Published in final edited form as:
Curr Opin Neurobiol. 2010 April ; 20(2): 205–211. doi:10.1016/j.conb.2010.01.009.

How do you (estimate you will) like them apples? Integration as a defining trait of orbitofrontal function

Geoffrey Schoenbaum[1,2,*] and **Guillem R Esber**[3]

[1]Department of Anatomy and Neurobiology, University of Maryland School of Medicine, 20 Penn St, HSF-2 S251, Baltimore, MD 21201

[2]Department of Psychiatry, University of Maryland School of Medicine, 20 Penn St, HSF-2 S251, Baltimore, MD 21201

[3]Department of Psychological and Brain Sciences, Johns Hopkins University, 3400 N Charles St, Ames Hall, Baltimore, MD 21209

Summary

The past 15 years have seen a rapid increase in our understanding of orbitofrontal function. Today this region is the focus of an enormous amount of research, including work on such complex phenomena as regret, ambiguity and willingness to pay. The orbitofrontal cortex is also credited as a major player in a host of neuropsychiatric diseases. This transformation arguably began with the application of concepts derived from animal learning theory. We will review data from studies emphasizing these approaches to argue that the orbitofrontal cortex forms a critical part of a network of structures that signals information about expected outcomes. Further we will suggest that, within this network, the orbitofrontal cortex provides the critical ability to integrate information in real-time to make what amounts to actionable predictions or estimates about future outcomes. As we will show, the influence of these estimates can be demonstrated experimentally in appropriate behavioral settings, and their operation can also readily explain the role of orbitofrontal cortex in much more complex phenomena such as those cited above.

Introduction

In the past 15 years, we have seen a rapid and exponential increase in our understanding of orbitofrontal function. Today this region is the focus of an enormous amount of research, including work on such complex phenomena as regret, ambiguity and willingness to pay [1–3]. The orbitofrontal cortex is also credited as a major player in a host of neuropsychiatric diseases, including diverse disorders such as addiction, obsessive compulsive disease, mania and depression, and even schizophrenia [4–8]. Here we will argue that this transformation began with the application of concepts derived from animal learning theory. We will review data from studies emphasizing these approaches to argue that the orbitofrontal cortex forms a critical part of a network of structures that signals information about expected outcomes. Further we will suggest that, within this network, the orbitofrontal cortex provides the critical ability to integrate information in real-time to make what amounts to actionable predictions or

*corresponding author (schoenbg@schoenbaumlab.org).

estimates about future outcomes. As we will show, the influence of these estimates can be demonstrated experimentally in appropriate behavioral settings, and their operation can also readily explain the role of orbitofrontal cortex in much more complex phenomena such as those cited above. Additionally we will lay out a number of areas where critical questions remain to be answered.

Orbitofrontal cortex signals information about specific outcomes

An overarching principle of learning theory is that even very simple learning situations involve the acquisition of multiple parallel associative representations [9]. In other words, when an animal is presented with a simple predictive relationship between a cue (e.g. light) and an outcome (e.g. food), the animal does not learn a simple unitary association. Rather the animal actually learns to represent different features of the relationship. While it is unclear how finely the unitary association may be sliced, what is clear is that some gross aspects of it can be dissociated, suggesting they are represented independently in the brain.

A key division is between information about the specific outcome versus information about the general affective or emotional properties that outcome shares with other outcomes; both types of information seem to be accessed independently by the predictive cue after training [10]. This can be shown experimentally by training an animal that a cue predicts a particular outcome and then devaluing the outcome, either by satiation or pairing with illness. If this is done, the animal will subsequently respond less to the predictive cue, even if this cue is never again explicitly paired with the devalued outcome. Further if one pairs the cue with illness, it is possible to induce an aversion to the food [11]. These data show that the cue is able to evoke a representation of the outcome. Changes in behavior are specific to the particular outcome that is devalued and thus suggest that the evoked representation is also specific.

At the same time, however, the predictive cue is still sufficient to drive many behaviors that would seem to reflect its acquired value or significance. For example, conditioned stimuli can still support approach (and avoidance), second order conditioning, conditioned reinforcement, and Pavlovian-to-instrumental transfer after devaluation of their associated outcome [12–15]. These data suggest that in addition to evoking a representation of the specific outcome, a cue can also directly trigger the general affective or emotional properties that the outcome shares with other outcomes.

The application of this idea to the study of neural circuits has led to remarkable advances in our detailed understanding of what different areas do to support associative learning. This is particularly true for the orbitofrontal cortex. Orbitofrontal cortex has long been implicated in associative learning. For example, for decades it has been clear that monkeys with orbitofrontal damage are impaired in learning to respond appropriately to cues predictive of reward, particularly in reversal tasks, and neurons in orbitofrontal cortex signal associative information in this setting [16,17]. However the precise function mediated by orbitofrontal cortex in associative learning has remained elusive. Insight into this question has come from a series of studies examining the role of the orbitofrontal cortex in devaluation paradigms.

These studies have shown clearly that the orbitofrontal cortex is critical for changes in cue-evoked responding caused by devaluation of the predicted outcome. Both monkeys and rats with orbitofrontal lesions learn to respond normally to cues that predict rewards. They also learn to stop eating the associated rewards when they are devalued. However they fail to change their behavior later when presented with cues that predict these outcomes [18–21] (Figure 1A and 1B). These studies have been corroborated by single-unit and fMRI data indicating that the neural activity in orbitofrontal cortex anticipates specific features – including the value – of expected outcomes. For example, in monkeys, unit activity in orbitofrontal cortex to different food items declines with feeding on that food [22]. Similarly BOLD signal in the orbitofrontal

cortex to odors of foods in hungry subjects declines selectively if the subject is given that food to eat [23,24] (Figure 1C).

At the same time, a growing number of reports suggest that the orbitofrontal cortex is often not necessary for normal behavior that is not critically dependent on information about specific outcomes. This is clearly true for a host of behaviors, such as simple discrimination learning and Pavlovian conditioning, in which orbitofrontal lesioned animals perform normally [18, 20,25,26]. In these tasks, behavior is almost certainly mediated by multiple associative structures, thus the necessity for a representation of the specific outcome is likely minimal. Similarly, although orbitofrontal lesions have been reported to affect conditioned reinforcement [27], these effects were complex with some animals exhibiting increased responding. Further orbitofrontal lesions do not have any effect on conditioned reinforcement if the outcome is first devalued, thereby removing any contribution of outcome-related information [28]. On the other hand, orbitofrontal cortex is important when training procedures are used that emphasize the contribution of outcome-specific information in these settings. For example, while orbitofrontal cortex is not necessary for discrimination learning, it is important for the normal facilitation of learning that occurs when different responses lead to different outcomes [29]. Similarly when special training procedures are utilized to minimize associations between the cue and general affective information and force the animal to rely on outcome representations, orbitofrontal cortex becomes essential for conditioned reinforcement [28].

Orbitofrontal cortex signals estimates about future outcomes

Yet even as associative significance was a poor descriptor of the function of the orbitofrontal cortex 20 years ago, the idea that the orbitofrontal cortex is critical for signaling information about specific outcomes also seems scarcely sufficient for describing the complex role this area plays in behavior. For one, the orbitofrontal cortex is only one of a number of areas that is critical to outcome-guided behaviors revealed by devaluation tasks [30–34]. Moreover some of the more complex phenomena suggest that the value-added contribution of orbitofrontal cortex goes beyond signaling of existing information about specific outcomes.

One attractive proposal is that the orbitofrontal cortex may cooperate with downstream associative learning nodes in creating actionable predictions or estimates about future outcomes [35]. Critically, in doing so orbitofrontal cortex would integrate existing knowledge to create new information. This proposal is consistent with the idea that orbitofrontal cortex is a prefrontal region and with proposals from a number of other groups that this region plays a role in executive function or working memory within the domain of value or outcomes [36–38]. According to this model, downstream areas like basolateral amygdala or striatum would store associative information, reflecting acquired knowledge about how the world has worked in the past, whereas the orbitofrontal cortex would integrate this information with other input to create predictions about the future. The orbitofrontal cortex should be particularly important when past experience is insufficient to correctly predict the occurrence of an outcome. In other words, the orbitofrontal cortex should be necessary when behavior requires information about likely future outcomes to be generated on the fly, by combining retrospective rules about how the world has worked in the past in order to generate *estimates* about future outcomes.

A number of lines of evidence are consistent with this hypothesis. For example, as noted above, the orbitofrontal cortex is not necessary for simple Pavlovian conditioning nor is it required for animals to learn to avoid a food that has been paired with illness or satiated [18–21]. However it is essential for changes in conditioned responding when a predicted food has been paired with illness. Critically, this effect of devaluation requires integration of existing representations concerning outcomes. This is evident in the fact that conditioned responding after devaluation is different the very first time the cue is encountered. It does not require new

learning. Instead it reflects the novel combination of two pieces of associative information acquired previously. Consistent with the proposition that this integration requires orbitofrontal cortex, single unit activity [22] and BOLD signal [23,24] in orbitofrontal cortex related to expected outcomes changes with devaluation of that outcome, and damage or inactivation of orbitofrontal cortex immediately prior to the critical probe test prevents the normal effect of devaluation on conditioned responding [19,39]. Notably this is not true for basolateral amygdala, which is another region involved in this effect [32]. If damage to basolateral amygdala is made after the original associations are acquired, devaluation can proceed normally [19].

Another example comes from examining the role of orbitofrontal cortex in extinction learning using a Pavlovian over-expectation task. In this task, rats are trained to associate different cues with reward. After learning, two of the cues are presented in compound, followed by the same amount of reward. Subsequently normal animals will respond less to either of these cues if they are presented alone (Figure 2A). Like devaluation, this decline in responding is observed on the very first trial of the probe test, thus it does not reflect new learning in the probe test. Instead it is thought to reflect the difference between the summed expectations for reward, produced by the two cues, and the actual reward delivered on those compound trials, which is smaller. The resultant prediction error is proposed to lead to extinction of responding in much the same way omitting an expected reward does in a conventional extinction task. In support of this idea, both over-expectation and conventional extinction exhibit spontaneous recovery and renewal [40,41].

The orbitofrontal cortex is critical for changes in responding after over-expectation [42]. Specifically inactivation of orbitofrontal cortex during the compound phase, when signaling of information about expected outcomes is necessary for learning, prevents the later decline in responding in a probe test conducted a day after the last inactivation session (Figure 2B). In the compound sessions, control rats showed summation; that is they respond more to the compounded cues than to the individual control cues. By contrast, rats in whom orbitofrontal cortex is inactivated failed to show summation, instead continuing to respond at normal levels to the two cues. This is consistent the proposal that orbitofrontal cortex is essential for integrating existing information – in this case across cues with unique associative histories - to create estimates about expected outcomes. Contralateral inactivation of orbitofrontal cortex and midbrain also blocked learning (Figure 2C), suggesting that summed outcome expectancies signaled by orbitofrontal cortex might contribute to extinction by supporting error signaling by midbrain dopamine neurons.

In this regard, it is notable that the same rats that fail to summate and show extinction of responding due to over-expectation subsequently exhibited normal extinction of responding when orbitofrontal cortex was inactivated in a conventional setting [43]. Thus orbitofrontal cortex is necessary only when extinction requires integration across cues to derive a novel estimate of the expected outcome; orbitofrontal cortex is not necessary for extinction when elemental associations that simply reflect past experience are sufficient. Of course, normal extinction may often be facilitated by summation of expectancies, as for example when the context or environment also becomes associated with reward. This might be particularly likely in discrimination settings in which inter-trial intervals are often very short, resulting in contextual conditioning. This might explain why in some instances orbitofrontal lesions have been reported to impair conventional extinction learning.

A third example comes from Pavlovian-instrumental transfer. Pavlovian-instrumental transfer refers to the increase in instrumental responding that occurs when a separately trained Pavlovian cue is superimposed on responding. To the extent that this effect results from the integration of unique expectations about an impending outcome, independently derived from

the instrumental response and the Pavlovian cue, then it should depend on orbitofrontal cortex. Consistent with this, it has been reported that orbitofrontal lesions made after training disrupt transfer [44]. Interestingly this effect is more subtle than over-expectation in that lesions only affected transfer when the outcomes were explicitly different. It is possible that the use of an instrumental procedure may have allowed other areas implicated in instrumental learning, such as medial prefrontal cortex, to participate in the summation process. Nevertheless the critical involvement of the orbitofrontal cortex when outcome information is preeminent is consistent with the overall hypothesis.

Conclusions

Here we have reviewed evidence regarding the role of orbitofrontal cortex in associative learning, focusing on studies that have employed concepts from learning theory to specify more precisely the function mediated by this area. Based on this work, we have suggested that the orbitofrontal cortex is critical for signaling information about the specific outcomes that can be expected in a particular situation. This function seems particularly necessary (ie not redundant with that provided by other areas or brain circuits) either when information about specific outcomes is necessary for the behavior or perhaps when information must be integrated to create an estimate of future outcomes.

This function explains a great deal of the data regarding orbitofrontal function. This is certainly true for the vast majority of single unit and fMRI data concerning this area. Data that fails to fit this model is typically open to interpretation either with regard to the function being assessed, the source of the signal, or whether the technique is appropriate for drawing conclusions from negative results.

One notable exception to this is a recent report that neural activity in monkey orbitofrontal cortex represents general utility or value of outcomes rather than signaling information about specific outcomes [45]. This result on its face contradicts simple predictions of this model. However more recent data from this group [46] combined with earlier results [47] suggests that even this apparent general value coding is subject to menu or framing effects. Thus the activity evoked in orbitofrontal cortex to a particular value outcome changes to reflect the context in which that outcome is available; this is evident when different trial types are blocked rather than being randomly interleaved. This effect is consistent with the proposal that signaling in orbitofrontal cortex represents the value of a specific outcome, in this case that available in a particular context or setting. In this regard, a key variable in these studies that is not explored is the influence of training. These monkeys had an enormous amount of experience making selections among marginally different outcomes in the same behavioral setting. Like a currency trader, who can convert euros and dollars and yen effortlessly, the monkeys had learned over tens or even hundreds of thousands of trials to trade between different juice types. We would suggest that the training caused the specific outcome features to be compressed in orbitofrontal cortex, just as the specific context features are compressed if trials are interleaved. Consistent with this, we have found much more specific aspects of outcome value in rats performing a comparable task but given much less training [48].

In any case, correlates are just that – correlates. Even if more general coding is observed in some neurons in orbitofrontal cortex, it remains to be shown that this neural activity is critical to behavior. Indeed all of the behavioral tests of which we are aware that have directly addressed this question suggest it is not. On the other hand, the model we have proposed accounts well for most of these behavioral data. This includes effects of orbitofrontal cortex in the tasks we have described, such as Pavlovian devaluation, conditioned reinforcement, Pavlovian-instrumental transfer, over-expectation, extinction, blocking and unblocking, as well as a

number of other settings that we have not described in detail, including reversal learning, latent inhibition, blocking and unblocking, and even delayed discounting [16,28,49–52].

Further the same logic that explains the role of orbitofrontal cortex in these simple tasks would also account for its involvement in a variety of much more complicated behavioral phenomena studied in humans. For example, as we pointed out earlier, the orbitofrontal cortex has been implicated in regret and counterfactual reasoning [1]. Obviously one cannot suffer regret or consider the implications of a path not taken if one cannot integrate information about how the world works to create estimates of likely outcomes for these options. Indeed we would argue that this functions place a premium on the ability to create novel combinations of information, since often these alternatives have not been experienced. Similarly the preferential involvement of orbitofrontal cortex in signaling what might be called ambiguity (uncertain uncertainty) versus risk (certain uncertainty) [2,53,54] would be expected if this area was particularly necessary for formulating estimates in situations where experience alone is insufficient. Interestingly this function is selectively impaired in patients with obsessive-compulsive disorder [6], a disorder thought to involve orbitofrontal cortex, whereas the opposite dissociation is observed in Parkinson's patients [55], which is not thought to directly impact this region. And finally, willingness to pay, which has been linked to orbitofrontal signaling [3], clearly involves a general if somewhat unspecified integration of existing knowledge and desires in order to imagine some future outcome (e.g. satisfaction). Of course this is only one possible explanation for the role of orbitofrontal cortex in these behavioral phenomena, but it is an explanation most consistent with data from paradigms in which behavior can be more rigorously constrained.

Of course serious questions remain. In addition to questions about effects of training and the role of orbitofrontal cortex in signaling general value, noted above, it is also unclear to what extent there is correspondence between rats and primates. Devaluation and reversal effects appear similar across species. However there is at least one report that even these functions may be topographically segregated in primate orbitofrontal cortex in a manner that would require changes in the specifics of the model we have presented [56]. Further many of the other effects we have highlighted in rat work have not been directly replicated in primates.

Another serious question concerns how closely the role of orbitofrontal cortex in integrating information to estimate likely outcomes is related to its role in signaling outcome specific information. Effects of orbitofrontal inactivation on over-expectation, described above, highlight this issue since the outcomes predicted by the two cues are nominally the same. However although the two outcomes seem similar, they are not identical because they differ in their context and antecedents. Thus they may be perceived and coded by non-overlapping neural circuits or populations in the rat. This suggestion should not be too surprising; context clearly influences the qualities of outcomes. This is apparent when you experience an expensive meal in a novel location versus the same meal in your own kitchen and also in so-called framing effects. Similarly a banana food pellet predicted by a light may be perceived differently – and represented by different neural populations - than the same banana food pellet predicted by a tone. Indeed summation may depend on the extent to which training procedures emphasize or discourage such non-overlapping representations of different outcomes. In other words, if two cues activate precisely the same neural population to represent the outcome, it may be much more difficult to get summation effects than if the cues activate different or even partially different populations. If this testable prediction is true, then the role of orbitofrontal cortex in producing summation would be essentially a by-product of its importance in integrating presentations of unique outcomes, whether those outcomes are unique due to their innate features or due to any other factor causing them to be uniquely represented.

Of course as famously noted by George Box, all models are wrong in some regard. This surely is true here. However, as Dr Box also noted, some models can be useful to the extent they frame existing data in a manner that makes testable predictions. We believe this is the case for this model. It provides testable predictions regarding other situations in which the orbitofrontal cortex and related brain regions should be critical. In testing these predictions, we may come closer to identifying the true functions of this important brain region.

Acknowledgments

This work was supported by grants to Dr Schoenbaum from the NIDA (R01-DA015718), NIMH (R01-MH080865), and NIA (R01-AG027097). In addition, Dr Esber was supported on an NIMH grant to Dr Peter Holland (R01-MH053667). The authors have no competing interests.

References

1. Camille N, Coricelli G, Sallet J, Pradat-Diehl P, Duhamel J-R, Sirigu A. The involvement of the orbitofrontal cortex in the experience of regret. Science 2004;304:1168–1170.
2. Venkatraman V, Payne JW, Bettman JR, Luce MF, Heuttel SA. Separate neural mechanisms underlie choices and strategic preferences in risky decision making. Neuron 2009;62:593–602. [PubMed: 19477159]
3. Plassmann H, O'Doherty J, Rangel A. Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. Journal of Neuroscience 2007;27:9984–9988. [PubMed: 17855612]
4. Schoenbaum G, Shaham Y. The role of orbitofrontal cortex in drug addiction: a review of preclinical studies. Biological Psychiatry 2007;63:256–262. [PubMed: 17719014]
5. Gur RE, Cowell PE, Latshaw A, Turetsky BI, Grossman RI, Arnold SE, Bilker WB, Gur RC. Reduced dorsal and orbital prefrontal gray matter volumes in schizophrenia. Archives of General Psychiatry 2000;57:761–768. [PubMed: 10920464]
6. Starcke T, Tuschen-Caffier B, Markowitsch HJ, Brand M. Dissociation of decisions in ambiguous and risky situations in obsessive-compulsive disorder. Psychiatry Research. 2009 epub ahead of print.
7. Cavedini P, Gorini A, Bellodi L. Understanding obsessive-compulsive disorder: focus on decision making. Neuropsychology Review 2006;16:3–15. [PubMed: 16708289]
8. Drevets WC. Functional anatomical abnormalities in limbic and prefrontal cortical structures in major depression. Progress in Brain Research 2000;126:413–431. [PubMed: 11105660]
9. Rescorla RA. Pavlovian conditioning: it's not what you think it is. American Psychology 1988;43:151–160.
10. Holland PC, Straub JJ. Differential effects of two ways of devaluing the unconditioned stimulus after Pavlovian appetitive conditioning. Journal of Experimental Psychology: Animal Behavior Processes 1979;5:65–78. [PubMed: 528879]
11. Holland PC. Acquisition of representation-mediated conditioned food aversions. Learning and Motivation 1981;12:1–18.
12. Holland PC, Rescorla RA. The effects of two ways of devaluing the unconditioned stimulus after first and second-order appetitive conditioning. Journal of Experimental Psychology: Animal Behavior Processes 1975;1:355–363. [PubMed: 1202141]
13. Holland PC. Relations between Pavlovian-Instrumental transfer and reinforcer devaluation. Journal of Experimental Psychology: Animal Behavior Processes 2004;30:104–117. [PubMed: 15078120]
14. Parkinson JA, Roberts AC, Everitt BJ, Di Ciano P. Acquisition of instrumental conditioned reinforcement is resistant to the devaluation of the unconditioned stimulus. Quarterly Journal of Experimental Psychology 2005;58:19–30. [PubMed: 15844375]
15. Rescorla RA. Transfer of instrumental control mediated by a devalued outcome. Animal Learning and Behavior 1994;22:27–33.
16. Jones B, Mishkin M. Limbic lesions and the problem of stimulus-reinforcement associations. Experimental Neurology 1972;36:362–377. [PubMed: 4626489]
17. Thorpe SJ, Rolls ET, Maddison S. The orbitofrontal cortex: neuronal activity in the behaving monkey. Experimental Brain Research 1983;49:93–115.

18. Gallagher M, McMahan RW, Schoenbaum G. Orbitofrontal cortex and representation of incentive value in associative learning. Journal of Neuroscience 1999;19:6610–6614. [PubMed: 10414988]
19. Pickens CL, Setlow B, Saddoris MP, Gallagher M, Holland PC, Schoenbaum G. Different roles for orbitofrontal cortex and basolateral amygdala in a reinforcer devaluation task. Journal of Neuroscience 2003;23:11078–11084. [PubMed: 14657165]
20. Izquierdo AD, Suda RK, Murray EA. Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. Journal of Neuroscience 2004;24:7540–7548. [PubMed: 15329401]
21. Machado CJ, Bachevalier J. The effects of selective amygdala, orbital frontal cortex or hippocampal formation lesions on reward assessment in nonhuman primates. European Journal of Neuroscience 2007;25:2885–2904. [PubMed: 17561849]
22. Critchley HD, Rolls ET. Hunger and satiety modify the responses of olfactory and visual neurons in the primate orbitofrontal cortex. Journal of Neurophysiology 1996;75:1673–1686. [PubMed: 8727405]
23. Gottfried JA, O'Doherty J, Dolan RJ. Encoding predictive reward value in human amygdala and orbitofrontal cortex. Science 2003;301:1104–1107. [PubMed: 12934011]
24. O'Doherty J, Rolls ET, Francis S, Bowtell R, McGlone F, Kobal G, Renner B, Ahne G. Sensory-specific satiety-related olfactory activation of the human orbitofrontal cortex. Neuroreport 2000;11:893–897. [PubMed: 10757540]
25. 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]
26. Dias R, Robbins TW, Roberts AC. Dissociation in prefrontal cortex of affective and attentional shifts. Nature 1996;380:69–72. [PubMed: 8598908]
27. Pears A, Parkinson JA, Hopewell L, Everitt BJ, Roberts AC. Lesions of the orbitofrontal but not medial prefrontal cortex disrupt conditioned reinforcement in primates. Journal of Neuroscience 2003;23:11189–11201. [PubMed: 14657178]
28. Burke KA, Franz TM, Miller DN, Schoenbaum G. The role of orbitofrontal cortex in the pursuit of happiness and more specific rewards. Nature 2008;454:340–344. [PubMed: 18563088]
29. McDannald MA, Saddoris MP, Gallagher M, Holland PC. Lesions of orbitofrontal cortex impair rats' differential outcome expectancy learning but not conditioned stimulus-potentiated feeding. Journal of Neuroscience 2005;25:4626–4632. [PubMed: 15872110]
30. Malkova L, Gaffan D, Murray EA. Excitotoxic lesions of the amygdala fail to produce impairment in visual learning for auditory secondary reinforcement but interfere with reinforcer devaluation effects in rhesus monkeys. Journal of Neuroscience 1997;17:6011–6020. [PubMed: 9221797]
31. Baxter MG, Parker A, Lindner CCC, Izquierdo AD, Murray EA. Control of response selection by reinforcer value requires interaction of amygdala and orbitofrontal cortex. Journal of Neuroscience 2000;20:4311–4319. [PubMed: 10818166]
32. Hatfield T, Han JS, Conley M, Gallagher M, Holland P. Neurotoxic lesions of basolateral, but not central, amygdala interfere with Pavlovian second-order conditioning and reinforcer devaluation effects. Journal of Neuroscience 1996;16:5256–5265. [PubMed: 8756453]
33. Mitchell AS, Browning PG, Baxter MG. Neurotoxic lesions of the medial mediodorsal nucleus of the thalamus disrupt reinforcer devaluation effects in rhesus monkeys. Journal of Neuroscience 2007;27:11289–11295. [PubMed: 17942723]
34. de Borchgrave R, Rawlins JNP, Dickinson A, Balleine BW. Effects of cytotoxic nucleus accumbens lesions on instrumental conditioning in rats. Experimental Brain Research 2002;144:50–68.
35. Schoenbaum G, Roesch MR, Stalnaker TA, Takahashi YK. A new perspective on the role of the orbitofrontal cortex in adaptive behaviour. Nature Reviews Neuroscience. 2009 AOP.
36. Goldman-Rakic, PS. Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In: Mountcastle, VB.; Plum, F.; Geiger, SR., editors. Handbook of Physiology: The Nervous System. Vol. vol V. American Physiology Society; 1987. p. 373-417.
37. Wallis JD. Orbitofrontal cortex and its contribution to decision-making. Annual Review of Neuroscience 2007;30:31–56.

38. Montague PR, Berns GS. Neural economics and the biological substrates of valuation. Neuron 2002;36:265–284. [PubMed: 12383781]
39. Pickens CL, Saddoris MP, Gallagher M, Holland PC. Orbitofrontal lesions impair use of cue-outcome associations in a devaluation task. Behavioral Neuroscience 2005;119:317–322. [PubMed: 15727536]
40. Rescorla RA. Spontaneous recovery from overexpectation. Learning and Behavior 2006;34:13–20. [PubMed: 16786880]
41. Rescorla RA. Renewal from overexpectation. Learning and Behavior 2007;35:19–26. [PubMed: 17557388]
42. Takahashi Y, Roesch MR, Stalnaker TA, Haney RZ, Calu DJ, Taylor AR, Burke KA, Schoenbaum G. The orbitofrontal cortex and ventral tegmental area are necessary for learning from unexpected outcomes. Neuron 2009;62:269–280. [PubMed: 19409271]
43. 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. European Journal of Neuroscience. 2009 epub ahead of print.
44. Ostlund SB, Balleine BW. Orbitofrontal cortex mediates outcome encoding in Pavlovian but not instrumental learning. Journal of Neuroscience 2007;27:4819–4825. [PubMed: 17475789]
45. Padoa-Schioppa C, Assad JA. Neurons in orbitofrontal cortex encode economic value. Nature 2006;441:223–226. [PubMed: 16633341]
46. Padoa-Schioppa C. Range-adapting representation of economic value in the orbitofrontal cortex. Journal of Neuroscience 2009;29:14004–14014. [PubMed: 19890010]
47. Tremblay L, Schultz W. Relative reward preference in primate orbitofrontal cortex. Nature 1999;398:704–708. [PubMed: 10227292]
48. 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]
49. Winstanley CA, Theobald DEH, Cardinal RN, Robbins TW. Contrasting roles of basolateral amygdala and orbitofrontal cortex in impulsive choice. Journal of Neuroscience 2004;24:4718–4722. [PubMed: 15152031]
50. Rudebeck PH, Walton ME, Smyth AN, Bannerman DM, Rushworth MF. Separate neural pathways process different decision costs. Nature Neuroscience 2006;9:1161–1168.
51. Schiller D, Weiner I. Lesions to the basolateral amygdala and the orbitofrontal cortex but not to the medial prefrontal cortex produce an abnormally persistent latent inhibition in rats. Neuroscience 2004;128:15–25. [PubMed: 15450350]
52. Mobini S, Body S, Ho M-Y, Bradshaw CM, Szabadi E, Deakin JFW, Anderson IM. Effects of lesions of the orbitofrontal cortex on sensitivity to delayed and probabilistic reinforcement. Psychopharmacology 2002;160:290–298. [PubMed: 11889498]
53. Kepecs A, Uchida N, Zariwala HA, Mainen ZF. Neural correlates, computation and behavioural impact of decision confidence. Nature 2008;455:227–231. [PubMed: 18690210]
54. van Duuren E, van der Plasse G, Lankelma J, Joosten RN, Feenstra MG, Pennartz CM. Single-cell and population coding of expected reward probability in the orbitofrontal cortex of the rat. Journal of Neuroscience 2009;29:8965–8976. [PubMed: 19605634]
55. Euteneuer F, Schaefer F, Stuermer R, Boucsein W, Timmermann L, Barbe MT, Ebersbach G, Otto J, Kessler J, Kalbe E. Dissociation of decision-making under ambiguity and decision-making under risk in patients with Parkinson's disease: a neuropsychological and psychophysiological study. Neuropsychologia 2009;47:2882–2890. [PubMed: 19545579]
56. 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. Journal of Neuroscience 2009;29:2794–2804. [PubMed: 19261875]

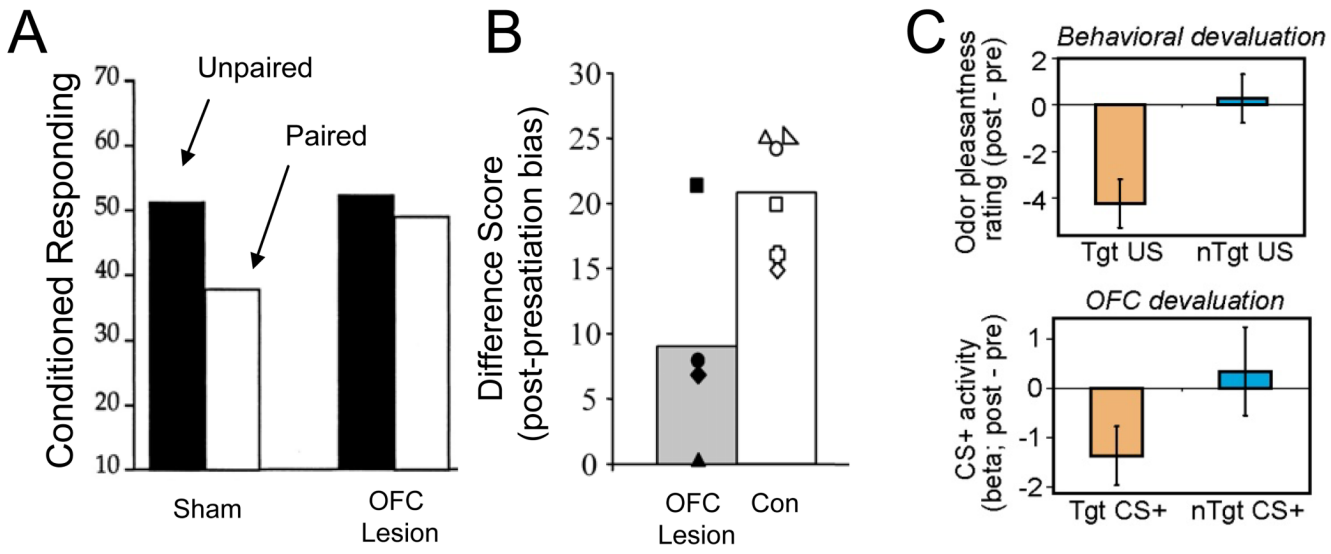

Figure 1.

The role of orbitofrontal cortex in changing conditioned responding as a result of reinforcer devaluation. A. Changes in Pavlovian conditioned responding in sham and orbitofrontal-lesioned rats after reinforcer devaluation. Rats were trained to associate a light cue with food. Subsequently the food was devalued by pairing it with illness, then responding to the cue was assessed in a final probe session. Orbitofrontal-lesioned rats showed normal conditioning to the cues and stopped eating the food when it was paired with illness but, as shown in the figure, failed to change conditioned responding as a result of devaluation in the final probe test. B. Changes in discriminative responding in sham and orbitofrontal lesioned monkeys after reinforcer devaluation. Monkeys were trained to associate different objects with different food rewards. Subsequently one food was devalued by over-feeding, then discrimination performance was assessed in a probe test. As illustrated in the figure by a difference score comparing pre- and post-satiation bias, orbitofrontal-lesioned monkeys failed to bias their choices away from objects associated with the satiated food. C. Changes in BOLD signal in human orbitofrontal cortex after reinforcer devaluation. Subjects were scanned during presentation of odors of different foods. Subsequently one food was devalued by over-feeding, then subjects were re-scanned. As illustrated in the figure, appetive ratings of the odor and BOLD response in orbitofrontal cortex declined to odors of satiated (Tgt CS/US) but not non-satiated (nTgt CS/US) foods. Adapted from Gallagher et al., Journal of Neuroscience, 1999, Izquierdo et al, Journal of Neuroscience, 2004, and Gottfried et al, Science, 2003.

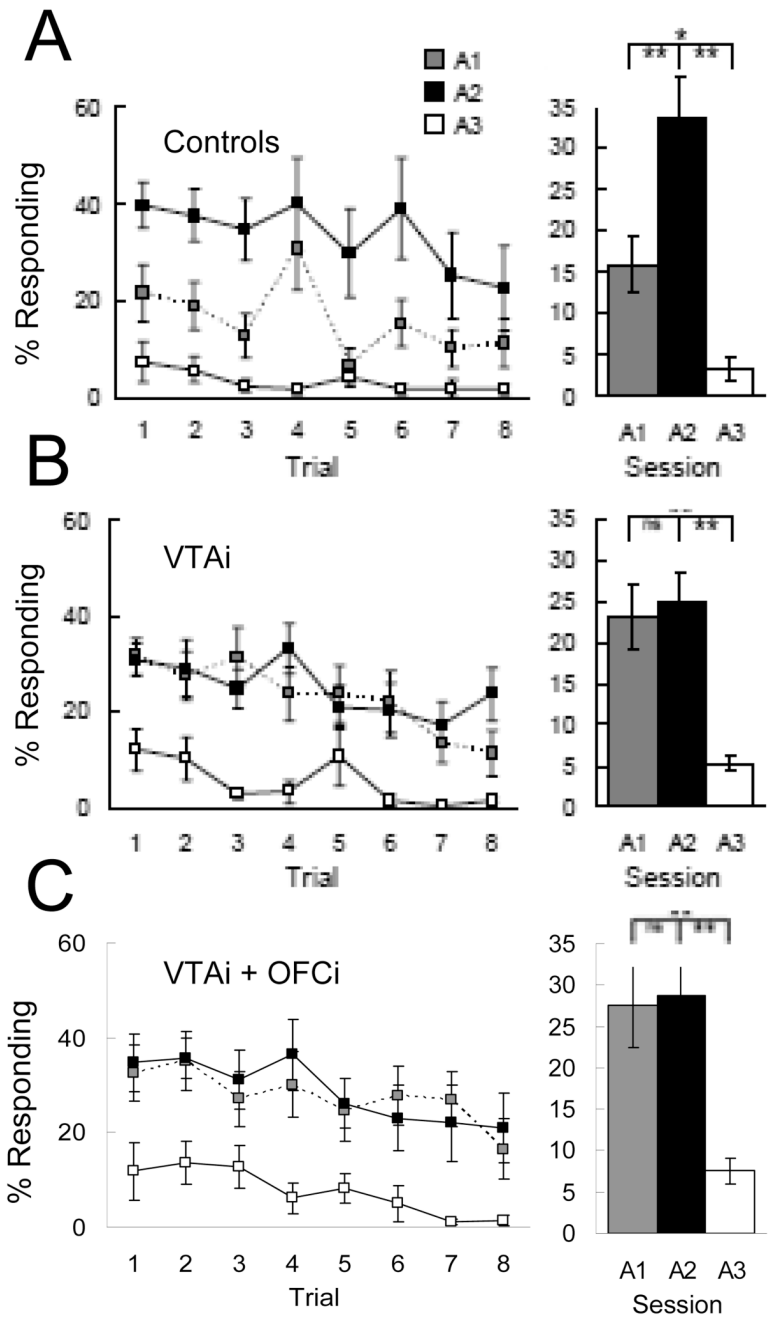

Figure 2.

Effect of inactivation of bilateral inactivation of ventral tegmental area or contralateral inactivation orbitofrontal cortex + ventral tegmental area on changes in behavior after over-expectation. Rats in all groups conditioned normally and maintained responding during compound training (though only controls showed summation to the compound cue). Shown is food cup responding to the auditory cues during the critical probe test. As in Figure 13, controls (A) exhibited weaker responding in this probe test to the cue that had been compounded. This decline in responding was not observed if ventral tegmental area had been inactivated bilaterally (B) during prior compound training or if ventral tegmental area and orbitofrontal

cortex were disconnected via contralateral inactivation (C). Adapted from Takahashi et al, Neuron, 2009.