Schizophrenia Bulletin vol. 36 no. 3 pp. 465–471, 2010
doi:10.1093/schbul/sbq005
Advance Access publication on February 12, 2010

The Neural Underpinnings of Associative Learning in Health and Psychosis: How Can Performance Be Preserved When Brain Responses Are Abnormal?

Graham K. Murray*[,1–3], **Philip R. Corlett**[1–4], **and Paul C. Fletcher**[1–3]

[1]Brain Mapping Unit; [2]Department of Psychiatry; [3]Behavioural and Clinical Neuroscience Institute, University of Cambridge, Cambridge, UK; [4]Department of Psychiatry, Yale University, New Haven, CT

*To whom correspondence should be addressed; Department of Psychiatry, Box 189, Addenbrooke's Hospital, University of Cambridge, Box 189, Cambridge CB2 0QQ, UK; tel: +44-0-1223-764678, fax: +44-0-1223-764675, e-mail: gm285@cam.ac.uk.

Associative learning experiments in schizophrenia and other psychoses reveal subtle abnormalities in patients' brain responses. These are sometimes accompanied by intact task performance. An important question arises: How can learning occur if the brain system is not functioning normally? Here, we examine a series of possible explanations for this apparent discrepancy: (1) standard brain activation patterns may be present in psychosis but partially obscured by greater noise, (2) brain signals may be more sensitive to real group differences than behavioral measures, and (3) patients may achieve comparable levels of performance to control subjects by employing alternative or compensatory neural strategies. We consider these explanations in relation to data from causal- and reward-learning imaging experiments in first-episode psychosis patients. The findings suggest that a combination of these factors may resolve the question of why performance is sometimes preserved when brain patterns are disrupted.

Key words: reinforcement/causal/dopamine/striatum/adaptation/prediction error

Introduction

The formation of inappropriate associations between stimuli, thoughts, and percepts is increasingly recognized as a possible factor underpinning certain features of mental illnesses,[1,2] particularly the positive symptoms of schizophrenia.[3,4] Studies of associative learning in schizophrenia in the 1950s and 1960s produced mixed results,[5] and for many years this topic fell out of favor. The majority of research into cognition in schizophrenia has concentrated instead on assessing executive control, declarative memory, attention, and problem solving. However, recent years have seen a resurgence of interest in the importance of associative learning in schizophrenia research. This revival may relate to developments in our understanding of the role of dopamine in associative learning in preclinical neuroscience[6] and to increased recognition of the importance of this cognitive domain for survival and environmental adaptation across species.[7] Furthermore, some theories specifically link associative learning to psychotic symptoms.[8–14] While the latter ideas have yet to be fully developed, some patterns are emerging. One notable point is that, in psychosis, there may be both strengths and weaknesses in learning. This in itself raises a number of questions about the nature of successful learning and its neural underpinnings. In this article, we selectively review studies on this topic by our own and other groups and present new analyses of previously published imaging studies in which we consider relationships between brain and behavior during learning.

The Utility of Associative Learning

The British empiricist philosopher David Hume asserted that all our reasoning is based on associations that we form.[15] Associative learning may underpin our ability to represent and recall the causal and predictive structure of our environment. It thus provides us with a powerful means of predicting and therefore perhaps manipulating the impact of our environment upon us.[16]

Experiments that present human volunteers with exposure to contingencies between causes and effects indicate that it is possible to engender the same learning phenomena and biases in human responding upon which formal animal learning theories are predicated. The same cognitive and neural processes that govern learning about appetitive and aversive events (as studied in experimental animals or indeed humans) also contribute to human causal learning and belief formation.[7,17,18] Key among these processes is prediction error, the mismatch between expected and actual experience. Prediction error is signaled by the mesocorticolimbic dopamine and glutamate systems[6,19] and contributes to learning about rewards and punishments[20] and the predictive validity of information.[21] Theories that appeal to these processes in the generation of psychotic symptoms suggest that the neural and cognitive mechanisms that engage new learning are deployed erroneously. This would lead to spurious learning about internal and external events and thereby, ultimately, to delusional beliefs and hallucinations.

Behavioral Studies of Associative Learning in Schizophrenia and Other Psychoses Show Specific Patterns of Strengths and Weakness

When considered as a group, people with psychosis show an intriguing pattern of strengths and weaknesses in associative learning.[13,22–27] While patients are generally impaired, this is not always manifest as a slowed or weakened learning. Rather, there are instances in which patients can learn things faster than nonpsychotic individuals. This can occur in case subjects when prior experience would retard learning in healthy individuals but fails to do so in patients. For example, in latent inhibition,[28] when a healthy individual is preexposed to a stimulus without any consequences, this leads to a retardation of learning when the preexposed stimulus subsequently has important consequences. The effect is thought to reflect an adaptive mechanism through which attention is deflected away from redundant environmental cues and toward more informative stimuli. The attenuation of latent inhibition in patients with psychosis means that they learn faster[29] (though the literature on this is not consistent[30]). A consequence of this rapid acquisition is that patients may attend to and learn about irrelevant stimuli erroneously.[31]

During other feedback-dependent learning tasks, patients learn slightly more slowly (on average) than control subjects. However, many schizophrenia patients can form and maintain the appropriate associations in order to meet learning criteria for behavioral tasks. For example, a recent article documented performance on the intradimensional/extradimensional (IDED) test from the Cambridge Neuropsychological Test Automated Battery on 262 patients from the West London first-episode psychosis study (232 with schizophrenia) and 76 control subjects.[32] Although many studies using this test focus on aspects of attentional set shifting that characterize its latter stages, in fact the task involves participants learning from feedback while passing through various stages of learning. Notably, it has stages of initial simple discrimination learning, several stages of reversal learning (where a previously unreinforced stimulus becomes reinforced), rule abstraction (intradimensional shifting), and shifting attention to a previously irrelevant stimulus dimension (extradimensional set shifting). While patients showed relatively preserved performance on the simple discrimination test (though they did, on average, make significantly more errors than control subjects), there was a more prominent deficit in reversal learning, which requires cognitive flexibility and the ability to learn from negative feedback.

These results are consistent with our own study of 119 first-episode psychosis patients using the IDED test, where we found that psychosis patients made few errors (though statistically more than control subjects) on simple reinforcement (simple discrimination learning) in addition to making more reversal errors.[22] Other studies have found analogous results.[33,34] Interestingly, although patients made only a few errors, preliminary reports suggest that this fairly normal behavior may be accompanied by strikingly different patterns of brain activity when compared with control subjects.[24,26,35] We will now discuss further how (at the brain level) patients with psychosis are able to learn simple associations and what limits there are on the success of this learning.

Neuroimaging Studies of Associative Learning Show Markedly Different Brain Activation Patterns in Schizophrenia and Other Psychoses Compared With Control Subjects

In 2 previous experiments in (mainly) the same set of first-episode psychosis case subjects and control subjects, we examined associative learning in the functional magnetic resonance imaging (fMRI) scanner.[24,26] One study involved casual learning, the other probabilistic reinforcement learning with financial rewards. The studies produced convergent results. Importantly, all the patients had active psychotic symptoms around the time of the experiment. Some were taking antipsychotic medication. In both forms of learning, there was no significant difference in learning rates between patients and control subjects, but there were remarkable differences in brain activation between groups. Both studies revealed learning driven by a mesocorticolimbic network in control subjects but showed that in patients there was an absence of the normal neural distinction between important and unimportant events; there was even some evidence in reward learning that in patients there was a reversal of the normal pattern of brain activation in part of the midbrain. These results are consistent with theories that posit a role for aberrant incentive salience or dysfunctional prediction error learning in the pathogenesis of psychotic symptoms. Furthermore, the greater the magnitude of this dysfunction (during causal learning), the more severe the delusional ideation of the patient. The findings were not secondary to antipsychotic medication, as when the analysis was restricted to a small sample of patients who were not taking medication, the abnormalities remained. Twelve months after presentation, the majority of the patients had received diagnoses of schizophrenia.

A Paradox: Patients Show Abnormal Brain Responses but May Learn Normally

We have seen that a proportion of patients with schizophrenia perform differently from control subjects in some tests of associative learning and that this form of learning may relate to psychotic symptoms. Moreover, the neural circuitry underlying associative learning appears to be substantially altered in psychosis. These points prompt an important residual question: namely, if the neurophysiology of learning is so different in psychosis, how do patients learn successfully?

There are a number of possible answers to this question. Perhaps, the standard patterns of activation that underpin successful learning are present, but the signal-to-noise ratio of the recorded neural responses is lower, leading to significant group differences. One way of addressing this question is to examine patterns of neural activation at a low statistical threshold in patients to see whether a ‘‘normal’’ pattern of activation is present at a lower grade.

A second possible explanation is that overt behavioral responses (which quantify learning) are a cruder measure than that provided by imaging, which is multidimensional. Under more demanding conditions (outside the scanner), patients may begin to fail but, using reasonably simple tasks in the scanner, sensitive neural measures are able to show group differences not detectable in behavioral outcome variables.

A third explanation is that patients may be engaging alternative neural systems to achieve a level of performance comparable to that of control subjects. Such a possibility could be explored by capitalizing on the whole-brain information available with functional neuroimaging. This allows us to look beyond the regions of interest and determine whether patients are engaging brain regions beyond the normal circuitry. That is, we ignore responses in the traditional mesocorticolimbic circuitry, which we know is engaged during prediction error–driven causal learning and reward learning, and focus instead on activity outside of these circuits of interest. Care must be taken here, however. This analysis will not only reveal regions whose activity is compensatory but also regions whose responses may be causing the learning dysfunction, ie, brain areas whose activation interferes with and is deleterious for successful learning. These possibilities can be explored by relating responses in these regions to learning competence.

Below, we consider each of these 3 proposed explanations in more detail, using reexamination of imaging data during reward learning[24] and causal learning[26] to establish evidence supporting them.

Do Patients Activate the Normal Neural Circuitry During Learning, Albeit to a Lesser Degree Than Control Subjects?

We reexamined our data and first found support for the signal-to-noise ratio hypothesis. For example, at a more lenient statistical threshold than we used previously, we found clusters of activity in patients in bilateral ventral striatum and medial prefrontal cortex during reward learning—suggesting that patients were activating these areas, just not as robustly as control subjects (figure 1). It is interesting that these classic learning-related regions are identified at a lower threshold, indicative of less robust activation. Perhaps this modest activity was sufficient for our fairly simple task.

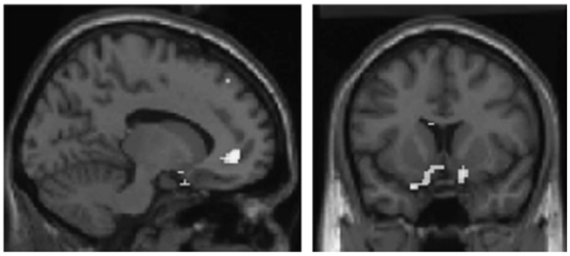

Fig. 1. Reward Prediction Error in Psychosis Patients, Revealing Activation in Bilateral Ventral Striatum and Medial Prefrontal Cortex ($P < .005$ Uncorrected, Minimum Cluster Size 10 Voxels). See online supplementary material for a color version of this figure.

An analogous result was found in the dorsal striatum by Weickert et al[35] using an implicit associative learning task—‘‘the weather prediction’’ task. Here, while patients did activate dorsal striatum in this nonrewarding learning task and their performance did not significantly differ from control subjects, nevertheless, in patients, the dorsal, associative striatum, was significantly less active than in control subjects.

How can patients with reduced or noisy activity responses show preserved learning? One might ask: Does the degree of brain activation matter for behavior? Initial evidence from studies of healthy humans suggests that reinforcement learning prediction error signal strength in the dorsal striatum does indeed distinguish

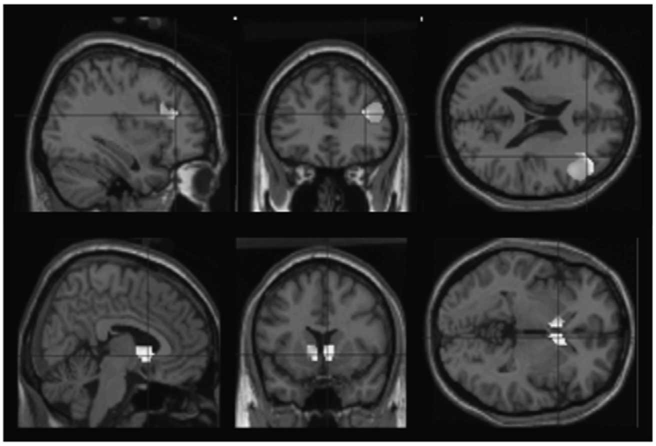

Fig. 2. During Causal Learning, Better Performing Patients More Strongly Activate the A Priori Network of Interest of Head of Caudate and Right Prefrontal Cortex Compared With Worse Performing Patients ($P < .005$ Uncorrected). See online supplementary material for a color version of this figure.

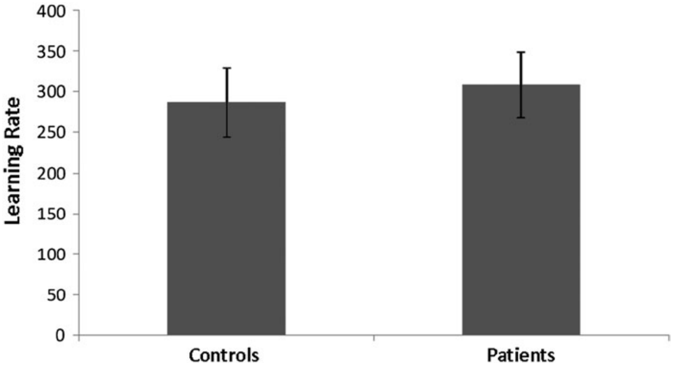

Fig. 3. Equivalent Learning Rates in First-Episode Psychosis Patients and Control Subjects During Causal Learning. Error bars represent SEs.

better from worse learners during probabilistic learning. For example, Schonberg et al[36] examined reward-based decision making in a sample of more than 30 healthy control subjects and considered the relationship between learning performance and brain activation. They found that striatal prediction error signals during learning differentiated learners from nonlearners and that, across subjects, the magnitude of prediction error signals in the dorsal striatum correlated significantly with behavioral performance. We further explored the role of "signal" and "noise" in the effects we observed, by comparing brain responses during causal learning in the system of interest in patients who learned well with responses in those subjects who were poor learners. We noted that the better learning patients engaged frontal cortex and dorsal striatum (head of caudate) to a greater extent than did poorer learners (figure 2). In this respect, a closer look at the data suggests that patients do show measurable activations and that these activations, being smaller, may be sufficient only to sustain weaker levels of behavioral performance.

Are Brain Signals More Sensitive to Group Differences Than Behavioral Learning Measures?

There is some evidence in favor of this explanation. In the causal learning task, the nonsignificant difference in learning measures between patients and control subjects (figure 3) is consistent with the notion that fMRI measures may indeed represent a more sensitive multidimensional assay of learning. Indeed, in other work, we have exploited this sensitivity to adjudicate between competing mechanistic accounts of causal learning.[17] In our reward learning task,[24] on closer inspection, there was a trend for control subjects to learn the reward task better than patients (with a mean of 80% correct choices as opposed to 67% in patients), but this difference was not statistically significant (figure 4, left panel). Furthermore, if reaction times (as opposed to choice behavior) were used as an index of learning, there were indeed statistically significant differences between case subjects and control subjects (figure 4, right panel). Thus, when viewed across both experiments, behavioral results in these tasks are less sensitive than imaging results at differentiating diagnostic groups, but they do reveal some evidence suggestive of group differences.

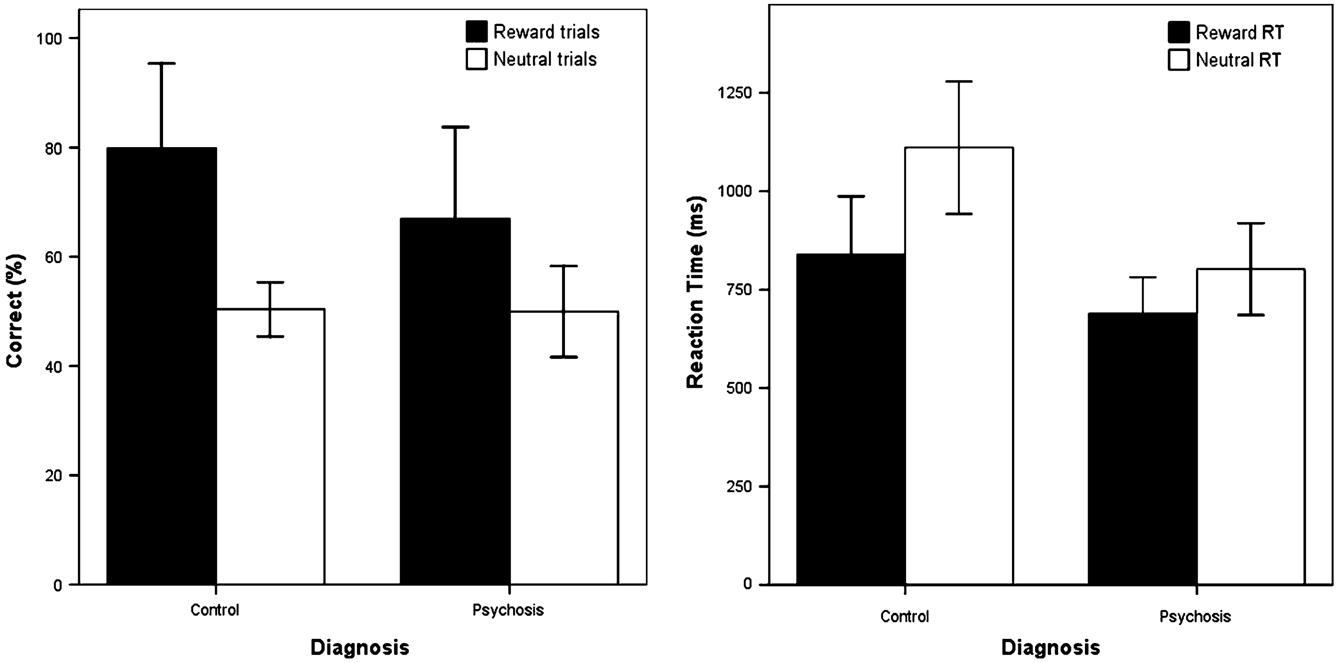

Fig. 4. Behavioral Results for Reward Learning. Left panel shows that although control subjects made more correct choices than patients, this difference was not statistically significant. Right panel shows an adaptive reinforcement-related speeding effect in control subjects (faster responses on reward trials) and a significant attenuation of this effect in patients. Moreover, patients were significantly faster than control subjects on the irrelevant, neutral condition. Error bars represent 95% confidence intervals.

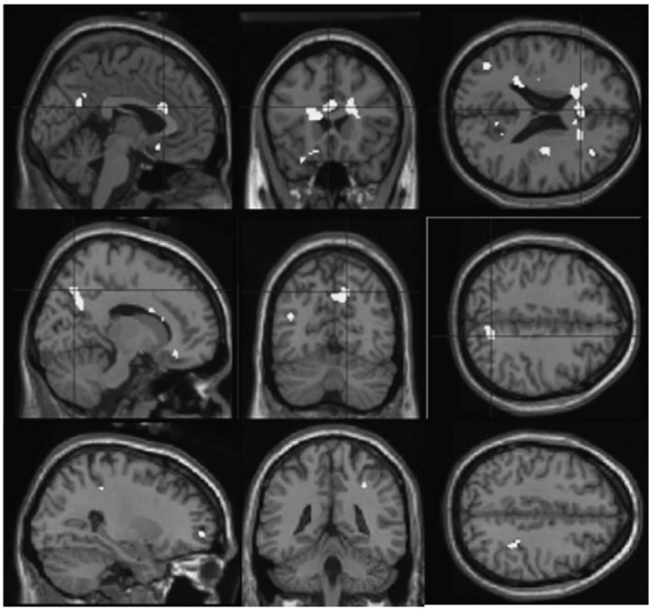

Fig. 5. Regions in Which Better Performing Patients Activate More Strongly Than Worse Performing Patients, Masking Out Areas Activated by Control Subjects, So Indicating Patient Specificity. Upper and middle panels show causal learning; lower panels show reward learning ($P < .005$). The contrast identified regions that differentiate better from worse performing patients more than they differentiate better from worse performing control subjects, masked inclusively by better learning patients greater than worse performing patients and masked exclusively by better learning control subjects greater than worse learning control subjects. See online supplementary material for a color version of this figure.

Do Patients Learn Using Alternative Strategies?

Learning in patients may not be so much impaired as different, at least in a proportion of patients. Here, neuroimaging offers the unique opportunity of building up an overall picture of not just how patients fail in the task but how they succeed. It is possible, eg, that patients may engage additional or alternative neural strategies in order to achieve behavioral success. If we look at those patients who are successful (ie, perhaps more likely to be applying compensatory or alternative mechanisms successfully), we can tell whether good performance in patients is upheld by differing neural systems to those responsible for good performance in control subjects. If patients engage in additional/alternative neural activity, and if this extra activity is associated with preserved performance, then we could infer that the extra neural activation represents a strategic or compensatory change. Thus, as above, this might explain why patients show (relative) failure of activation in the traditional neural system in the face of apparently preserved performance.

We examined whether the high-performing patients (defined using a median split) specifically activated extra regions making the assumption that a criterion for identifying such compensatory activity would be that it would involve regions that were active neither in the control subjects (where compensatory activity was not required) nor in the low-performing patients (where failing performance suggests that compensatory activity is not occurring or is less prevalent). Patients who were better learners did not differ from worse learners in symptomatology but did have higher estimated premorbid IQ. In our causal learning data,[26] we identified regions (outside of our a priori circuit of interest) that were more active in competent learning patients (compared with their poor learning groupmates) and furthermore that were not engaged preferentially by better learning compared with worse learning control subjects. This combination of contrasts enables us to rule out activity that is generically related to better performance, identifying areas specifically related to performance boost underpinned by an ''extra'' or compensatory activation. We identified significant foci in the parietal lobe and the anterior cingulate cortex (see figure 5, upper and middle panels). We found that during reward learning,[24] the same analytical approach revealed clusters in visual cortex, frontal pole, and right parietal lobe (figure 5, lower panel).

General Conclusion

Given the relative paucity of experimental studies of associative learning in schizophrenia and taking into account the increasing theoretical importance of this field, we believe that considerable further work is required. Schizophrenia patients demonstrate surprisingly intact learning in some aspects of associative learning and impaired learning elsewhere (especially in cognitive flexibility), but patterns of abnormal brain activity often emerge in response to tasks where behavioral performance does not differ from control subjects. We have put forward 3 explanations for how patients may appear to perform well in some learning tasks in spite of aberrant brain activation. Reexamining data more closely, we found partial experimental support for the 3 suggested possibilities. First, patients do show some normative brain activation at a lower threshold than control subjects, indicating a partially intact neural learning system. This is unsurprising as associative learning is so crucial and so basic a form of learning that a completely broken system may be almost incompatible with survival. Second, there is some evidence that during simple forms of learning, the brain provides a more sensitive index of learning than behavior does, with the behavioral abnormality elucidated only at a more relaxed statistical threshold. Finally, there is also some evidence that successful learning in patients is upheld by compensatory mechanisms, ie, they may engage different strategies from control subjects. Thus, successful learning in patients in reward and causal learning paradigms may be driven by activation in the frontal pole, anterior cingulate cortex, visual cortex, and parietal cortex. Ultimately, close examinations of the relationship between brain response and learning performance are likely to provide richer insights to psychosis than are provided by exploring either measure in isolation.

Supplementary Material

Supplementary material is available at http: //schizophreniabulletin.oxfordjournals.org.

Funding

Medical Research Council (Clinician Scientist Award G0701911 to G. K. M.); NARSAD (Young Investigator award to G. K. M.); NARSAD (Young Investigator award to P. R. C.); Medical Research Council and Wellcome Trust (joint award to University of Cambridge, Behavioural and Clinical Neuroscience Institute, G0001354); University of Cambridge (Parke Davis Exchange Fellowship to P. R. C.); Wellcome Trust (Senior Research Fellowship in Clinical Science award 064351 to P. C. F.); Bernard Wolfe Professorial award in Health Neuroscience to P. C. F.

References

1. Pavlov IP. *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex.* New York, NY: Dover Publications Inc; 1927.
2. Locke J. *An Essay Concerning Human Understanding.* London: Printed by E. Holt, For T. Basset; 1690.
3. Bleuler E. *Dementia Praecox or the Group of Schizophrenias.* New York, NY: International University Press; 1911/1950.
4. Schneider K. *Clinical Psychopathology.* New York, NY: Grune and Stratton; 1959.
5. Buss AH, Lang PJ. Psychological deficit in schizophrenia: I. Affect, reinforcement, and concept attainment. *J Abnorm Psychol*. 1965;70:2–24.
6. Schultz W, Dickinson A. Neuronal coding of prediction errors. *Annu Rev Neurosci*. 2000;23:473–500.
7. Dickinson A. The 28th Bartlett Memorial Lecture. Causal learning: an associative analysis. *Q J Exp Psychol B*. 2001;54:3–25.
8. Gray JA, Feldon J, Rawlins JNP, Smith AD. The neuropsychology of schizophrenia. *Behav Brain Sci*. 1991;14:1–19.
9. Miller R. Schizophrenic psychology, associative learning and the role of forebrain dopamine. *Med Hypotheses*. 1976; 2:203–211.
10. Corlett PR, Honey GD, Fletcher PC. From prediction error to psychosis: ketamine as a pharmacological model of delusions. *J Psychopharmacol*. 2007;21:238–252.
11. Kapur S. Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. *Am J Psychiatry*. 2003;160:13–23.
12. Heinz A, Schmidt LG, Reischies FM. Anhedonia in schizophrenic, depressed, or alcohol-dependent patients–neurobiological correlates. *Pharmacopsychiatry*. 1994;27(suppl 1):7–10.
13. Murray GK, Fletcher PC. Can models of reinforcement learning help us to understand symtpoms of schizophrenia? In: Dreher J-C, Tremblay L, eds. *Handbook of Reward and Decision-Making.* Oxford, UK: Elsevier; 2009:251–269.
14. Murray G. Dopamine dysfunction and delusions, hallucinations and anhedonia. *Eur Psychiatr Rev*. 2009;2:21–23.
15. Hume D. *An Enquiry Concerning Human Understanding.* London: A. Millar; 1748.
16. Dickinson A. Conditioning and associative learning. *Br Med Bull*. 1981;37:165–168.
17. Corlett PR, Aitken MR, Dickinson A, et al. Prediction error during retrospective revaluation of causal associations in humans: fMRI evidence in favor of an associative model of learning. *Neuron*. 2004;44:877–888.
18. Aitken MR, Larkin MJ, Dickinson A. Super-learning of causal judgements. *Q J Exp Psychol B*. 2000;53:59–81.
19. Lavin A, Nogueira L, Lapish CC, Wightman RM, Phillips PE, Seamans JK. Mesocortical dopamine neurons operate in distinct temporal domains using multimodal signaling. *J Neurosci*. 2005;25:5013–5023.
20. Matsumoto M, Hikosaka O. Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature*. 2009;459:837–841.
21. Bromberg-Martin ES, Hikosaka O. Midbrain dopamine neurons signal preference for advance information about upcoming rewards. *Neuron*. 2009;63:119–126.
22. Murray GK, Cheng F, Clark L, et al. Reinforcement and reversal learning in first-episode psychosis. *Schizophr Bull*. 2008; 34:848–855.
23. Murray GK, Clark L, Corlett PR, et al. Incentive motivation in first-episode psychosis: a behavioural study. *BMC Psychiatry*. 2008;8:34.
24. Murray GK, Corlett PR, Clark L, et al. Substantia nigra/ventral tegmental reward prediction error disruption in psychosis. *Mol Psychiatry*. 2008;13:239–267–276.
25. Gray JA. Integrating schizophrenia. *Schizophr Bull*. 1998;24: 249–266.
26. Corlett PR, Murray GK, Honey GD, et al. Disrupted prediction-error signal in psychosis: evidence for an associative account of delusions. *Brain*. 2007;130(pt 9):2387–2400.
27. Waltz JA, Frank MJ, Robinson BM, Gold JM. Selective reinforcement learning deficits in schizophrenia support predictions from computational models of striatal-cortical dysfunction. *Biol Psychiatry*. 2007;62:756–764.
28. Lubow RE. Latent inhibition: effects of frequency of nonreinforced preexposure of the CS. *J Comp Physiol Psychol*. 1965; 60:454–457.
29. Martins Serra A, Jones SH, Toone B, Gray JA. Impaired associative learning in chronic schizophrenics and their first-degree relatives: a study of latent inhibition and the Kamin blocking effect. *Schizophr Res*. 2001;48:273–289.
30. Rascle C, Mazas O, Vaiva G, et al. Clinical features of latent inhibition in schizophrenia. *Schizophr Res*. 2001;51:149–161.
31. Lubow RE, Kaplan O. The visual search analogue of latent inhibition: implications for theories of irrelevant stimulus processing in normal and schizophrenic groups. *Psychon Bull Rev*. 2005;12:224–243.
32. Leeson VC, Robbins TW, Matheson E, et al. Discrimination learning, reversal, and set-shifting in first-episode schizophrenia: stability over six years and specific associations with medication type and disorganization syndrome. *Biol Psychiatry*. 2009;66:586–593.

33. McKirdy J, Sussmann JE, Hall J, Lawrie SM, Johnstone EC, McIntosh AM. Set shifting and reversal learning in patients with bipolar disorder or schizophrenia. *Psychol Med.* 2009;39:1289–1293.
34. Waltz JA, Gold JM. Probabilistic reversal learning impairments in schizophrenia: further evidence of orbitofrontal dysfunction. *Schizophr Res.* 2007;93:296–303.
35. Weickert TW, Goldberg TE, Callicott JH, et al. Neural correlates of probabilistic category learning in patients with schizophrenia. *J Neurosci.* 2009;29:1244–1254.
36. Schonberg T, Daw ND, Joel D, O'Doherty JP. Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. *J Neurosci.* 2007;27:12860–12867.