Reconciling the Roles of Orbitofrontal Cortex in Reversal Learning and the Encoding of Outcome Expectancies

GEOFFREY SCHOENBAUM[a,b], **MICHAEL P. SADDORIS**[c], and **THOMAS A. STALNAKER**[a]

a*Departments of Anatomy and Neurobiology and Psychiatry, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA*

b*Department of Psychology, University of Maryland Baltimore County, Baltimore, Maryland 21228, USA*

c*Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, Maryland 21218, USA*

Abstract

Damage to orbitofrontal cortex (OFC) has long been associated with decision-making deficits. Such deficits are epitomized by impairments in reversal learning. Historically, reversal learning deficits have been linked to a response inhibition function or to the rapid reversal of associative encoding in OFC neurons. However here we will suggest that OFC supports reversal learning not because its encoding is particularly flexible—indeed it actually is not—but rather because output from OFC is critical for flexible associative encoding downstream in basolateral amygdala (ABL). Consistent with this argument, we will show that reversal performance is actually inversely related to the flexibility of associative encoding in OFC (i.e., the better the reversal performance, the less flexible the encoding). Further, we will demonstrate that associative correlates in ABL are more flexible during reversal learning than in OFC, become less flexible after damage to OFC, and are required for the expression of the reversal deficit caused by OFC lesions. We will propose that OFC facilitates associative flexibility in downstream regions, such as ABL, for the same reason that it is critical for outcome-guided behavior in a variety of setting—namely that processing in OFC signals the value of expected outcomes. In addition to their role in guiding behavior, these outcome expectancies permit the rapid recognition of unexpected outcomes, thereby driving new learning.

Keywords

orbitofrontal cortex; basolateral amygdala; reversal; associative learning; expectancies

---

Orbitofrontal cortex (OFC) has long been implicated in cognitive flexibility.[1,2] The loss of this ability is evident in impairments in rapid reversal learning that are observed after OFC damage in a variety of species.[1,3-13] In these settings, animals are first taught to respond to one cue to receive reward and to withhold or inhibit a similar response to avoid punishment or non-reward. After animals are responding correctly based on these contingencies, the meaning of the cues is reversed, such that animals must respond to the cue to which they had previously withheld responding and withhold responding to the cue to which they had previously responded. Animals with OFC damage typically acquire the initial discrimination normally but require many more trials than controls to relearn the discrimination after the cue-outcome associations are reversed. This deficit has been taken as evidence that OFC promotes flexible responding. Yet why is OFC important for this function?

Address for correspondence: Geoffrey Schoenbaum, M.D., Ph.D., 20 Penn Street HSF-2, Rm S251, Baltimore, MD 21201. Voice: 410-706-3814; fax: 410-706-2512. schoenbg@schoenbaumlab.org.

THE ORBITOFRONTAL CORTEX AND RESPONSE INHIBITION

The role of OFC in promoting reversal learning has been explained historically as a result of the involvement of OFC in inhibiting responses.[1,2] While this has been proposed to be a general prefrontal function, reports often identify disinhibited, perseverative, or impulsive responding as hallmarks of the "orbitofrontal syndrome." Thus, much of the behavior of orbitofrontal damaged humans has been conceptualized as an inability to inhibit so-called "prepotent" responses. To the extent that this is more than just a restatement of the deficit, this proposal suggests that output from OFC somehow directly inhibits expression of the response, as illustrated in Figure 1A. Unfortunately this account is contradicted by the observation that, in most of the discrimination reversal studies in which OFC lesions impair reversal, lesioned animals are fully capable, before reversal, of withholding a response that is identical to the one that they are unable to withhold after reversal. This is particularly true in go, no-go tasks, in which this initial response is often highly ingrained. For example, in our studies the rats are shaped for many hundreds of trials to respond for reward on every trial. Yet rats with OFC lesions learn to inhibit this response at the same rate as controls when presented with a series of discrimination problems, both on the initial and subsequent problems. Only after reversal are these rats unable to withhold responding.

Moreover, animals with OFC lesions are also able to inhibit responses normally in many other settings. For example, although OFC-lesioned monkeys and rats fail to suppress cue-evoked conditioned responses after reinforcer devaluation, they do inhibit consummatory responses like controls.[8,14,15] In these tasks, the animals are trained to associate cues with rewards. If the rewards are devalued by overfeeding or pairing with illness, normal animals exhibit reduced responding to the cues that predict those rewards. OFC-lesioned animals fail to show this decline in conditioned responding. However, they are able to suppress actual consumption of the devalued food item after this critical probe test. Further, during the probe test, which is conducted under extinction conditions, lesioned rats show normal extinction of conditioned responding. Thus while OFC-lesioned rats fail to suppress conditioned responding as a result of prior devaluation of the outcome, they are able to suppress conditioned responding as a result of non-reward. In fact, it has recently been reported that rats with OFC lesions can even suppress instrumental responses normally after reinforcer devaluation.[16]

Furthermore, it has recently been shown that OFC-lesioned monkeys are able to inhibit prepotent responses normally in the so-called reversed reward contingency task.[17] In this task, the monkeys were asked to select the smaller of two available food quantities in order to obtain presentation of the larger quantity. Thus monkeys were presented with a choice between one half peanut or four half peanuts. In order to obtain the four half peanuts, the monkeys had to select the one half peanut. This requires the monkeys to suppress their natural tendency to select the thing that they want to receive, which is the larger of the two rewards. OFC-lesioned monkeys learned to do this at the same rate as controls.

These studies show that there are circumstances in which OFC is not required for response inhibition. Thus while there may be situations in which OFC promotes response inhibition, response inhibition per se is not the underlying function of this area. In other words, disinhibited and impulsive responding after OFC lesions is a symptom, rather than an explanation.

THE ORBITOFRONTAL CORTEX AS AN ASSOCIATIVE LOOK-UP TABLE

More recently, the role of OFC in promoting reversal learning has been linked to the finding that cue-selective neurons in OFC reverse firing selectivity for cues during reversal learning. [18-20] Reversal of cue-selective activity in OFC neurons presumably reflects acquisition of the reversed associations. An example of such flexible associative encoding is given in Figure 2A, which shows a single neuron recorded in OFC in a rat learning and reversing a novel odor

discrimination problem. During initial learning, the neuron becomes selective to odor two, which predicts quinine, then after reversal, the same neuron switches to fire to odor one now that this odor predicts quinine. Similar correlates are also observed for the positive odor cue. When these neurons were first reported by Rolls and colleagues in great numbers in OFC,[19] flexible associative encoding had not been prominently demonstrated in other brain areas. The apparent uniqueness of these correlates, in combination with the close correspondence between their apparent function and the effects of OFC damage, led to the readily accepted proposal that the critical contribution of OFC to behavior was to be a rapidly modifiable associative look-up table of sorts. By this account, illustrated in Figure 1B, the reversed associations would be learned most rapidly in OFC and then employed to drive the behavioral reversal.

However more recent neurophysiological findings are inconsistent with this hypothesis. Specifically, though single-unit recording studies in rats and primates have shown that about a quarter of cue-selective OFC neurons reverse their stimulus selectivity after reversal of the stimulus-outcome associations,[18,20,21] the population of cue-selective neurons actually becomes non-selective after reversal and is, in effect, replaced by the emergence of a new cue-selective population. This is illustrated in Figure 3A, which shows the effect of reversal of a two-odor discrimination problem on the firing of OFC neurons selective for each odor cue during learning. Both populations fire selectively during sampling of their respective cue before reversal but not after reversal. Loss of selectivity is also shown graphically for each neuron in the two populations by the inset scatter plot in Figure 3A. This figure plots a cue-selectivity index for each neuron before and after reversal. This calculation consisted of the difference in firing rate to the odor cues divided by the sum of those rates. If a population of neurons maintained the same cue-selectivity before and after reversal, one would expect a regression with a positive correlation. On the other hand, if a population of neurons reversed cue-selectivity, one would expect a regression with a negative correlation. However, this plot shows that cue-selective neurons in OFC are actually equally likely to fire to either odor cue after reversal. As a result, there is no correlation between the cue-selectivity indices of these OFC neurons before and after reversal. This encoding pattern is at odds with the proposal that OFC rapidly encodes reversals of cue-outcome associations.

Furthermore, we have recently reported that the probability of observing reversal of cue-selectivity in OFC neurons is actually inversely related to the rate of reversal learning.[22] This relationship, illustrated in Figure 2B, is inconsistent with the notion that flexible encoding in OFC drives behavioral reversal, since if it did one would expect the opposite relationship (i.e., better performance when reversing neurons are plentiful). Instead, this relationship suggests that reversal of cue-selectivity in OFC reflects feedback regarding discrepancies between expected and actual outcomes. Such feedback would be greatest when reversal performance is poor. Indeed, OFC BOLD response is sensitive to violations of expected outcomes, and a small number of neurons in OFC respond strongly when errors are detected.[23-25]

Finally, recent work shows that associative encoding in other brain regions can be more flexible than that in OFC. The best example of this is activity in basolateral amygdala (ABL), a region that receives strong projections from OFC.[26-28] ABL is critical for associative learning,[29-32] and neurons in ABL rapidly become selective to cues that predict biologically relevant appetitive or aversive outcomes.[18,33-35] An analysis of the effect of reversal on cue-selectivity in ABL[18,34,36] is presented in Figure 3B. This figure, which parallels the analysis of OFC neurons in Figure 3A, shows the effect of reversal of a two-odor discrimination problem on the firing of ABL neurons selective for each odor cue during learning. Both populations switch their cue-selectivity across reversal, thereby rapidly encoding the reversed cue-outcome associations. This pattern is also evident in the inset scatterplot, which plots the cue-selectivity index for each neuron before and after reversal. These values were inversely correlated across the population, consistent with a reversal of cue-selectivity. Comparison of data in Figure 3A

and B is particularly revealing, showing that though OFC has received the greatest attention for the flexibility of its neural correlates, encoding in ABL neurons actually appears to provide a more rapidly flexible associative look-up table. Importantly this impression has been confirmed in primates in the past year in an elegant study by Salzman and colleagues showing that primate amygdala neurons reverse cue-selectivity in large proportions during reversal learning.[37]

THE ORBITOFRONTAL CORTEX AND ENCODING OF OUTCOME-EXPECTANCIES

So what then is the contribution of OFC to flexible responding in situations such as reversal learning? An alternative explanation for the role of OFC in promoting flexible responding is found in the neural activity in OFC that anticipates expected outcomes.[25,38-45] This is evident across different species and tasks; neural activity or BOLD changes in OFC are often triggered by cues and events that convey information about the value of impending outcomes.[25,38,40,43,44,46-51] (Note that this is different from prediction error signaling thought to characterize dopaminergic activity.[52]) The importance of OFC for signaling such information is evident in lesion studies in which damage to OFC causes selective deficits in the ability of animals to use information about the value of expected outcomes to guide existing or established behavior. For example, as described earlier OFC-lesioned animals fail to modify responding to cues after reinforcer devaluation.[8,14,15] Deficits in these settings, which do not involve new learning, likely reflect a critical function for OFC in modulating the expression of previously acquired associative information, perhaps in downstream areas such as the basal ganglia or amygdala.[53]

We have proposed that neural encoding of expected outcomes by OFC might also provide a signal that could be compared to actual outcomes to drive new learning.[54] When an actual outcome fails to match expectations, this comparison could promote changes in old associative representations and the acquisition of new ones. Presumably this function would be helpful even in initial learning, but it might be particularly critical during reversals, in which the discrepancy between actual and expected outcomes is maximal. By this account, illustrated in Figure 1C, OFC would support reversal learning not directly due to its role in rapid encoding new associations but rather indirectly, by facilitating changes in associative encoding in other brain regions. Such old or persistent memories might otherwise slow or impede behavioral reversal. Perhaps the best candidate region for showing this effect is ABL; as outlined above, ABL is critically involved in associative learning, and associative encoding in ABL normally changes rapidly after reversal. This proposal makes at least two testable predictions:

1. The flexibility of associative correlates in ABL should depend on input from OFC.

2. Expression of the reversal deficit caused by OFC lesions should be mediated through ABL.

We will describe these predictions—and the data that addresses them—in the remaining sections of this chapter.

Flexibility of Associative Encoding in Basolateral Amygdala Depends on Input from Orbitofrontal Cortex

The first prediction of our hypothesis is that associative encoding in amygdala should depend on OFC. Consistent with this proposal, we have recently reported that cue-selective firing in ABL neurons fails to reverse in rats with unilateral OFC lesions.[34] These data are illustrated in Figure 4, which shows the effects of reversal of a two-odor discrimination on the firing of cue-selective ABL neurons recorded in rats with ipsilateral neurotoxic lesions of OFC. Because the connections between OFC and ABL are largely (though not completely) ipsilateral, this

lesion would have substantially diminished the influence of OFC on ABL in the lesioned hemisphere while leaving processing in the intact hemisphere largely unaffected. This manipulation caused associative correlates in ABL to become both less prevalent and, when followed across reversal, significantly less flexible. This is evident in the population histograms, in which neither population of cue-selective neurons shows any selectivity to the odor cues after reversal, and in the scatterplot, where there was no correlation between cue-selectivity before and after reversal in the cue-selective neurons. Thus removal of OFC dramatically reduced the flexibility of encoding in ABL.

Of course associative encoding (and flexibility of that encoding) was not completely eliminated by OFC lesions. Obviously such preserved encoding could reflect input from the intact contralateral hemisphere or compensatory mechanisms in the lesioned brain. However the partially preserved encoding is also consistent with recent studies showing that some outcome-guided behaviors depend on ABL but not OFC. For example, ABL lesions cause deficits in fear conditioning,[30,55-57] but OFC lesions typically do not.[58] ABL lesions also abolish conditioned stimulus-potentiated feeding,[59,60] while OFC lesions do not.[61] Thus ABL neurons may be able to encode associative information, albeit at a slower rate, even in the absence of signaling from OFC. This would be consistent with the idea that other brain areas may signal aspects of impending outcomes independent of contributions from OFC.

Reversal Deficits Caused by Orbitofrontal Lesions Are Mediated through Basolateral Amygdala

The second prediction of our hypothesis is that the reversal deficits caused by OFC lesions should be mediated by the inflexible correlates in ABL. In other words, if OFC normally promotes reversal by facilitating changes in associative encoding in ABL, then damage to ABL in OFC-lesioned animals should mitigate or even abolish the reversal impairment. Note that this prediction contrasts with the prediction that other ideas regarding OFC function would make. For example, if OFC normally promotes reversal directly either by inhibiting responding or by rapidly acquiring the reversed associations, then damage to these downstream areas should either worsen or have no effect on the OFC-dependent reversal deficit. To test this, we assessed reversal learning in rats with bilateral neurotoxic lesions of either OFC or ABL and rats with bilateral OFC lesions combined with bilateral ABL lesions or unilateral ABL lesions plus inactivation of contralateral ABL during reversal.[62] The results are shown in Figure 5. Consistent with our hypothesis, we found that removal or inactivation of ABL completely abolished the reversal deficit caused by OFC lesions.

Importantly, lesions of ABL in otherwise intact rats did not facilitate reversal learning, and neither did ABL lesions have any effect on the acquisition of the initial discriminations. This argues against an account by which ABL lesions corrected the reversal impairment by an effect on initial acquisition of the discriminations or by independently improving cognitive flexibility. The absence of effects of ABL lesions on reversal learning is consistent with our own prior work and other reports that have used fiber-sparing lesion techniques to examine the role of amygdala in reversal learning.[6,8,61,63] In these studies, damage to amygdala did not facilitate or impair reversal learning performance, although ABL damage did affect differential changes in response latencies, consistent with a role in encoding the associative information acquired with learning. (Note that OFC + ABL lesions do not fix the latency effects observed with either lesion in isolation—see supplemental material in Ref. 62.) The simplest explanation of these results is that changes in such associative encoding in ABL during reversal learning normally occur more rapidly than those in other regions (at least in paradigms similar to ours). As a result, removing ABL has little effect in otherwise intact rats. However, OFC lesions impair reversal of associative encoding in ABL, rendering it slow enough to retard the rate of behavioral reversal.

CONCLUSIONS

Damage to OFC has long been associated with deficits in adaptive decision making. Such deficits are epitomized by impairments in reversal learning. Here we have shown that reversal performance is actually inversely related to the flexibility of associative encoding in OFC. Further, we have demonstrated that associative correlates in ABL are more flexible than those in OFC during reversal learning, become less flexible after damage to OFC, and are required for the expression of the reversal deficit caused by OFC lesions. This is consistent with the proposal that OFC supports rapid reversal learning by facilitating associative flexibility in downstream regions, such as ABL, rather than due to any role in directly driving reversals either through a role in response inhibition or in rapidly encoding the reversed associations. This proposal is consistent with the idea that OFC is critical for signaling the value of expected outcomes. In addition to guiding behavior, these outcome expectancies permit the rapid recognition of unexpected outcomes, thereby driving new learning.

References

1. Jones B, Mishkin M. Limbic lesions and the problem of stimulus-reinforcement associations. Exp Neurol 1972;36:362–377. [PubMed: 4626489]

2. Mishkin, M. Perseveration of central sets after frontal lesions in monkeys. In: Warren, JM.; Akert, K., editors. The Frontal Granular Cortex and Behavior. McGraw-Hill; New York: 1964. p. 219-241.

3. Schoenbaum G, et al. Orbitofrontal lesions in rats impair reversal but not acquisition of go, no-go odor discriminations. Neuroreport 2002;13:885–890. [PubMed: 11997707]

4. Bohn I, Giertler C, Hauber W. Orbital prefrontal cortex and guidance of instrumental behavior in rats under reversal conditions. Behav Brain Res 2003;143:49–56. [PubMed: 12842295]

5. Dias R, Robbins TW, Roberts AC. Dissociation in prefrontal cortex of affective and attentional shifts. Nature 1996;380:69–72.

6. Schoenbaum G, et al. Lesions of orbitofrontal cortex and basolateral amygdala complex disrupt acquisition of odor-guided discriminations and reversals. Learn Mem 2003;10:129–140. [PubMed: 12663751]

7. Rolls ET, et al. 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]

8. Izquierdo AD, 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]

9. 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]

10. 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]

11. 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]

12. Brown VJ, McAlonan K. Orbital prefrontal cortex mediates reversal learning and not attentional set shifting in the rat. Behav Brain Res 2003;146:97–130. [PubMed: 14643463]

13. Teitelbaum H. A comparison of effects of orbitofrontal and hippocampal lesions upon discrimination learning and reversal in the cat. Exp Neurol 1964;9:452–462. [PubMed: 14188532]

14. Gallagher M, McMahan RW, Schoenbaum G. Orbitofrontal cortex and representation of incentive value in associative learning. J Neurosci 1999;19:6610–6614. [PubMed: 10414988]

15. Pickens CL, et al. Different roles for orbitofrontal cortex and basolateral amygdala in a reinforcer devaluation task. J Neurosci 2003;23:11078–11084. [PubMed: 14657165]

16. Ostlund SB, Balleine BW. Orbitofrontal cortex mediates outcome encoding in Pavlovian but not instrumental learning. J Neurosci 2007;27:4819–4825. [PubMed: 17475789]

17. 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 2006;17:1154–1159. [PubMed: 16774961]

18. 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]

19. Thorpe SJ, Rolls ET, Maddison S. The orbitofrontal cortex: neuronal activity in the behaving monkey. Exp Brain Res 1983;49:93–115. [PubMed: 6861938]

20. Rolls ET, et al. Orbitofrontal cortex neurons: role in olfactory and visual association learning. J Neurophysiol 1996;75:1970–1981. [PubMed: 8734596]

21. 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]

22. Stalnaker TA, et al. Abnormal associative encoding in orbitofrontal neurons in cocaine-experienced rats during decision-making. Euro J Neurosci 2006;24:2643–2653.

23. Nobre AC, et al. Orbitofrontal cortex is activated during breaches of expectation in tasks of visual attention. Nat Neurosci 1999;2:11–12. [PubMed: 10195173]

24. Tobler PN, et al. Human neural learning depends on reward prediction errors in the blocking paradigm. J Neurophysiol 2006;95:301–310. [PubMed: 16192329]

25. Feierstein CE, et al. Representation of spatial goals in rat orbitofrontal cortex. Neuron 2006;51:495–507. [PubMed: 16908414]

26. Krettek JE, Price JL. Projections from the amygdaloid complex to the cerebral cortex and thalamus in the rat and cat. J Comp Neurol 1977;172:225–254.

27. Kita H, Kitai ST. Amygdaloid projections to the frontal cortex and the striatum in the rat. J Comp Neurol 1990;298:40–49. [PubMed: 1698828]

28. Shi CJ, Cassell MD. Cortical, thalamic, and amygdaloid connections of the anterior and posterior insular cortices. J Comp Neurol 1998;399:440–468. [PubMed: 9741477]

29. Weiskrantz L. Behavioral changes associated with ablations of the amygdaloid complex in monkeys. J Comp Physiol Psychology 1956;9:381–391.

30. Davis, M. The role of the amygdala in conditioned and unconditioned fear and anxiety. In: Aggleton, JP., editor. The Amygdala: a Functional Analysis. Oxford University Press; Oxford: 2000. p. 213-287.

31. Gallagher, M. The amygdala and associative learning. In: Aggleton, JP., editor. The Amygdala: a Functional Analysis. Oxford University Press; Oxford: 2000. p. 311-330.

32. Everitt, BJ., et al. Differential involvement of amygdala subsystems in appetitive conditioning and drug addiction. In: Aggleton, JP., editor. The Amygdala: a Functional Analysis. Oxford University Press; New York: 2000. p. 353-390.

33. Quirk GJ, Armony JL, LeDoux JE. Fear conditioning enhances different temporal components of tone-evoked spike trains in auditory cortex and lateral amygdala. Neuron 1997;19:613–624. [PubMed: 9331352]

34. Saddoris MP, Gallagher M, Schoenbaum G. Rapid associative encoding in basolateral amygdala depends on connections with orbitofrontal cortex. Neuron 2005;46:321–331. [PubMed: 15848809]

35. Quirk GJ, Repa JC, LeDoux JE. Fear conditioning enhances short-latency auditory responses of lateral amygdala neurons: parallel recordings in the freely behaving rat. Neuron 1995;15:1029–1039. [PubMed: 7576647]

36. Stalnaker TA, et al. Cocaine-induced decision-making deficits are mediated by miscoding in basolateral amygdala. Nat Neurosci 2007;10:949–951. [PubMed: 17603478]

37. Patton JJ, et al. The primate amygdala represents the positive and negative value of visual stimuli during learning. Nature 2006;439:865–870. [PubMed: 16482160]

38. Schoenbaum G, Chiba AA, Gallagher M. Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. Nat Neurosci 1998;1:155–159. [PubMed: 10195132]

39. Padoa-Schioppa C, Assad JA. Neurons in orbitofrontal cortex encode economic value. Nature 2006;441:223–226. [PubMed: 16633341]

40. Schoenbaum G, et al. Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala. Neuron 2003;39:855–867. [PubMed: 12948451]

41. Schoenbaum G, et al. Encoding changes in orbitofrontal cortex in reversal-impaired aged rats. J Neurophysiol 2006;95:1509–1517. [PubMed: 16338994]

42. Hikosaka K, Watanabe M. Delay activity of orbital and lateral pre-frontal neurons of the monkey varying with different rewards. Cereb Cortex 2000;10:263–271. [PubMed: 10731221]

43. 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]

44. Hikosaka K, Watanabe M. Long-and short-range reward expectancy in the primate orbitofrontal cortex. Eur J Neurosci 2004;19:1046–1054. [PubMed: 15009152]

45. Tremblay L, Schultz W. Reward-related neuronal activity during go-no go task performance in primate orbitofrontal cortex. J Neurophysiol 2000;83:1864–1876. [PubMed: 10758098]

46. Roesch MR, Olson CR. Neuronal activity related to reward value and motivation in primate frontal cortex. Science 2004;304:307–310. [PubMed: 15073380]

47. O'Doherty J, et al. Neural responses during anticipation of a primary taste reward. Neuron 2002;33:815–826. [PubMed: 11879657]

48. Gottfried JA, O'Doherty J, Dolan RJ. Encoding predictive reward value in human amygdala and orbitofrontal cortex. Science 2003;301:1104–1107. [PubMed: 12934011]

49. Roesch MR, Olson CR. Neuronal activity in primate orbitofrontal cortex reflects the value of time. J Neurophysiol 2005;94:2457–2471. [PubMed: 15958600]

50. Tremblay L, Schultz W. Relative reward preference in primate orbitofrontal cortex. Nature 1999;398:704–708. [PubMed: 10227292]

51. Blair K, et al. Choosing the lesser of two evils, the better of two goods: specifying the roles of ventromedial prefrontal cortex and dorsal anterior cingulate in object choice. J Neurosci 2006;26:11379–11386. [PubMed: 17079666]

52. Calu DJ, Roesch MR, Schoenbaum G. Orbitofrontal cortex does not signal reward prediction errors. Society for Neuroscience Abstracts 2007:749.16.

53. Frank MJ, Claus ED. Anatomy of a decision: striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. Psychol Rev 2006;113:300–326. [PubMed: 16637763]

54. Schoenbaum G, Roesch MR, Stalnaker TA. Orbitofrontal cortex, decision-making, and drug addiction. Trends Neurosci 2006;29:116–124. [PubMed: 16406092]

55. Maren S, Goosens KA. Contextual and auditory fear conditioning are mediated by the lateral, basal, and central amygdaloid nuclei in rats. Learn Mem 2001;8:148–155. [PubMed: 11390634]

56. Amorapanth P, Ledoux JE, Nader MA. Different lateral amygdala outputs mediate reactions and actions elicited by a fear-arousing stimulus. Nat Neurosci 2000;3:74–79. [PubMed: 10607398]

57. LeDoux JE, et al. The lateral amygdaloid nucleus: sensory interface of the amygdala in fear conditioning. J Neurosci 1990;10:1062–1069. [PubMed: 2329367]

58. Morgan MM, Ledoux JE. Contribution of ventrolateral prefrontal cortex to the acquisition and extinction of conditioned fear in rats. Neurobiol Learn Mem 1999;72:244–251. [PubMed: 10536101]

59. Petrovich GD, et al. Amygdalo-hypothalamic circuit allows learned cues to override satiety and promote eating. J Neurosci 2002;22:8748–8753. [PubMed: 12351750]

60. Holland PC, Gallagher M. Double dissociation of the effects of lesions of basolateral and central amygdala on conditioned stimulus-potentiated feeding and Pavlovian-instrumental transfer. E J Neurosci 2003;17:1680–1694.

61. McDannald MA, et al. Lesions of orbitofrontal cortex impair rats' differential outcome expectancy learning but not conditioned stimulus-potentiated feeding. J Neurosci 2005;25:4626–4632. [PubMed: 15872110]

62. Stalnaker TA, et al. Basolateral amygdala lesions abolish orbitofrontal-dependent reversal impairments. Neuron 2007;54:51–58. [PubMed: 17408577]

63. Baxter MG, Murray EA. The amygdala and reward. Nat Rev Neurosci 2002;3:563–573. [PubMed: 12094212]
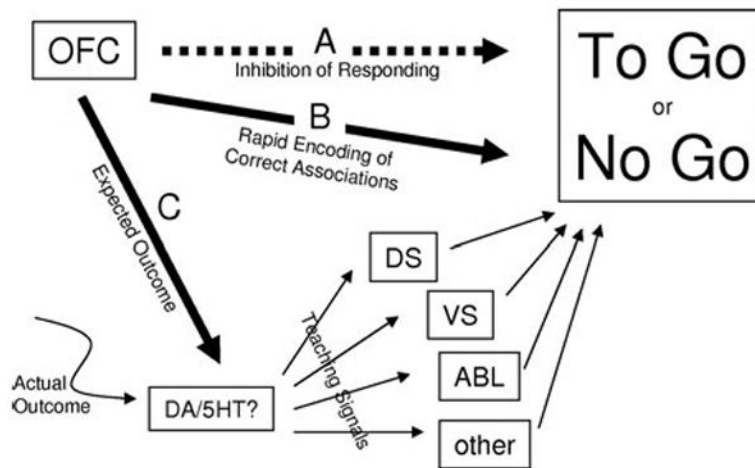
FIGURE 1.
Schematic summarizing proposed roles for the OFC in reversal learning. **(A)** Historical proposal that output from OFC directly inhibits "prepotent" responses and thus is critical for reversal learning. **(B)** More recent idea that OFC acts as a highly flexible associative look-up table to guide correct responding and thus is critical for reversal learning. **(C)** Our proposal that OFC is critical for reversal learning because it facilitates associative learning in other structures, by facilitating the generation of teaching signals when actual outcomes do not match OFC signals of expected outcomes. DS, dorsal striatum; VS, ventral striatum; DA, dopamine; 5-HT, serotonin.
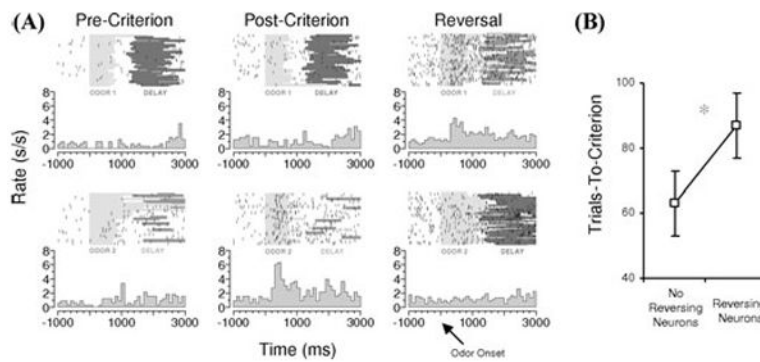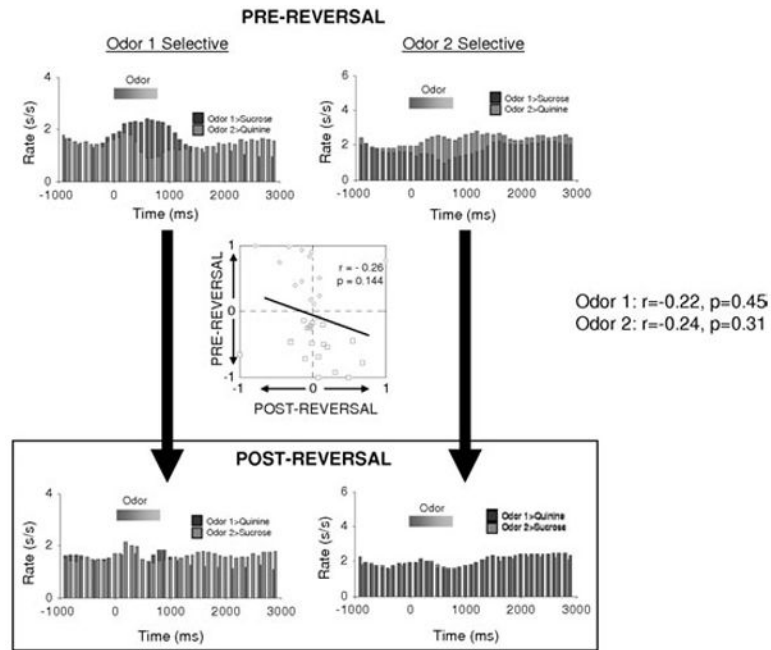
FIGURE 2.
Flexibility of associative encoding in OFC is inversely related to speed of reversal learning.
(A) Example of associative encoding in a rat OFC neuron recorded during acquisition and
reversal of a two-odor discrimination problem. Neural activity is shown synchronized to odor
onset on each trial, in both raster and peri-event time histograms, for trials before learning (pre-
criterion), after learning (post-criterion), and after reversal (reversal). Before reversal, odor 1
predicts sucrose, and odor 2 predicts quinine. The neuron exhibits a phasic response during
odor sampling that develops with learning and tracks the predicted outcome (quinine) across
reversal. **(B)** Average trials required to attain criterion on the reversal in sessions in which
reversing cue-selective neurons were recorded versus sessions in which cue-selective neurons
were recorded that did not reverse. Rats performed significantly worse in sessions with
reversing neurons. (Data adapted from Stalnaker *et al.*[22]) (In color in *Annals* online.)
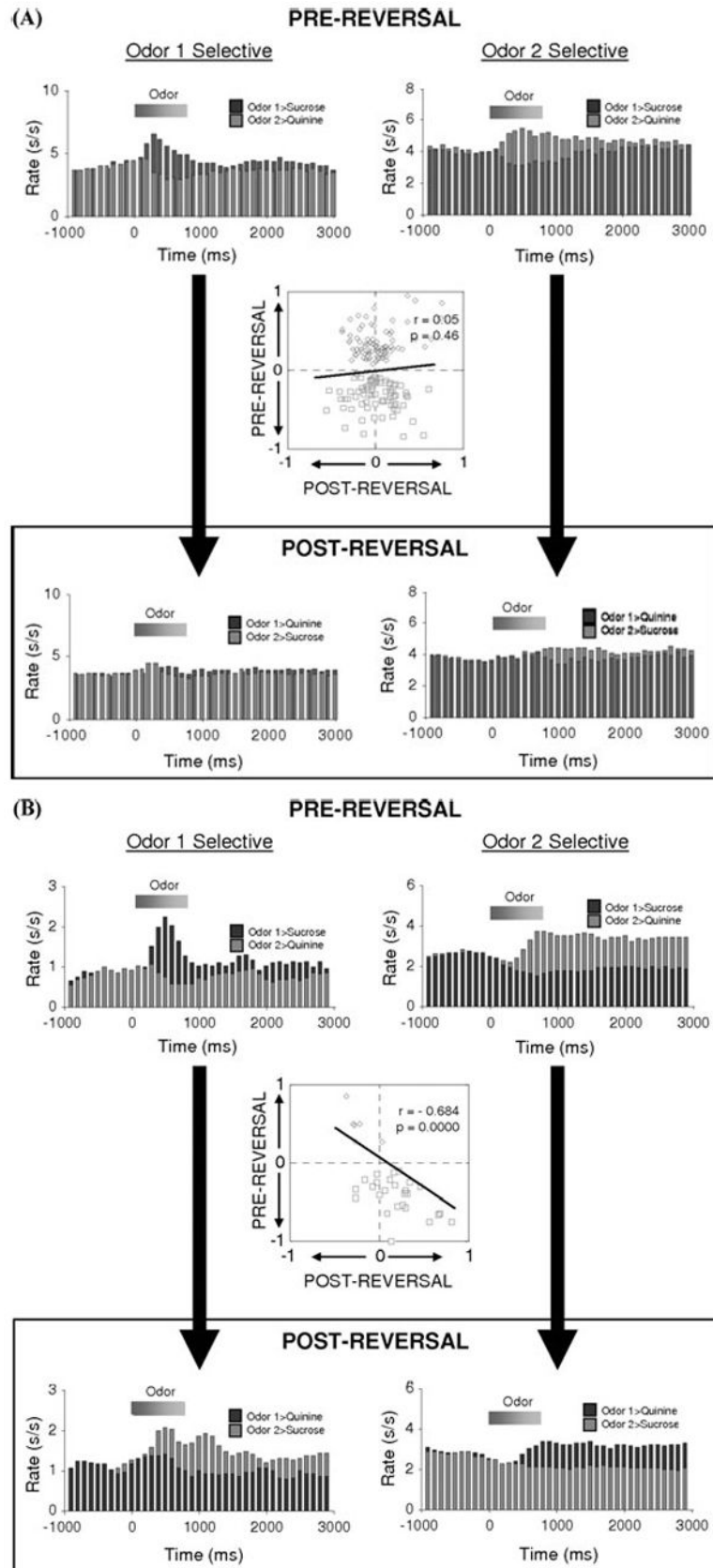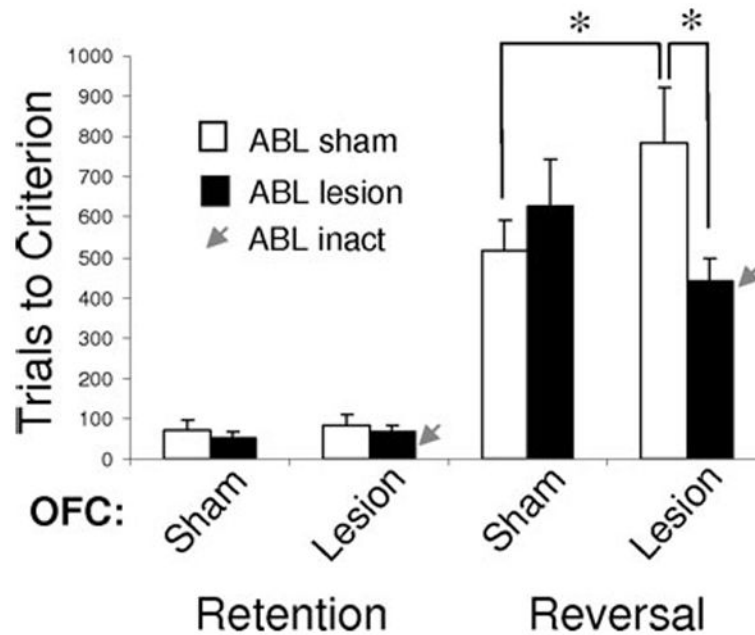
FIGURE 3.
Flexibility of associative encoding is greater in ABL than in OFC. Population response of neurons in OFC **(A)** and ABL **(B)** identified as cue-selective during learning. Average activity per neuron is shown, synchronized to odor onset, during and after reversal. The population response reverses cue-selectivity in ABL but not in OFC. Inset scatterplot compares the cue-selectivity indices before (X-axis) and after reversal (Y-axis) for all the cue-selective neurons used to construct the population histogram. *Blue* and *red symbols* show data for "Odor One Selective" neurons and "Odor Two Selective" neurons, respectively. The cue-selectivity indices are inversely correlated in ABL, consistent the reversal of cue-selectivity, whereas they show no correlation in OFC. (Data adapted from Stalnaker *et al.*.[22,36]) (In color in *Annals* online.)

FIGURE 4.
Flexibility of associative encoding in ABL depends on input from OFC. Population response of cue-selective neurons in ABL in rats with ipsilateral lesions of OFC. Average activity per neuron is shown, synchronized to odor onset, during and after reversal. Unlink the populations recorded in intact rats, illustrated in Figure 3B, the population response recorded in OFC-lesioned rats does not reverse cue-selectivity. Inset scatterplot compares the cue-selectivity indices before (X-axis) and after reversal (Y-axis) for all the cue-selective neurons used to construct the population histograms. *Blue* and *red symbols* show data for "Odor One Selective" neurons and "Odor Two Selective" neurons, respectively. Again in contrast to the inverse correlation in intact rats, illustrated in Figure 3B, the cue-selectivity indices in OFC-lesioned rats showed no correlation, indication that ABL neurons in OFC-lesioned rats had lost their tendency to reverse. (Data adapted from Saddoris *et al.*[34]) (In color in *Annals* online.)

FIGURE 5.
Reversal impairments caused by OFC lesions are abolished by lesions or inactivation of ABL.
Bars show the average number of trials (±SEM) required to retain and reverse a two-odor
discrimination problem for controls, rats with OFC lesions, and rats with OFC lesions
combined with bilateral ABL lesions or unilateral ABL lesions and inactivation of contralateral
ABL via infusions of muscimol + baclofen immediately prior to reversal (*red arrows*). As
expected, OFC lesions impaired reversal learning. This impairment was abolished by pre-
training lesions of ABL or by inactivation at the time of reversal. ABL lesions alone had no
effect. *Significant difference with $P < 0.05$. (Data adapted from Stalnaker *et al.*,[62]) (In color
in *Annals* online.)