frontiers in
HUMAN NEUROSCIENCE

HYPOTHESIS AND THEORY ARTICLE
published: 10 July 2009
doi: 10.3389/neuro.09.012.2009

Why do delusions persist?

Philip R. Corlett[1,2]*, John H. Krystal[1], Jane R. Taylor[1] and Paul C. Fletcher[2]

[1] *Department of Psychiatry, School of Medicine, Yale University, New Haven, CT, USA*
[2] *Brain Mapping Unit, Department of Psychiatry, Behavioural and Clinical Neurosciences Institute, School of Clinical Medicine, University of Cambridge, Cambridge, UK*

Edited by:
Neal J. Cohen, University of Illinois, USA

Reviewed by:
Anthony Wagner, Stanford University, USA
Stephan Heckers, Vanderbilt University, USA

****Correspondence:***
Philip R. Corlett, Department of Psychiatry, Abraham Ribicoff Research Facility, Connecticut Mental Health Centre, Yale University School of Medicine, 34 Park Street, New Haven, CT 06519, USA.
e-mail: philip.corlett@yale.edu

Delusions are bizarre and distressing beliefs that characterize certain mental illnesses. They arise without clear reasons and are remarkably persistent. Recent models of delusions, drawing on a neuroscientific understanding of learning, focus on how delusions might emerge from abnormal experience. We believe that these models can be extended to help us understand why delusions persist. We consider prediction error, the mismatch between expectancy and experience, to be central. Surprising events demand a change in our expectancies. This involves making what we have learned labile, updating and binding the memory anew: a process of memory reconsolidation. We argue that, under the influence of excessive prediction error, delusional beliefs are repeatedly reconsolidated, strengthening them so that they persist, apparently impervious to contradiction.

Keywords: salience, delusions, prediction error, extinction, habit, reconsolidation

INTRODUCTION

In a recent and highly influential paper (Kapur, 2003), Kapur related the aberrant experiences characteristic of early psychosis to inappropriate dopamine signaling by appealing to the concept of motivational salience (Berridge and Robinson, 1998), a quality bestowed on objects or events by virtue of their behavioral relevance. Salient stimuli grab attention, drive action and influence goal-directed behavior; processes that are usurped by drug related stimuli in addiction (Berridge and Robinson, 1998). Due to dopamine dysregulation, psychotic individuals attribute salience inappropriately to stimuli, thoughts and percepts. The world of an individual with emerging psychosis is strange and sinister, pregnant with new meaning. These experiences can crystallize into bizarre causal attributions and beliefs: delusions. Under the influence of antipsychotic medication, aberrant salience is dampened, patients are less perturbed by their experiences and symptom resolution begins (Kapur, 2003). That is, the treatment alters the experiences such that the delusion can extinguish.

Kapur's ideas relate to a number of previous suggestions that key symptoms of schizophrenia emerge from abnormal attention and learning processes (Corlett et al., 2007a). This re-consideration of symptoms in the context of a contemporary model for the role of dopamine has had a big impact on cognitive neuroscientists studying psychosis and supportive evidence has already emerged (Corlett et al., 2007b, Murray et al., 2008). We believe that it can be taken further and embedded more deeply in the principles of formal animal learning theory. In doing so, we provide an account for features of delusions that are not readily explicable in terms of motivational salience alone: their persistence, which is often as striking as their bizarreness.

The salience of an event can be parsed into several components (Horvitz, 2002), some of which are represented by dissociable neural substrates (El-Amamy and Holland, 2007). Salience can relate to attention and motivation (Redgrave and Gurney, 2006). Moreover, it is apparent that prediction error, a fundamental parameter in associative learning models, is a driving force in salience attribution, an observation that is reflected in a growing literature within computational modeling (Smith et al., 2006) and empirical investigation (Menon et al., 2007). Prediction error represents the mismatch between what we expect in a given situation and what we actually experience. By working to reduce this mismatch, we improve our understanding of the causal structure of the world (Dickinson, 2001). That is, a prediction error is a signal that our understanding of, or belief about, the world must be updated. Furthermore, those stimuli that engender prediction errors become more salient and this will be reflected in greater allocation of attention when they next occur (Schultz and Dickinson, 2000).

Prediction error theories of delusion formation suggest that under the influence of inappropriate prediction error signal, possibly as a consequence of dopamine dysregulation, events that are insignificant and merely coincident seem to demand attention, feel important and relate to each other in meaningful ways. Delusions ultimately arise as a means of explaining these odd experiences (Kapur, 2003; Maher, 1974). The insight relief gained by arriving at an explanatory scheme leads to strong consolidation of the scheme in memory (**Figure 1**).

In support of this view, aberrant prediction error signals during learning in patients with first-episode psychosis have been confirmed experimentally (Corlett et al., 2007b, Murray et al., 2008). Furthermore, the magnitude of aberrant prediction error signal correlated with delusion severity across a group of patients with first-episode psychosis (Corlett et al., 2007b). However, there are important characteristics of delusions that still demand explanation: notably their persistence. Normal associations can extinguish if they prove erroneous, normal beliefs can be challenged and modified. But delusions are noteworthy for the fact that they remain even in the absence of support and in the face of strong contradictory evidence. We believe that this striking clinical phenomenon can be

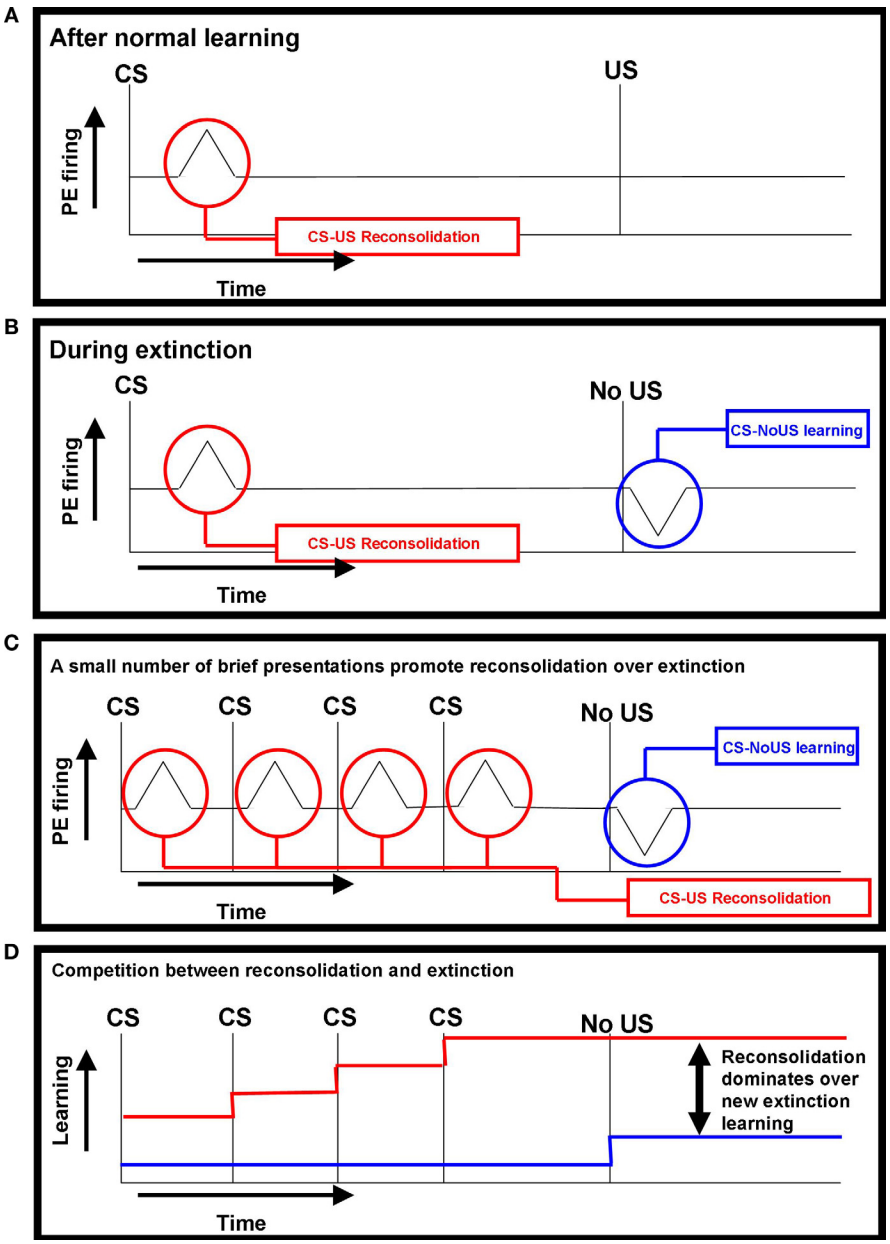

FIGURE 1 | Prediction error responses during learning, extinction and reconsolidation. In these schematics, an upward deflection from the baseline reflects a positive prediction error and a downward deflection represents a negative prediction error (PE). **(A)** After normal learning – Dopamine neurons come to respond to the conditioned stimulus, the best predictor of the salient outcome (Schultz and Dickinson, 2000). When the CS is presented, the organism is reminded of the salient event (represented in red). **(B)** During extinction – When the salient event does not occur, a negative prediction error promotes new extinction learning: the formation of a competing CS–NoUs association (represented in blue) (Bouton, 2000; Eisenhardt and Menzel, 2007). **(C)** A small number of brief presentations promote reconsolidation over extinction – According to the IRH, brief presentations of the conditioned stimulus serve as a reminder and promote reconsolidation of the CS–US association, which overrides any new extinction learning that occurs when the salient event does not occur (Eisenhardt and Menzel, 2007). **(D)** Competition between reconsolidation and extinction – This schematic shows the competition between representations of the CS–US association (strengthened by reconsolidation) and the CS–NoUS association (formed during extinction). With brief reminder presentations, reconsolidation dominates over extinction.

explained within the same framework by considering key findings from the animal learning literature, a literature that has been formerly invoked to explain chronic relapse to drug abuse; extinction (Crombag et al., 2008) and reconsolidation (Lee et al., 2005).

THE PERSISTENCE AND ELASTICITY OF DELUSIONS

If delusion formation may be explained in terms of associative learning then perhaps extinction may represent the process through which delusions are resolved (Miller, 1989). Extinction involves

a decline in responding to a stimulus that has previously been a consistent predictor of a salient outcome (Lovibond, 2004).

Prediction error is also central to extinction (Pedreira et al., 2004). A recent account (Redish et al., 2007) suggests that negative prediction error (a reduction in baseline firing rate of prediction error coding neurons) leads the organism to categorize the extinction situation as different from the original, reinforced, situation and it now learns not expect the salient event in that situation. This learning focuses on contextual cues, allowing the animal to distinguish the newly non-reinforced context from the old, reinforced one. Extinction does not involve unlearning of the original association, but rather the formation a new association between the absence of reinforcement and the extinction situation (Pavlov, 1927). Extinction experiences (the absence of expected reinforcement) invoke an inhibitory learning process which eventually overrides the original cue response in midbrain dopamine neurons (Pan et al., 2008). Individuals with psychosis do not learn well from these absent but expected events (Waltz et al., 2007), nor do they consolidate the learning that does occur (Holt et al., 2009).

But there is more to delusion maintenance than persistence in the absence of supportive evidence: delusions persist even when there is evidence that directly contradicts them. When confronted with counterfactual evidence, deluded individuals do not simply disregard the information. Rather, they may make further erroneous extrapolations and even incorporate the contradictory information into their belief (Joseph, 1986). So, while delusions are fixed, they are also elastic and may incorporate new information without shifting their fundamental perspective. We now go on to consider these characteristics with respect to the internal reinforcement hypothesis (IRH) of extinction learning and memory reconsolidation (Eisenhardt and Menzel, 2007).

MEMORY AND BELIEF

The traditional view of memory as a store that consolidates over time has been challenged by demonstrations that recall or reactivation of a memory briefly restores it into a labile state, rendering it susceptible to interference by amnestic agents. The process – through which memories are recalled, become labile, are combined with new information and finally consolidated once more – has been termed reconsolidation (Misanin et al., 1968; Nader et al., 2000).

This phenomenon suggests a more dynamic aspect to memory function, one perhaps geared toward dealing with incumbent information rather than merely retrospection. Furthermore, reconsolidation has been argued to result in the enhancement of salient memories (Tronson and Taylor, 2007), a claim supported by empirical data. (Lee, 2008; Tronson et al., 2006). Since beliefs must overlap with memories (Eichenbaum and Bodkin, 2000), what might we learn from the brain bases of memory formation and maintenance that will enhance our understanding of delusions?

The IRH contends that extinction and reconsolidation are two parallel processes and that an organism's behavior is based upon the balance between the strength of their two memory representations following memory retrieval. This balance may be modulated by prediction error (Eisenhardt and Menzel, 2007). During conditioning, an organism learns a relationship between a previously neutral stimulus (CS; e.g., a tone) and a reinforcer (food or electric shock, US). In a subsequent training session, an unreinforced exposure to the CS (presenting the tone in the absence of the food reward) might, besides inducing extinction learning, remind the animal about the reinforced situation (when food followed the tone). This kind of reminder might itself lead to new learning and new consolidation, i.e., reconsolidation (Eisenhardt and Menzel, 2007). This reconsolidation process appears to be preferentially engaged in situations when updating occurs, that is, when additional information needs to be incorporated into memory (Morris et al., 2006; Pedreira et al., 2004).

The midbrain dopamine neuron response to conditioned stimuli (Schultz, 1998) might drive reminder learning and reconsolidation of the CS–US relationship. In parallel, the surprising absence of a US engages extinction, through which a new CS–noUS association is formed and consolidated. Which of these representations controls behavior depends on the temporal characteristics of the learning situation (Pedreira et al., 2004). If the CS is presented exactly as it was during training, but no reinforcer is presented, then the organisms' expectancies are definitively disconfirmed and extinction dominates. However, if the CS is presented, say, more briefly than previously, the organism's expectancies are neither definitively confirmed nor completely violated. In this case, reconsolidation of the CS–US association dominates, driven by the midbrain dopamine neuron response which reminds the organism of the reinforced situation (Eisenhardt and Menzel, 2007).

INTERNAL REINFORCEMENT HYPOTHESIS AND DELUSIONS

We have conceptualized delusions as a network of associations formed by aberrant prediction error signals (Corlett et al., 2007a). Given the features of the IRH, perhaps we might interpret the persistence of delusions, even in the absence of supportive evidence and the presence of contradictory evidence, by positing inappropriate activity in the midbrain reminder system. That is, aberrant prediction errors might re-evoke the representation of the delusion without definitively disconfirming it. This would drive preferential reconsolidation over and above any new extinction learning (**Figure 2**). The net effect would be a strengthening of the delusion through reconsolidation rather than a weakening by extinction.

This hypothesis derives support from the phenomenology of delusional beliefs; with medication, delusions resolve gradually (Stanton and David, 2000), and, during that resolution patients describe an intermediate stage, a duality of belief and disbelief ["*a part of me wants to dispel the delusions whilst a part of me is frightened and resists*" (Ruocchio, 1991)].

During this phase, patients may describe how much less occupied they are by their delusions. They become less salient (Kapur, 2004). This transition would be aided by medications which prevent aberrant prediction error firing. The absence of aberrant error signals would gradually shift the balance in favor of the extinction of delusion-related material (**Figure 2**). The duality of patients' beliefs is consistent with our suggestion that multiple representations (in the simplest sense endorsing the belief and not endorsing it) compete for expression in behavior. With medication, the extinction trace (not endorsing the delusion) comes to win that competition. Some more chronically ill patients appear to show another kind of double-awareness, known as double-bookkeeping

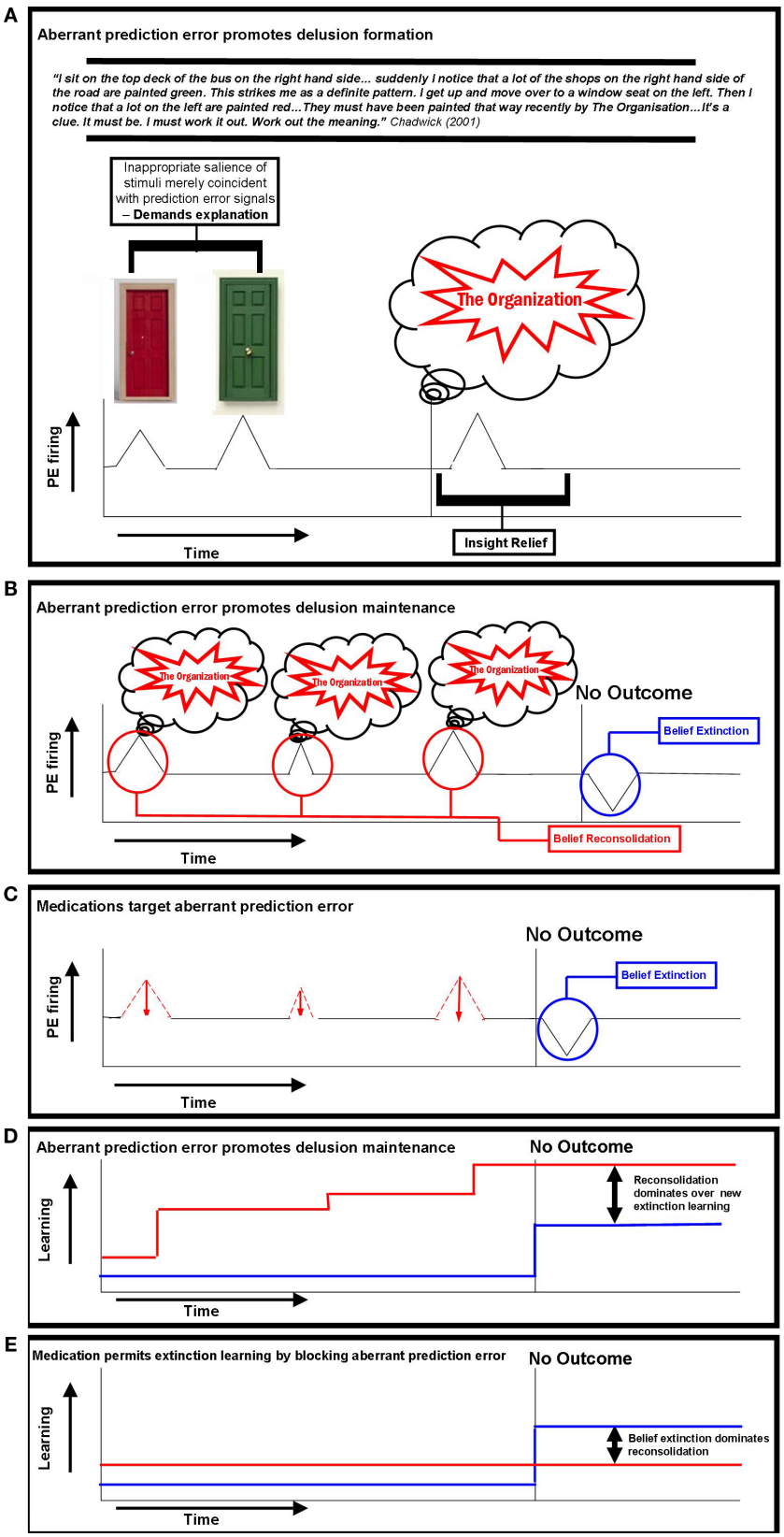

FIGURE 2 | See caption opposite.

(Sass, 2004). Here, patients appear to regain a modicum of insight: delusions persist but the patient does not consistently act upon them. This may be because chronically ill patients experience severe negative symptoms including avolition and emotional blunting, with the consequence that delusional beliefs cannot maintain their previously high degree of salience. Thus, they no longer feel the compulsion to act or ruminate upon the beliefs, though the memory of how salient the belief used to be is sufficient to maintain it (Sass, 2004). We argue, after Alberini (2007), that this salience is sufficient to ensure that the belief is not extinguished.

With medication, the memory trace mediating belief in the delusion is not erased. It is overshadowed by extinction learning. Hence, if medications are ceased, the delusion returns (Chadwick, 2001). Moreover, this model could explain why the contents of patients' delusions persist across consecutive psychotic episodes (Sinha and Chaturvedi, 1989): since the association persists, albeit in attenuated form, recurrence of the neurochemical disturbance that provoked it in the first place will now allow it to re-emerge. State dependency may provide further explanatory insight into the re-emergence of delusional beliefs when antipsychotic drugs are stopped. That is, when the neurochemical sequelae in which a memory was formed are reinstated, the probability and richness with which that memory is recalled are enhanced (Goodwin et al., 1969). While most experimental investigations of state-dependency have focused on administration of compounds that alter cognitive and mood states, the endogenous altered states of consciousness that attend mental illness have been suggested to engender state-dependent recall and therefore alter prognosis (Reus et al., 1979). Finally, extinguished behaviors and delusions spontaneously recover. Even after extinction, the passage of time and the presence of interoceptive or exteroceptive contextual cues can cause the old behavior or belief to re-occur (Bouton, 2000; see **Box 1**).

Applying the IRH to delusions suggests that reminding patients of their delusions in a therapeutic setting, by trying to disconfirm their beliefs might, paradoxically, lead to an increase in the conviction of the belief – if that reminder treatment occurred whilst patients were experiencing inappropriate prediction error firing (see **Figure 2**). There is some evidence that this may indeed be the case: confronting deluded patients with contradictory evidence does indeed strengthen delusional beliefs (Milton et al., 1978). Indeed, as we outline above, they may even incorporate challenging information into their belief (Simpson and Done, 2002). Aberrant prediction error driven reconsolidation might account for the voracity with which patients incorporate ostensibly unrelated information into their delusional scheme – since aberrant prediction error signals remind the patient of their delusion and render its representation highly associable, the subsequent reconsolidation process incorporates the new information, maintaining and strengthening the delusion rather than extinguishing it (**Figure 2**).

Memory reconsolidation may underpin the transition from a salient episodic experience to a habitual belief about the world (Alberini, 2007). The neurobiological systems responsible for coding the salience of an event are intimately related to those associated with learning and memory, such that modulations of those systems with pharmacological intervention can engender inappropriate salience attribution (Corlett et al., 2006). Such interventions can be substituted for training trials in experimental animals, and can strengthen the memory trace to the same extent as it is strengthened by actual repeated experience (Alberini, 2007). That is, pharmacological interventions which enhance the salience of certain associations will reinforce or strengthen those associations as if they had been repeatedly experienced. The effect of pharmacological interventions on the processing of reactivated memories depends on how the memory is reactivated (Lee et al., 2006; see **Box 1**). As such, novel therapeutic approaches may well involve bypassing the receptor level pharmacology of synaptic plasticity and instead targeting alternative mechanisms of memory maintenance (see **Box 2**). It is possible that reactivation of a memory trace for a salient event increases the stabilization of that trace. The most salient memories would be reactivated most frequently and would therefore engender the most reconsolidation based stabilization, increasing their fixity. The net effect would be that a salient experience more rapidly and profoundly updates knowledge and perhaps is more likely to become enshrined in a belief.

A belief, once formed, can be conceived as an habitual way of looking at the world and of interpreting incoming sensory information (see **Box 3**). It is, in other terms, a schema. Perhaps reconsolidation is a crucial component of building such schemata. The transition from salient experience to belief habit in psychotic states occurs rapidly due to excessive prediction error firing (Corlett et al., 2007a) inducing repeated recall of delusion-related material and reconsolidation of that material, updated with new information, forming a "schema-like" representation through which incumbent information is then

FIGURE 2 | Prediction error firing during delusion formation, maintenance and resolution. (A) Aberrant prediction error promotes delusion formation – Aberrant, internally generated prediction error signals imbue mere coincidences (the colors of doors and the sides of the street) with significance which demands explanation (Chadwick, 2001). When an explanation is arrived at, the insight relief may engender dopamine firing which further stamps in the association between the odd experiences and the Organization (Miller, 1993). **(B)** Aberrant prediction error promotes delusion maintenance – Once the explanatory scheme has been developed, it is invoked any time an aberrant prediction error occurs. These reminders, serve to reconsolidate the belief that odd experiences are a result of the malevolent intervention of The Organization. This reconsolidation overcomes any extinction learning that may occur when no harm comes to the sufferer from The Organization or when others try to convince the sufferer that The Organization is not ministrating against them. **(C)** Medications target aberrant prediction error – Antipsychotic medications target aberrant prediction error signals, reminders do not occur and as such reconsolidation is blocked. New extinction learning is allowed to prevail and the individual recovers from their delusion. The mechanism for this targeting of prediction error induced reconsolidation may be at the level of neurotransmission in the synapse (e.g., Aripiprazole blocking phasic dopamine signaling (Hamamura et al., 2007) or it may be intracellular, targeting the way in which the experience-explanation belief is stored epigenetically (Bredy and Barad, 2008). **(D)** Aberrant prediction error promotes delusion maintenance – Prediction errors modulate the representations that govern behavior. Under aberrant prediction error, beliefs reconsolidate rather than extinguish. **(E)** Medication permits extinction learning by blocking aberrant prediction error – Under antipsychotic medication, aberrant prediction errors are blocked, allowing extinction learning to dominate and the resolution of the delusional belief.

BOX 1 | Reconsolidation based therapies for delusions

Classical demonstrations of reconsolidation were based on observations of amnesia for associations that animals had been reminded about just prior to electroconvulsive shock (ECS, Misanin et al., 1968). ECS has been administered therapeutically in schizophrenia (electroconvulsive therapy or ECT). Whilst effective in first-episode patients, it is less so in the more chronically ill (Fink and Sackeim, 1996), perhaps because the intervention is applied after a delusion has become strongly consolidated and is less susceptible to disruption.

ECT is usually administered under general anesthesia making its targeted application to a pathogenic memory problematic. However, ECT has been applied, without anesthesia, with a reminder treatment, to patients with fixed intrusive beliefs (Rubin, 1976). Patients had their attention directed toward their most disturbing beliefs, and were instantly given ECT. All patients improved dramatically for periods of up to 10 years at the time of publication.

Of course, this approach would no longer be considered ethically acceptable. However, the findings lend some support to the notion that the reconsolidation/extinction balance is important for delusion maintenance. A less invasive intervention such as transcranial magnetic stimulation might be employed to modulate reactivated memories, an approach that has also been indicated for motor memory rehabilitation following stroke (Bernad and Doyon, 2008).

Alternatively, it may be possible to target specific a reactivated memory pharmacologically, reactivating a memory and administering a drug to impair its reconsolidation whilst its representation is still labile. The impact of pharmacological manipulations on memory reconsolidation depends on the nature of what is recalled and how the memory is re-evoked (Lee et al., 2006): The effects of a glutamatergic agonist and an antagonist on reconsolidation were modulated by the way in which memories were reactivated. Following a long reactivation session [favoring extinction], blocking NMDA receptors blocked extinction and enhancing their function promoted extinction. However, following a brief reactivation session [favoring reconsolidation], the opposite was true: NMDA blockade enhanced and an NMDA agonist impaired extinction.

D-cycloserine (DCS) an NMDA receptor partial agonist has been trialed as an adjunct to traditional dopaminergic antipsychotics, but only modest improvements were observed, and in some patients, DCS aggravated positive symptoms (van Berckel et al., 1999). This inconsistency may have implications for the therapeutic targeting of delusions through reconsolidation blockade. When attempting to block reconsolidation, a brief re-exposure session should be used to render the belief labile. If the reminder session is too extensive, the amnestic agent will impair new extinction learning that is invoked and will lead to maintenance of the maladaptive memory underpinning the belief.

An alternative therapeutic approach might involve enhancing extinction of the delusional memory. However, this would have a limited effect as extinction learning is context dependent (Bouton, 2000). Following the addiction example, an addict will extinguish drug seeking behaviors in a rehabilitation treatment center but they may relapse outside this context (Taylor et al., 2009). Training in multiple different extinction contexts does not seem to improve this (Bouton et al., 2006). The ideal therapeutic approach would involve a combined approach which impaired reconsolidation but also encouraged new extinction learning.

Despite the observation that directly challenging a delusion may lead to a strengthening of the belief (Milton et al., 1978; Simpson and Done, 2002) cognitive therapies for delusions have been applied with reported success (Chadwick and Lowe, 1990, 1994; Rector and Beck, 2001; Turkington and Dudley, 2004). It is telling, with respect to the current model, that certain principles have emerged, (Chadwick and Lowe, 1994) notably:

1. Modification should begin with the least strongly held beliefs.
2. Patients are encouraged to consider the alternative to the delusional belief rather than to try to accept this alternative immediately.
3. Evidence for the belief, rather than the belief itself, should be challenged.
4. The patient should be encouraged to voice the arguments against his or her beliefs.

In the context of our model, principles 1 and 2 might attenuate the excessive reactivation and reconsolidation of the delusion while principles 3 and 4 would tip the balance in favor of its extinction.

Given our preference for a combined approach that enhances extinction and impairs reconsolidation, we note with interest the recent demonstration of significant weakening of a fear memory in rats by inducing extinction following reactivation (Monfils et al., 2009). These data could perhaps be mapped onto the suggested parameters for CBT targeting delusions. That is, the delusional memory could be re-engaged and, whilst it is labile, the data that are considered supportive could be challenged resulting in its extinction.

interpreted. To quote Jaspers; "*the trail is blazed and the preparedness for the significant experience then permeates almost all perceived contents. The now dominant delusion motivates the apperceptive schema for all future percepts*." (Jaspers, 1963) – see **Box 1**.

CONCLUSION

We have argued that aberrant prediction error signals may be important not only for delusion formation (Corlett et al., 2007a) but also for delusion maintenance since they drive the retrieval and reconsolidation based strengthening (Lee, 2008) of delusional beliefs, even in situations when extinction learning ought to dominate (Eisenhardt and Menzel, 2007; Pedreira et al., 2004). Given the proposed function of reconsolidation, in driving automaticity of behavior (Stickgold and Walker, 2007) we argue that in an aberrant prediction error system, delusional beliefs rapidly become inflexible habits. Taking this translational approach will enhance our understanding of psychotic symptoms and may move us closer to the consilience between the biology and phenomenology of delusions that Kapur sought in his article (Kapur, 2003).

ACKNOWLEDGEMENTS

The authors would like to thank Trevor Robbins, Jonathan Lee, Amy Milton, David Shanks and Graham Murray who gave insightful criticism of the hypothesis. Philip Corlett is the University of Cambridge Parke Davis Exchange Fellow in Biomedical Sciences and a NARSAD Young Investigator. Jane Taylor is supported by NARSAD and NIDA. John Krystal is supported by NIMH, NIAAA and NARSAD, he reports the following: Consulting: AstraZeneca Pharmaceuticals, LP, Cypress Bioscience, Inc., HoustonPharma, Schering-Plough Research Institute, Shire Pharmaceuticals, and Pfizer Pharmaceuticals; Advisory Boards: Bristol-Myers Squibb,

BOX 2 | The biology of memory maintenance

Changes in synaptic plasticity are widely believed to underpin learning and memory. However, the role of synaptic changes in memory maintenance was ascertained from preparations assuming Hebbian rules; put simply, neurons that fire together wire together. Simple Hebbian rules deal well with learning about contiguous events but they cannot explain the role of prediction error in learning (McLaren and Dickinson, 1990), nor can they explain the memory reconsolidation phenomena critical to the current hypothesis (Arshavsky, 2006). Furthermore, changes in synaptic plasticity may not be the mechanism for memory (or belief) maintenance across the lifespan – the components of a synapse that confer its strength and excitability are frequently completely recycled (Arshavsky, 2006). A more permanent memory storage mechanism is necessary to ensure the accurate restoration of synaptic proteins following such recycling events (Arshavsky, 2006). One relatively stable and information rich candidate is DNA. Changes in how the genome can be expressed could mediate long-term retention of information by an organism (Crick, 1984; Holliday, 1999). Such epigenetic changes represent a mechanism through which an organism can adapt to its environment and maintain that adaptation throughout its lifetime, processes that are likely disturbed in a number of mental illnesses (Bredy, 2007).

Much of the genome may be silenced in a particular cell, whilst other portions will be highly expressed; a process mediated in part through changes in proteins called histones which combine with DNA to form chromatin. Histone proteins can be modified in a variety of ways which impact upon how readily the DNA with which they are associated can be expressed as protein. Histone modification makes an important contribution to learning related gene expression. One particular histone modification, changes in acetylation, is mediated by an enzyme called histone deacetylase (HDAC). Inhibition of HDAC enhances long-term memory (Levenson et al., 2004).

Barad and colleagues recently demonstrated that the HDAC inhibitor valproic acid modulate extinction and reconsolidation of conditioned fear. They hypothesize that HDAC inhibitors: "*overcome the reconsolidation-like incubation of fear by spaced CS-presentation to allow effective extinction to take place*" (Bredy and Barad, 2008). This is exactly the desired effect for an antipsychotic drug within the context of the IRH. Whilst valproate, has not been trialed as a standalone antipsychotic, it does benefit patients with schizophrenia as an adjunct to antipsychotics (Basan and Leucht, 2004), future research should explore the role of HDAC inhibition as an antipsychotic mechanism, particularly its potential application in combination with reminder treatments as a therapeutic approach to delusions.

BOX 3 | Perception, learning and belief

We have recently speculated on the interaction between prior beliefs and current sensory experiences (Corlett, 2009, Fletcher and Frith, 2009) and how delusions may arise from an imbalance in these interactions. We argue that, in the earliest stages of psychosis, prior to delusion formation, noise in the nervous system leads to excessive free energy or prediction error (Friston, 2005). Subjectively, this experience is of a mutated world, one which feels strange, perhaps sinister and beset by new significance. Priors are updated in order to minimize free energy and this must necessarily lead to the formation of new beliefs and expectancies. The new model of the world is built to accommodate or explain away these strange experiences. With continued aberrant prediction error those new priors are strengthened further. In the terms of the present model we suggest that memory reconsolidation provides a potential mechanism for this updating process. With sufficient reconsolidation, the priors become so strong as to be resistant to contradictory evidence.

Pavlov believed that one could equate his conditioned reflexes with Helmholtz' learned perceptual expectancies (Barlow, 1990; Helmholtz, 1871/1971), an assertion supported by empirical evidence (Davies et al., 1982). Recently, computational neuroscientists have also come to appreciate the overlap between learning and perception, emphasizing the consilience between formal theories of conditioning, Bayesian accounts of learning and signal detection theories of perceptual decision making (Dayan and Daw, 2008). According to such combined models, organisms must learn the rules of the game they are playing and, in the context of those rules, they must discern meaningful events and contingencies from meaningless noise. Organisms achieve this by maintaining a robust but flexible set of expectancies about the world that tune their sensitivity and responses biases.

Estes drew similar parallels between learning, memory and signal detection theory in his stimulus sampling theory of memory (Estes, 1997). Here, after a memory trace is encoded, the occurrence of the same or similar events reactivates the trace, necessitating some re-encoding or reconsolidation process, arguing that this process is subject to random error, due to the physiological noise that permeates the nervous system and provides the analytical basis for signal detection theory (Green and Swets, 1966). Estes hypothesizes that random error (or physiological noise) at re-encoding may account for the loss of detailed information, the incorporation of spurious details and the shift toward familiar information that occur with successive reproductions of information from memory (Bartlett, 1932). Unlike Estes' model, in our account prediction error signal (both appropriate and inappropriate) drives memory reactivation and also impacts on how that memory is re-encoded.

We argue that delusions occur in the context of a noisy nervous system that is attempting to form and maintain a robust set of priors. Such excessive noise would engender more cycles of reactivation and subsequent reconsolidation, leading to a bizarre and maladaptive set of expectancies about the world, expectancies strong enough to vitiate normal sensory and cognitive experience. These learned expectancies have much in common with Jaspers' apperceptive schema' – beliefs so strong that they are impervious to sensation – which he believed were the basis for the maintenance of delusional beliefs (Jaspers, 1963).

Eli Lilly and Co., Forest Laboratories, GlaxoSmithKline, Lohocla Research Corporation, Merz Pharmaceuticals, Takeda Industries, and Transcept Pharmaceuticals, Inc.; Exercisable Warrant Options: Tetragenex Pharmaceuticals Inc.; Research Support: Janssen Research Foundation; Pending Patents: glutamatergic agents for psychiatric disorders (depression, OCD), antidepressant effects of oral ketamine, and oral ketamine for depression. All other authors report no biomedical financial interests or potential conflicts of interest. Paul Fletcher is supported by the Bernard Wolfe Health Neuroscience Fund and the Wellcome Trust.

REFERENCES

Alberini, C. M. (2007). Reconsolidation: the samsara of memory consolidation. *Debates Neurosci.* 1, 17–24.

Arshavsky, Y. I. (2006). "The seven sins" of the Hebbian synapse: can the hypothesis of synaptic plasticity explain long-term memory consolidation? *Prog. Neurobiol.* 80, 99–113.

Barlow, H. (1990). Conditions for versatile learning, Helmholtz's unconscious inference, and the task of perception. *Vis. Res.* 30, 1561–1571.

Bartlett, F. C. (1932). Remembering. Cambridge, Cambridge University Press.

Basan, A., and Leucht, S. (2004). Valproate for schizophrenia. *Cochrane Database Syst. Rev.* CD004028.

Bernad, D. M., and Doyon, J. (2008). The role of noninvasive techniques in stroke therapy. *Int. J. Biomed. Imaging* 2008, 672582.

Berridge, K. C., and Robinson, T. E. (1998). What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Res. Brain Res. Rev.* 28, 309–369.

Bouton, M. E. (2000). A learning theory perspective on lapse, relapse, and the maintenance of behavior change. *Health Psychol.* 19, 57–63.

Bouton, M. E., Garcia-Gutierrez, A., Zilski, J., and Moody, E. W. (2006). Extinction in multiple contexts does not necessarily make extinction less vulnerable to relapse. *Behav. Res. Ther.* 44, 983–994.

Bredy, T. W. (2007). Behavioural epigenetics and psychiatric disorders. *Med. Hypotheses* 68, 453.

Bredy, T. W., and Barad, M. (2008). The histone deacetylase inhibitor valproic acid enhances acquisition, extinction, and reconsolidation of conditioned fear. *Learn. Mem.* 15, 39–45.

Chadwick, P. K. (2001). Psychotic consciousness. *Int. J. Soc. Psychiatry* 47, 52–62.

Chadwick, P. D., and Lowe, C. F. (1990). Measurement and modification of delusional beliefs. *J. Consult. Clin. Psychol.* 58, 225–232.

Chadwick, P. D., and Lowe, C. F. (1994). A cognitive approach to measuring and modifying delusions. *Behav. Res. Ther.* 32, 355–367.

Corlett, P. R., Frith, C. D., and Fletcher, P. C. (2009). From drugs to deprivation: a Bayesian framework for understanding models of psychosis. *Psychopharmacology (Berl)*. [Epub ahead of print].

Corlett, P. R., Honey, G. D., Aitken, M. R., Dickinson, A., Shanks, D. R., Absalom, A. R., Lee, M., Pomarol-Clotet, E., Murray, G. K., McKenna, P. J., Robbins, T. W., Bullmore, E. T., and Fletcher, P. C. (2006). Frontal responses during learning predict vulnerability to the psychotogenic effects of ketamine: linking cognition, brain activity, and psychosis. *Arch. Gen. Psychiatry* 63, 611–621.

Corlett, P. R., Honey, G. D., and Fletcher, P. C. (2007a). From prediction error to psychosis: ketamine as a pharmacological model of delusions. *J. Psychopharmacol.* 21, 238–252.

Corlett, P. R., Murray, G. K., Honey, G. D., Aitken, M. R., Shanks, D. R., Robbins, T. W., Bullmore, E. T., Dickinson, A., and Fletcher, P. C. (2007b). Disrupted prediction-error signal in psychosis: evidence for an associative account of delusions. *Brain* 130, 2387–2400.

Crick, F. (1984). Memory and molecular turnover. *Nature* 312, 101.

Crombag, H. S., Bossert, J. M., Koya, E., and Shaham, Y. (2008). Context-induced relapse to drug seeking: a review. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 363, 3233–3243.

Davies, P., Davies, G. L., and Bennett, S. (1982). An effective paradigm for conditioning visual perception in human subjects. *Perception* 11, 663–669.

Dayan, P., and Daw, N. D. (2008). Decision theory, reinforcement learning, and the brain. *Cogn. Affect. Behav. Neurosci.* 8, 429–453.

Dickinson, A. (2001). The 28th Bartlett Memorial Lecture. Causal learning: an associative analysis. *Q. J. Exp. Psychol. B* 54, 3–25.

Eichenbaum, H., and Bodkin, J. A. (2000). Belief and knowledge as distinct forms of memory. In Memory, Brain and Belief, D. L. Schacter and E. Scarry, eds (Cambridge, MA, Harvard University Press).

Eisenhardt, D., and Menzel, R. (2007). Extinction learning, reconsolidation and the internal reinforcement hypothesis. *Neurobiol. Learn. Mem.* 87, 167–173.

El-Amamy, H., and Holland, P. C. (2007). Dissociable effects of disconnecting amygdala central nucleus from the ventral tegmental area or substantia nigra on learned orienting and incentive motivation. *Eur. J. Neurosci.* 25, 1557–1567.

Estes, W. K. (1997). Processes of memory loss, recovery, and distortion. *Psychol. Rev.* 104, 148–169.

Fink, M., and Sackeim, H. A. (1996). Convulsive therapy in schizophrenia? *Schizophr. Bull.* 22, 27–39.

Fletcher, P. C., and Frith, C. D. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nat. Rev. Neurosci.* 10, 48–58.

Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 360, 815–836.

Goodwin, D. W., Powell, B., Bremer, D., Hoine, H., and Stern, J. (1969). Alcohol and recall: state-dependent effects in man. *Science* 163, 1358–1360.

Green, D. M., and Swets, J. A. (1966). Signal Detection Theory and Psychophysics. New York, Wiley.

Hamamura, T., and Harada, T. (2007). Unique pharmacological profile of aripiprazole as the phasic component buster. *Psychopharmacology (Berl)*. 191, 741–743.

Helmholtz, H. (1871/1971). The facts of perception. In: Selected Writings of Hermann von Helmholtz, K. Russell, ed. (Middletown, Wesleyan University Press).

Holliday, R. (1999). Is there an epigenetic component in long-term memory? *J. Theor. Biol.* 200, 339–341.

Holt, D. J., Lebron-Milad, K., Milad, M. R., Rauch, S. L., Pitman, R. K., Orr, S. P., Cassidy, B. S., Walsh, J. P., and Goff, D. C. (2009). Extinction memory is impaired in schizophrenia. *Biol. Psychiatry* 65, 455–463.

Horvitz, J. C. (2002). Dopamine gating of glutamatergic sensorimotor and incentive motivational input signals to the striatum. *Behav. Brain Res.* 137, 65–74.

Jaspers, K. (1963). General Psychopathology. Manchester, Manchester University Press.

Joseph, R. (1986). Confabulation and delusional denial: frontal lobe and lateralized influences. *J. Clin. Psychol.* 42, 507–520.

Kapur, S. (2003). Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. *Am. J. Psychiatry* 160, 13–23.

Kapur, S. (2004). How antipsychotics become anti-"psychotic" – from dopamine to salience to psychosis. *Trends Pharmacol. Sci.* 25, 402–406.

Lee, J. L. (2008). Memory reconsolidation mediates the strengthening of memories by additional learning. *Nat. Neurosci.* 11, 1264–1266.

Lee, J. L., Di Ciano, P., Thomas, K. L., and Everitt, B. J. (2005). Disrupting reconsolidation of drug memories reduces cocaine-seeking behavior. *Neuron* 47, 795–801.

Lee, J. L., Milton, A. L., and Everitt, B. J. (2006). Reconsolidation and extinction of conditioned fear: inhibition and potentiation. *J. Neurosci.* 26, 10051–10056.

Levenson, J. M., O'Riordan, K. J., Brown, K. D., Trinh, M. A., Molfese, D. L., and Sweatt, J. D. (2004). Regulation of histone acetylation during memory formation in the hippocampus. *J. Biol. Chem.* 279, 40545–40559.

Lovibond, P. F. (2004). Cognitive processes in extinction. *Learn. Mem.* 11, 495–500.

Maher, B. A. (1974). Delusional thinking and perceptual disorder. *J. Individ. Psychol.* 30, 98–113.

McLaren, I. P., and Dickinson, A. (1990). The conditioning connection. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 329, 179–186.

Menon, M., Jensen, J., Vitcu, I., Graff-Guerrero, A., Crawley, A., Smith, M. A., and Kapur, S. (2007). Temporal difference modeling of the blood-oxygen level dependent response during aversive conditioning in humans: effects of dopaminergic modulation. *Biol. Psychiatry* 62, 765–772.

Miller, R. (1989). Hyperactivity of associations in psychosis. *Aust. NZ J. Psychiatry* 23, 241–248.

Miller, R. (1993). Striatal dopamine in reward and attention: a system for understanding the symptomatology of acute schizophrenia and mania. *Int. Rev. Neurobiol.* 35, 161–278.

Milton, F., Patwa, V. K., and Hafner, R. J. (1978). Confrontation vs. belief modification in persistently deluded patients. *Br. J. Med. Psychol.* 51, 127–130.

Misanin, J. R., Miller, R. R., and Lewis, D. J. (1968). Retrograde amnesia produced by electroconvulsive shock after reactivation of a consolidated memory trace. *Science* 160, 554–555.

Monfils, M. H., Cowansage, K. K., Klann, E., and LeDoux, J. E. (2009). Extinction-reconsolidation boundaries: key to persistent attenuation of fear memories. *Science* 324, 951–955.

Morris, R. G., Inglis, J., Ainge, J. A., Olverman, H. J., Tulloch, J., Dudai, Y., and Kelly, P. A. (2006). Memory reconsolidation: sensitivity of spatial memory to inhibition of protein synthesis in dorsal hippocampus during encoding and retrieval. *Neuron* 50, 479–489.

Murray, G. K., Corlett, P. R., Clark, L., Pessiglione, M., Blackwell, A. D., Honey, G., Jones, P. B., Bullmore, E. T., Robbins, T. W., and Fletcher, P. C. (2008). Substantia nigra/ventral tegmental reward prediction error disruption in psychosis. *Mol. Psychiatry* 13, 239, 267–276.

Nader, K., Schafe, G. E., and Le Doux, J. E. (2000). Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature* 406, 722–726.

Pan, W. X., Schmidt, R., Wickens, J. R., and Hyland, B. I. (2008). Tripartite

mechanism of extinction suggested by dopamine neuron activity and temporal difference model. *J. Neurosci.* 28, 9619–9631.

Pavlov, I. P. (1927). Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex. New York, Dover Publications Inc.

Pedreira, M. E., Perez-Cuesta, L. M., and Maldonado, H. (2004). Mismatch between what is expected and what actually occurs triggers memory reconsolidation or extinction. *Learn. Mem.* 11, 579–585.

Rector, N. A., and Beck, A. T. (2001). Cognitive behavioral therapy for schizophrenia: an empirical review. *J. Nerv. Ment. Dis.* 189, 278–287.

Redgrave, P., and Gurney, K. (2006). The short-latency dopamine signal: a role in discovering novel actions? *Nat. Rev. Neurosci.* 7, 967–975.

Redish, A. D., Jensen, S., Johnson, A., and Kurth-Nelson, Z. (2007). Reconciling reinforcement learning models with behavioral extinction and renewal: implications for addiction, relapse, and problem gambling. *Psychol. Rev.* 114, 784–805.

Reus, V. I., Weingartner, H., and Post, R. M. (1979). Clinical implications of state-dependent learning. *Am. J. Psychiatry* 136, 927–931.

Rubin, R. D. (1976). Clinical use of retrograde amnesia produced by electroconvulsive shock. A conditioning hypothesis. *Can. Psychiatr. Assoc. J.* 21, 87–90.

Ruocchio, P. J. (1991). First person account: the schizophrenic inside. *Schizophr. Bull.* 17, 357–360.

Sass, L. A. (2004). Some reflections on the (analytic) philosophical approach to delusion. *Philos. Psychiatr. Psychol.* 11, 71–80.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *J. Neurophysiol.* 80, 1–27.

Schultz, W., and Dickinson, A. (2000). Neural coding of prediction errors. *Annu. Rev. Neurosci.* 23, 473–500.

Simpson, J., and Done, D. J. (2002). Elasticity and confabulation in schizophrenic delusions. *Psychol. Med.* 32, 451–458.

Sinha, V. K., and Chaturvedi, S. K. (1989). Persistence of delusional content among psychotics over consecutive episodes. *Psychopathology* 22, 208–212.

Smith, A., Li, M., Becker, S., and Kapur, S. (2006). Dopamine, prediction error and associative learning: a model-based account. *Network* 17, 61–84.

Stanton, B., and David, A. (2000). First-person accounts of delusions. *Psychiatr. Bull.* 24, 333–336.

Stickgold, R., and Walker, M. P. (2007). Sleep-dependent memory consolidation and reconsolidation. *Sleep Med.* 8, 331–343.

Taylor, J. R., Olausson, P., Quinn, J. J., and Torregrossa, M. M. (2009). Targeting extinction and reconsolidation mechanisms to combat the impact of drug cues on addiction. *Neuropharmacology* 56(Suppl. 1), 186–195.

Tronson, N. C., and Taylor, J. R. (2007). Molecular mechanisms of memory reconsolidation. *Nat. Rev. Neurosci.* 8, 262–275.

Tronson, N. C., Wiseman, S. L., Olausson, P., and Taylor, J. R. (2006). Bidirectional behavioral plasticity of memory reconsolidation depends on amygdalar protein kinase A. *Nat. Neurosci.* 9, 167–169.

Turkington, D., and Dudley, R. (2004). Cognitive behavioral therapy in the treatment of schizophrenia. *Expert Rev. Neurother.* 4, 861–868.

van Berckel, B. N., Evenblij, C. N., van Loon, B. J., Maas, M. F., van der Geld, M. A., Wynne, H. J., van Ree, J. M., and Kahn, R. S. (1999). D-cycloserine increases positive symptoms in chronic schizophrenic patients when administered in addition to antipsychotics: a double-blind, parallel, placebo-controlled study. *Neuropsychopharmacology* 21, 203–210.

Waltz, J. A., Frank, M. J., Robinson, B. M., and Gold, J. M. (2007). Selective reinforcement learning deficits in schizophrenia support predictions from computational models of striatal–cortical dysfunction. *Biol. Psychiatry* 62, 756–764.

Conflict of Interest Statement: The authors have no competing financial interests.

Received: 01 April 2009; paper pending published: 17 May 2009; accepted: 16 June 2009; published online: 10 July 2009.
*Citation: Corlett PR, Krystal JH, Taylor JR and Fletcher PC (2009) Why do delusions persist? Front. Hum. Neurosci. (2009) **3**:12. doi: 10.3389/neuro.09.012.2009*