci*. 1999; 2:11–12. [PubMed: 10195173]
134. Sul JH, Kim H, Huh N, Lee D, Jung MW. Distinct roles of rodent orbitofrontal and medial prefrontal cortex in decision making. *Neuron*. 2010; 66:449–460. [PubMed: 20471357]
135. Wallis JD, Rich EL. Challenges of interpreting frontal neurons during value-based decision-making. *Front Neurosci*. 2011; 5:124. [PubMed: 22125508]
136. Takahashi YK, et al. Neural estimates of imagined outcomes in the orbitofrontal cortex drive behavior and learning. *Neuron*. 2013; 80:507–518. [PubMed: 24139047]
137. Schoenbaum G, Roesch MR, Stalnaker TA, Takahashi YK. A new perspective on the role of the orbitofrontal cortex in adaptive behaviour. *Nat Rev Neurosci*. 2009; 10:885–892. [PubMed: 19904278]
138. Takahashi YK, et al. Expectancy-related changes in firing of dopamine neurons depend on orbitofrontal cortex. *Nat Neurosci*. 2011; 14:1590–1597. [PubMed: 22037501]
139. Wheeler EZ, Fellows LK. The human ventromedial frontal lobe is critical for learning from negative feedback. *Brain*. 2008; 131:1323–1331. [PubMed: 18344561]
140. Walton ME, Behrens TE, Noonan MP, Rushworth MF. Giving credit where credit is due: orbitofrontal cortex and valuation in an uncertain world. *Ann NY Acad Sci*. 2011; 1239:14–24. [PubMed: 22145871]
141. Thorndike EL. A proof of the law of effect. *Science*. 1933; 77:173–175.
142. Tolman EC. There is more than one kind of learning. *Psychol Rev*. 1949; 56:144–155. [PubMed: 18128182]

143. Wilson RC, Takahashi YK, Schoenbaum G, Niv Y. Orbitofrontal cortex as a cognitive map of task space. *Neuron*. 2014; 81:267–279. [PubMed: 24462094]
144. Deacon TW, Eichenbaum H, Rosenberg P, Eckmann KW. Afferent connections of the perirhinal cortex in the rat. *J Comp Neurol*. 1983; 220:168–190. [PubMed: 6643724]
145. Voorn P, Vanderschuren LJMJ, Groenewegen HJ, Robbins TW, Pennartz CMA. Putting a spin on the dorsal-ventral divide of the striatum. *Trends Neurosci*. 2004; 27:468–474. [PubMed: 15271494]
146. Groenewegen HJ, Berendse HW, Wolters JG, Lohman AHM. The anatomical relationship of the prefrontal cortex with the striatopallidal system, the thalamus and the amygdala: evidence for a parallel organization. *Prog Brain Res*. 1990; 85:95–116. [PubMed: 2094917]
147. Klein-Flügge MC, Barron HC, Brodersen KH, Dolan RJ, Behrens TE. Segregated encoding of reward-identity and stimulus-reward associations in human orbitofrontal cortex. *J Neurosci*. 2013; 33:3202–3211. [PubMed: 23407973]
148. Schoenbaum G, Eichenbaum H. Information coding in the rodent prefrontal cortex. I Single-neuron activity in orbitofrontal cortex compared with that in pyriform cortex. *J Neurophysiol*. 1995; 74:733–750. [PubMed: 7472378]
149. Ongür D, Price JL. The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cereb Cortex*. 2000; 10:206–219. [PubMed: 10731217]
150. Schoenbaum, G., Setlow, B., Gallagher, M. Orbitofrontal cortex: modeling prefrontal function in rats. In: Squire, L., Schacter, D., editors. *The Neuropsychology of Memory*. Guilford Press; 2002. p. 463-477.



Kattie Vicari/Nature Publishing Group

**Figure 1.** The (lateral) OFC across species. The figure illustrates the OFC in humans and monkeys and the approximate analogous area in rats. The OFC in humans and monkeys is defined as the lateral orbital network (dark blue) and related intermediate areas (light blue) proposed by Price and colleagues<sup>149</sup>. Note that this region is distinct from the medial network, ventromedial prefrontal cortex or what has more recently been called medial OFC. An analogous area has been identified in rats based on connectivity with mediodorsal thalamus, striatum and amygdala, as well as functional criteria<sup>150</sup>.



**Figure 2.** Taxonomy of proposed OFC functions. The diagram shows possible relationships between the various ideas for orbitofrontal function discussed in the review. Although many of these ideas explain some of the data, they generally fail to explain all of the data. Thus, they are no longer viable freestanding explanations for OFC function. However, as illustrated, they may still be viewed as subfunctions of larger or more general concepts, as they are able to explain subsets of the experimental findings. S-O, stimulus-outcome.