



Illusions and delusions: relating experimentally-induced false memories to anomalous experiences and ideas

Philip R. Corlett^{1,2*}, Jon S. Simons³, Jennifer S. Pigott⁴, Jennifer M. Gardner¹, Graham K. Murray¹, John H. Krystal² and Paul C. Fletcher¹

¹ Department of Psychiatry, University of Cambridge, Brain Mapping Unit, Cambridge, UK

² Department of Psychiatry, Abraham Ribicoff Research Facility, Yale University Medical School, Connecticut Mental Health Centre, New Haven, CT, USA

³ Department of Experimental Psychology, University of Cambridge, Cambridge, UK

⁴ University College London Medical School, University College London, London, UK

Edited by:

Andreas Meyer-Lindenberg, Central Institute of Mental Health, Germany

Reviewed by:

Heike Tost, National Institute of Mental Health, USA

Joshua W. Buckholz, Vanderbilt University, USA

*Correspondence:

Philip R. Corlett, Yale University Medical School, Abraham Ribicoff Research Facility, Connecticut Mental Health Centre, New Haven, CT, USA.
e-mail: philip.corlett@yale.edu

The salience hypothesis of psychosis rests on a simple but profound observation that subtle alterations in the way that we perceive and experience stimuli have important consequences for how important these stimuli become for us, how much they draw our attention, how they embed themselves in our memory and, ultimately, how they shape our beliefs. We put forward the idea that a classical memory illusion – the Deese–Roediger–McDermott (DRM) effect – offers a useful way of exploring processes related to such aberrant belief formation. The illusion occurs when, as a consequence of its relationship to previous stimuli, a stimulus that has not previously been presented is falsely remembered. Such illusory familiarity is thought to be generated by the surprising fluency with which the stimulus is processed. In this respect, the illusion relates directly to the salience hypothesis and may share common cognitive underpinnings with aberrations of perception and attribution that are found in psychosis. In this paper, we explore the theoretical importance of this experimentally-induced illusion in relation to the salience model of psychosis. We present data showing that, in healthy volunteers, the illusion relates directly to self reported anomalies of experience and magical thinking. We discuss this finding in terms of the salience hypothesis and of a broader Bayesian framework of perception and cognition which emphasizes the salience both of predictable and unpredictable experiences.

Keywords: false memory, psychosis, salience, prediction error

INTRODUCTION

Understanding how our brains represent and misrepresent our world and how such representations guide our behavior is a key goal for neuroscience and psychiatry (Kandel, 1999). Symptoms such as hallucinations and delusions, associated with mental illnesses like schizophrenia, involve gross alterations in the experience of reality (Crow, 1980). For sufferers of these symptoms, the world can seem strange, sinister and sometimes inexplicable (Matussek, 1954; Conrad, 1958; McGhie and Chapman, 1961; Arieti, 1974). The drive to explain away odd experiences can result in delusion formation (Maher, 1974; Kapur, 2003), which then sculpts and vitiates future experiences and attributions (Jaspers, 1963; Pally, 2007; Corlett et al., 2009b).

Perception (Helmholtz, 1878/1971) and memory (Schacter et al., 1998) are substantially constructive; what we perceive and remember is strongly influenced by our prior experiences and expectancies. This interaction between efferent stimulation and top-down expectation reflects an efficient approach to representing reality (Rao and Ballard, 1999; Friston, 2005), but it leaves us susceptible to mistakes; manifest as perceptual and cognitive illusions: percepts and recollections that depart from objective reality (Gregory, 1996). Our memories are subject to experience-dependent biases (Bartlett, 1932) and are highly mutable (Loftus and Palmer, 1974; Estes, 1997). For example, with repeated retrieval of a memory comes repeated reconsolidation, a process

of re-stabilization which is susceptible to physiological noise that can sculpt a memory into a misrepresentation of what actually occurred (Estes, 1997; Corlett et al., 2009b).

Individuals with schizophrenia may experience illusions differently. For example, they are less susceptible to some perceptual illusions (Emrich, 1989; Dakin et al., 2005), but they sometimes form false memories about what they have perceived (Moritz and Woodward, 2002) and they can falsely attribute the source of those memories to entities external to themselves (Keefe et al., 1999). These observations could result from a poor specification of prior expectancies, leading perceptual illusions to fail, recollections to become confused, and on-going experiences to be unusually salient (Gray et al., 1991; Hemsley, 1994). It has been suggested that symptoms of psychosis can be explained in terms of these alterations (Gray et al., 1991; Kapur, 2003; Corlett et al., 2007a, 2009a,b) and this has led to a renewed interest in the how patients perform on the sorts of learning tasks that engage these processes (Corlett et al., 2007b; Jensen et al., 2008; Murray et al., 2008; Roiser et al., 2009). In particular, recent work has invoked abnormalities in prediction error signaling as a possible underlying factor in the emergence of psychosis (Corlett et al., 2006, 2007b; Schlagenhauf et al., 2009). However, while recent experimental examinations of the hypothesis have tended to involve trial-by-trial reinforcement learning tasks [with some notable exceptions (Menon et al., 2005; Holt et al., 2006; Gibbs et al., 2007, 2008)], it is also possible to explore these processes in the context of

a memory domain also known to be abnormal in schizophrenia, that of episodic memory (Tulving, 1987). If prediction errors govern false memory, and psychosis results from aberrant prediction error, then individuals who are prone psychotic experiences may be differentially susceptible to forming illusory false memories.

One well-established way of producing illusory memory was described by Deese (1959) and Roediger and McDermott (1995). This “DRM” effect manifests as a greater tendency to assert falsely that a prototype word (e.g. sleep) was seen when it is presented following a study list consisting of a number of its semantic associates (night, dream etc). That is, subjects report an illusory or false memory for non-presented words that are highly associated with the words that they were asked to remember.

The most recent theoretical perspectives on this effect relate strongly to the notions of surprise and prediction error that we have raised above. The discrepancy-attribution hypothesis (Whittlesea and Williams, 1998; Whittlesea, 2002) suggests that subjects monitor their cognitive processing by comparing expectations about their performance with their current processing. Both true and false familiarity are experienced when a comparison between expectations and outcomes within a processing episode elicits a surprise or prediction error. Like reward prediction errors and other expectancy violations, this motivates the person to seek an explanation, attributing the surprise to some plausible source in the current environment or to similar features in the past. Prototype words in the DRM feel surprisingly fluent, they exceed the fluency threshold being used to perform the task. Subjects ascribe this fluency to a source in the past, experiencing a conscious feeling of familiarity for words that were not presented to them (Whittlesea et al., 2005). In short, because the prototype word is strongly related to the list of words that were actually studied, it is processed with greater fluency and this fluency engenders the surprise signal that is interpreted incorrectly to mean that the word was actually presented.

Thus, this illusion offers the opportunity to explore prediction error-related function in the context of episodic memory. If surprise or prediction error is the basis for the DRM illusion then individuals with aberrant prediction error processing might be expected to experience the illusion in differing ways. To this end, we sought to determine whether there were individual differences in the DRM effect that could be related to attenuated psychosis-like experiences and beliefs (Eckblad and Chapman, 1983; Peters et al., 1999). This approach has proven successful in previous investigations of source memory impairments and how they relate to psychosis-like experiences (Simons et al., 2006, 2008). We assessed the DRM effect in healthy volunteers and asked them to complete questionnaires that assayed their susceptibility to unusual experiences and beliefs. Based on our previous finding that individuals most sensitive to prediction error were more likely to experience perceptual aberrations and delusional ideation following treatment with the psychotomimetic drug ketamine (Corlett et al., 2006), we predicted that those individuals who were most sensitive to the DRM illusion would have the most severe psychosis-like experiences including strange perceptual experiences and odd beliefs.

MATERIALS AND METHODS

PARTICIPANTS

All participants tested were undergraduate students at the University of Cambridge, recruited through poster and email advertisement. Participants were aged between 18 and 22 years (mean 21.16, SD

1.13 years). They participated in the study for the chance to win a prize for most accurate performance on the computer task. 50 participants were tested in total (18 males and 32 females). All subjects provided informed consent, and had a clear understanding of what the task involved before beginning. The study was approved by the Cambridge Psychological Research Ethics Committee, Cambridge, England.

COGNITIVE TASK

The DRM task replicated Whittlesea et al. (2005). Word stimuli were taken from Stadler et al. (1999). There were 36 semantic prototypes, each with a list of 15 high-frequency word associates (Stadler et al., 1999). For example, for the prototype word *cold*, associates (in descending order of associative frequency) would be *hot, snow, warm, winter, ice, wet, frigid, chilly, heat, weather, freeze, air, shiver, arctic* and *frost*. The words in each list tend to be elicited as associates of the prototype in free recall, the higher up the list the stronger the association with the prototype word (Stadler et al., 1999).

Two shortened lists of six words were created from each 15 word list. The first shortened list consisted of the odd-numbered items from each list, beginning with the third word (for the example above: warm, ice, frigid, heat, freeze, shiver), and the other shortened list consisted of the even-numbered items from the list, beginning with the fourth word (in this case winter, wet, chilly, weather, air, arctic). The 15th word, the prototype and the two highest associates from the list were not included in the shortened lists. This resulted in 72 lists of six words, each associated with a prototype and the two highest associates. Eighteen of these six-word lists were assigned to each of the four test conditions for each participant. This was done at random for each condition and for each participant.

The task was presented in the DMDX stimulus delivery programme, on an Advantech PPC-123T portable PC running Windows 98, with a 12.1” screen.

In each trial, participants were shown one list of six words. Each word was presented in white letters at the centre of a black screen for 120 ms, followed immediately by the next word in the list. Following the presentation of these six words, a question was displayed for 3 s asking whether a particular word had been seen in the list, for example ‘Did you see the word tiger?’ At the start of the task, two practice trials were presented, comprising words that were not present in the task lists. We adopted this method of rapid serial visual presentation (RSVP) because of our interest in the relationships between expectancy, experience and memory (Luo et al., 1998; McDermott and Watson, 2001; Whittlesea et al., 2005). Indeed, forming false memories about rapidly presented visual information and having an hallucination may have a great deal in common (Reed et al., 2008). In this respect, a secondary aim of our study is to relate the salience hypotheses of psychosis more directly to hallucination-like experiences.

Participants were instructed to respond as quickly and accurately as possible by pressing the computer key corresponding to their decision. The keys were the left and right shift key on a standard keyboard, the relevant key was labeled with a sticker saying ‘yes’ or ‘no’ to avoid confusion. The role of the particular keys was counter-balanced across subjects. After Moritz and Woodward (2002), we asked subjects to indicate how confident they were in their memory

decisions by holding the button down for longer if they were more confident. If they were guessing they tapped the key, if they were very confident they pressed and held for longer.

There were four different list and target word conditions, each consisting of 18 trials, presented in a different order for each subject:

- (1) *Congruent prototype lure* – a prototype was presented immediately after one of the lists of its six high-frequency associates. The prototype did not appear in the list of associates.
- (2) *Incongruent prototype lure* – a prototype was presented immediately after one of the lists of six high-frequency associates for a different prototype. The prototype did not appear in the list of associates.
- (3) *Highest associate presented* – a highest associate was presented immediately after one of the lists of six high-frequency associates for its prototype. In this case, the highest associate presented in the question was inserted into the list at random in locations 2–5, so that it was always preceded by at least one word and was never the final word.
- (4) *Highest associate not presented* – a highest associate was presented immediately after one of the lists of six high-frequency associates for a different prototype.

QUESTIONNAIRES

Subjects completed the following questionnaires:

- *Chapman Scales* (Eckblad and Chapman, 1983), which provided a measure of schizotypy, separated into subscales of magical ideation, physical anhedonia, social anhedonia, and perceptual aberration.
- *Peters Delusion Inventory* (Peters et al., 1999), which provided measures of unusual beliefs including endorsement and preoccupation.
- *Marlowe-Crowne Social Desirability Scale* (Strahan and Gerbasi, 1973) which provided a measure of the tendency of the participant to give answers that are perceived as acceptable by the investigator, rather than giving an account of their true experience. It poses questions about behaviours that are socially desirable but relatively uncommon, or that are socially undesirable but relatively common within the population.

Subjects also completed other questionnaires, the data from which will be reported elsewhere.

PLANNED ANALYSES

DRM task results

The DRM effect was summarized as the difference in probability of a false alarm to a related prototype and the probability of a false alarm to an unrelated prototype. Given the aim of the study; to explore the relationship between attenuated psychotic experiences and susceptibility to the DRM illusion, we also computed a more continuous measure of the illusion, incorporating subject's confidence in their memories. First, we corrected for inter-subject variability in confidence by subtracting each subject's mean confidence across all conditions from their response on every trial and then squaring this value, to express subjects' confidence for every

trial as a positive value. We then calculated a mean confidence in the errors that subjects made on DRM trials (when they were asked if they had been presented with a related prototype lure). This gave us a value for subject's confidence in having perceived prototype words that they were not actually presented.

Relating task and questionnaire data

Subjects' total scores on the Chapman Schizotypy scales provided a summary of their total schizotypy, incorporating both attenuated positive and attenuated negative symptoms. Subject's tendency to endorse odd beliefs as well as their preoccupation with and conviction were rated using the Peters' Delusion Inventory (PDI). Total scores on both scales were separately regressed onto both the probability and confidence assays of the DRM.

Given our *a priori* hypotheses about salience, expectancy and belief subsequent planned analyses ignored physical and social anhedonia (since they relate to attenuated negative symptoms) and examined the relationships between confidence in the DRM illusion and subjects' summed scores on the magical ideation and perceptual aberration subscales. Items on the Magical Ideation scale captured the sorts of aberrant experience that are crucial to salience based models of psychosis-like experiences (Gray et al., 1991; Hemsley, 1994; Kapur, 2003; Corlett et al., 2007a, 2009b). For example, the magical ideation scale includes items such as "I have noticed sounds on my records that are not there at other times", "I have felt that there were messages for me in the way things were arranged, like in a store window", and "I sometimes have a feeling of gaining or losing energy when certain people look at me or touch me." Items on the perceptual aberration subscale are more concerned with the intensity of percepts, unusual bodily experiences and aberrations of agency.

Signal detection theory and psychosis-like experiences

Subjects' decision making on this task may have been influenced by a number of different factors including their perceptual sensitivity (their ability to perceive very rapidly presented stimuli) and their response biases (e.g. how frequently they believed we would actually present them with words that we subsequently questioned them on). We captured subjects' sensitivity and response biases with a signal detection theory (SDT) analysis (Green and Swets, 1966). In brief, SDT conceives of decision making under uncertainty as a process of discerning between two distributions; that of internal physiological noise and the impact of real signals (plus noise) on the nervous system (see **Figure 3A**).

Perceptual sensitivity is summarized using d' (d-prime), the difference in the peaks of the internal noise and signal plus noise distributions. We estimated d-prime for each subject from our data as the difference between the normalized hit rate (correctly endorsing a word that was presented) and the normalized false alarm rate (incorrectly endorsing a prototype word that was not presented). The more sensitive the participant is at discriminating between presented target and prototype lures, the larger the d' value will be (MacMillan and Creelman, 1991; Shapiro, 1994; Fox, 2004).

We calculated subjects' response bias or criterion, c , by multiplying the sum of the standardized scores of the hit rates and the false alarm rates by -0.5 (MacMillan and Creelman 1991; Shapiro, 1994; Fox, 2004). When the false alarm rate for prototype lures equals the miss rate for targets that were presented, a participant is just as

likely to say “yes” as to say “no” in making recognition judgments (MacMillan, 1991; Shapiro, 1994; Fox, 2004). When this happens, the value for c is zero and the criterion is considered unbiased (MacMillan, 1991; Fox, 2004). When the false alarm rate is greater than the miss rate, the bias is toward answering “yes,” indicating a liberal criterion bias with a negative value (MacMillan, 1991; Fox, 2004). On the other hand, when the miss rate is greater than the false alarm rate, the bias is toward answering “no,” indicating a conservative criterion bias with a positive value (MacMillan, 1991; Fox, 2004). See **Figure 3B** Bias and Criterion scores were regressed onto our measures of positive schizotypy to assess the relationships between illusory memories and aberrant beliefs and experiences.

RESULTS

DRM TASK RESULTS

Our subjects’ demonstrated a DRM memory illusion; the rate of false alarms to prototype words that were semantically related to the preceding word list was significantly higher than the false alarm rate for prototype words that were unrelated to the preceding list [$t(49) = 10.828, p < 0.001$]. Recognition performance was good, subjects correctly reported having seen a word with which they were presented significantly better than chance [$t(49) = 10.678, p < 0.001$]. However, their performance was not perfect, on average, subjects incorrectly guessed that a presented word was in fact novel on 5 out of 18 trials.

RELATING TASK AND QUESTIONNAIRE DATA

We found no association between probability of making DRM errors and schizotypy scores or delusion-like ideation. However, this is perhaps unsurprising, given that Morris and Woodward found

no difference between the number of false memories formed by patients with schizophrenia and matched control subjects (Moritz and Woodward, 2002). Patients with schizophrenia were, however, more confident in the DRM memory errors they made (Moritz and Woodward, 2002). Similarly, we observed a significant correlation between subjects’ confidence in their false memories and their total score on the Chapman scales ($r = 0.271, p = 0.04, r^2 = 0.07$). [There was also a strong trend toward this measure predicting the number of unusual beliefs endorsed on the Peter’s delusion inventory ($r = 0.272, p = 0.06, r^2 = 0.07$).

When we examined the subscales most relevant to the cognitive models under investigation, we found a significant correlation between positive schizotypy and confidence in the DRM illusion ($r = 0.374, p = 0.008, r^2 = 0.11$). Subjects who were most confident in their false memories were most likely to have experienced perceptual aberrations or to have endorsed magical ideas (see **Figure 1**). In *post hoc* confirmatory analyses, neither social nor physical anhedonia were significantly associated with confidence in DRM errors ($r < 0.1, p > 0.5$).

Partial correlation analyses took into account potentially confounding factors; the correlation remained significant when age ($r = 0.374, p = 0.008$) and social desirability (as measured with the Marlow-Crowne scale, $r = 0.383, p = 0.007, r^2 = 0.15$) were taken into account. There were too few left-handed participants to examine the effects of handedness on our variables of interest, however, removal of left-handed participants did not eliminate the significant relationship between magical ideation scores and DRM susceptibility ($n = 45, r = 0.397, p = 0.007, r^2 = 0.16$).

We also explored the relationship between failure to recall events that were actually presented and schizotypal experiences. There were no significant associations between odd experiences and beliefs and

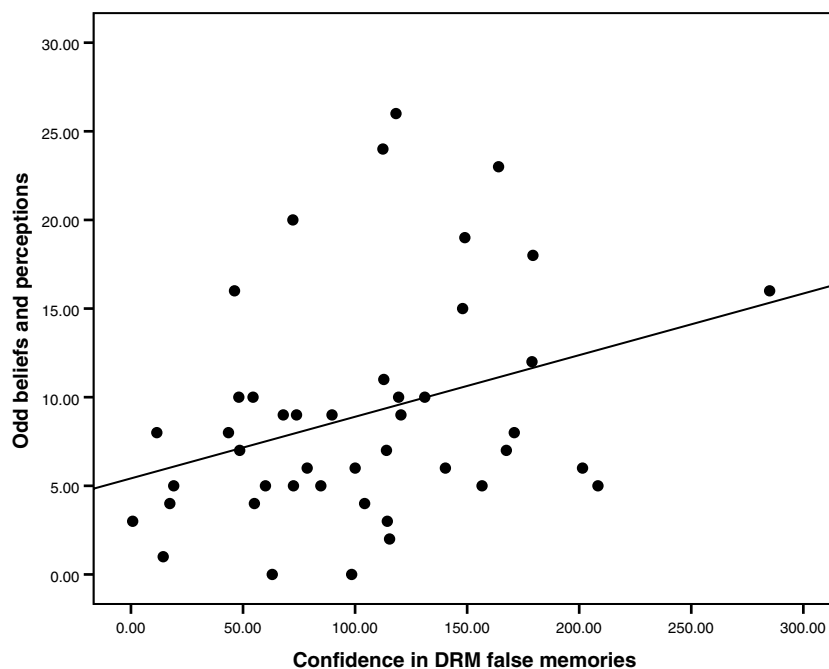


FIGURE 1 | The relationship between DRM confidence and odd beliefs/experiences: Scatter-plot of subject’s confidence in DRM errors and their self-rated positive schizotypy scores.

the number of failures to endorse experienced events, nor subjects' confidence in these errors and the severity of their unusual experiences and ideas. The association between confidence in memory errors and positive schizotypal traits appeared to be specific to DRM false memories rather than related to poor task performance or inappropriate confidence in false memories in general.

SDT AND PSYCHOSIS-LIKE EXPERIENCE

There was no significant correlation between d' and psychosis-like experiences or beliefs. However, there was a significant negative correlation between criterion and positive schizotypy score ($r = -0.353$, $p = 0.019$, $r^2 = 0.12$) that is, subjects with a more liberal acceptance bias (calculated from their decisions on the DRM trials) were also more likely to endorse having had unusual perceptual experiences and to hold unusual beliefs (see **Figures 2 and 3**).

Finally, as a test of internal consistency, we examined the correlation between criterion and confidence in DRM false memories, the significant relationship ($r = -0.362$, $p = 0.016$, $r^2 = 0.13$) unifies two of our measures of sensitivity to the illusion that relate to psychosis-like experiences and beliefs; subjects with a liberal acceptance bias on the task were more confident in the DRM errors that they made and more likely to have had odd perceptions and beliefs.

DISCUSSION

We produced illusory memories in healthy individuals using a well-established approach. As expected, participants were more likely to assert falsely that they recalled a word when it was semantically related to words with which they actually had been presented. Critically individuals who were prone to unusual experiences and beliefs were more confident in these illusory memories.

So, why might it be the case that these confident but false assertions tend to covary with the tendency to endorse beliefs about such things as supernatural occurrences or extraterrestrial creatures? The false sense of fluency or familiarity of prototype words in this task has been suggested to engage a surprise or prediction error: prototypes feel more familiar than subjects expect them to, leading them to conclude that the words had been previously presented (Whittlesea et al., 2005). The statistically significant association between susceptibility to this effect and the severity of schizotypal beliefs and experiences suggest that prediction error-driven false fluency may provide the basis both for susceptibility to the illusion and to the unusual experiences and ideas that characterize schizotypy. That is, task-engendered false familiarity might have a common cognitive and neural basis with false familiarity experiences in daily life, experiences which manifest as perceptual aberrations and odd beliefs.

In previous work we found that larger prediction error responses (engendered by violating learned expectancies) predisposed healthy subjects to perceptual aberrations and delusional ideation when they were administered the psychotomimetic drug ketamine (Corlett et al., 2006). Specifically, a region of right dorsolateral prefrontal cortex (rDLPFC) appears to be sensitive to prediction error during learning (Fletcher et al., 2001; Corlett et al., 2004; Turner et al., 2004), and, the greater the magnitude of that response, the more likely subjects were to experience aberrant salience and referential ideas on ketamine (Corlett et al., 2006). While we did not have personality measures on these individuals it is noteworthy that false memories on the DRM task are also associated with augmented rDLPFC activation (Kim and Cabeza, 2007) and lesions of the DLPFC impair familiarity-based memory processing (Kishiyama et al., 2009). Taken together, these data imply that DLPFC over activity might engage

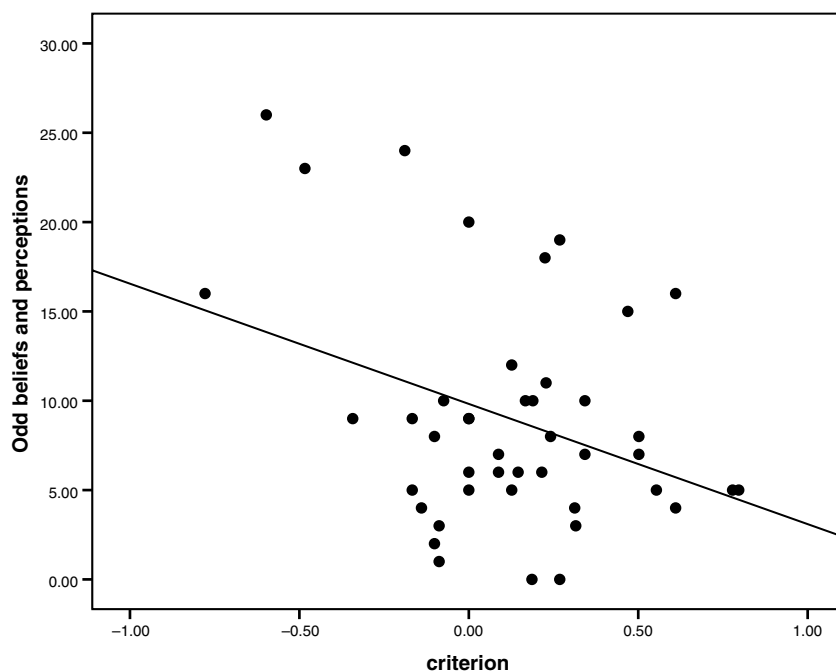
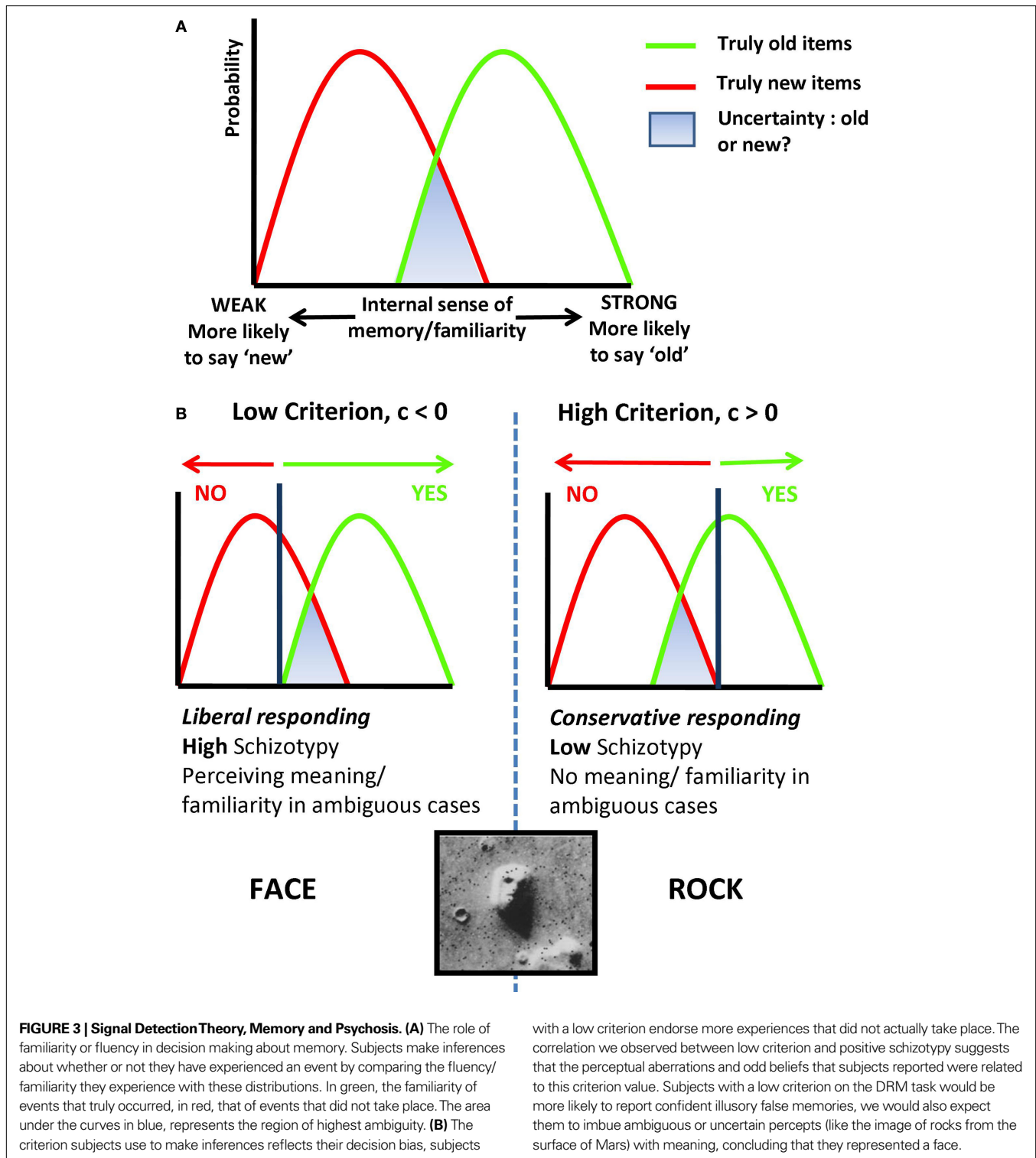


FIGURE 2 | The correlation between criterion and positive schizotypy: Scatter-plot of subject's criterion values and their susceptibility to odd beliefs. A low criterion or negative criterion corresponds to a liberal acceptance criterion or expectancy bias.



the prediction error-driven false familiarity that leads to DRM false memories as well as odd experiences and ideas. Further support for this comes from a subsequent observation that DLPFC prediction error activations are perturbed in individuals with early psychosis and the extent of that disruption is predictive of delusion severity across subjects (Corlett et al., 2007b).

The present finding is redolent of investigations of healthy individuals prone to hallucinations who show a similar bias toward reporting stimuli that did not occur (Bentall and Slade, 1985; Tsakanikos and Reed, 2005a,b; Reed et al., 2008). We show that a confirmatory or expectancy bias, possibly engendered by sensitivity to prediction errors (Whittlesea et al., 2005), relates not

only to perceptual aberrations but also delusion-like ideas. In what may be the earliest translational neuroscience theory of dopamine and positive psychotic symptoms, Robert Miller hypothesized that dopamine in the basal ganglia set a significance threshold on cognitive inferences, akin to a significance level in a statistical test; if dopamine were elevated in the brains of psychotic individuals, this threshold would decrease with two consequences; an increase in the total number of conclusions accepted, and an increase in the proportion of these which are, by chance, incorrect (Miller, 1976). We speculate that individual differences in dopamine release or regulation may underpin the relationships between susceptibility to the DRM illusion and attenuated psychosis-like symptoms that we report, that is, those subjects who reported more confident false memories, aberrant experiences and beliefs had sensitized dopamine systems which altered their threshold for concluding that a particular event occurred (see **Figure 3**).

The basal ganglia, in particular the striatum, are implicated in salience processing (Zink et al., 2003), prediction error-driven learning (McClure et al., 2003; O'Doherty et al., 2003) and aberrant salience theories of psychosis (Gray et al., 1991; Kapur, 2003). In previous work we found aberrant prediction error responses in the striatum, however, the signals in this region were not predictive of perceptual aberrations or delusional ideation (Corlett et al., 2006, 2007b). It is possible that DLPFC (the region whose inappropriate engagement we found to predict delusions) represents the confluence of inputs from a variety of regions, any noise from input regions like VTA and striatum drives maladaptive updating of expectations associated with delusions (Corlett et al., 2007b). However some patients with schizophrenia have DLPFC hypoactivity and increased striatal engagement (Meyer-Lindenberg et al., 2002). These conflicting studies may be examining different phases of psychosis or engaging different cognitive processes with different neural bases. Notably, neither the study of frontostriatal interactions (Meyer-Lindenberg et al., 2002) nor fMRI studies demonstrating aberrant striatal engagement during reward anticipation (Juckel et al., 2006) have demonstrated a relationship between inappropriate striatal signaling and delusions. Another study that explored the relationship between aberrant prediction error and delusions related delusion severity to the inappropriate engagement of frontal cortex (Schlagenhauf et al., 2009). Clearly the cognitive and neural mechanisms of frontostriatal interaction and their disruption in delusion formation is an important topic for future investigations.

We believe this false memory illusion and its relationship with aberrant beliefs and experiences favor a Bayesian model of processing in which expected and experienced fluency are compared in order to judge familiarity, the surprising fluency with which prototype words are processed predisposes subjects to conclude they are familiar (Benjamin et al., 1998). We argue that this Bayesian account extends to psychosis-like experiences. That is, certain experiences are inappropriately surprising and demand an explanation, with repeated surprising experiences, that explanation develops into an odd or inappropriate belief (Hemsley and Garety, 1986; Fleming, 1992; Corlett et al., 2009a; Fletcher and Frith, 2009). In this respect, it is noteworthy that subjects who were biased to endorse having experienced a word in the preceding RSVP list (criterion < 0) were more likely to have higher magical ideation and perceptual aberration.

Recent computational models in psychology and neuroscience recognize the overlap between Bayes theorem, SDT and associative learning (Courville et al., 2006; Dayan and Daw, 2008), in all types of theory, what is learned, remembered and perceived is governed not only by current inputs but also past experiences and expectancies (Alloy and Tabachnik, 1984). Our data support the notion that predictions and prediction errors interact in a Bayesian manner in the service of cognition and perception (Hemsley and Garety, 1986; Corlett et al., 2009a; Fletcher and Frith, 2009). Strong priors and strong prediction errors lead to false fluency of processing and an inappropriate feeling of familiarity that engenders perceptual aberrations and magical ideation. According to Estes, strong prior expectancy biases in memory can arise through repeated cycles of recollection and reconsolidation in which physiological noise vitiates the reconsolidated memory and biases it away from what actually occurred. We recently hypothesized that this mechanism may be responsible for the maintenance of delusional ideas and aberrant perceptions (Corlett et al., 2009b) we believe the observed association between DRM confidence, criterion bias and positive schizotypy provide initial support for that hypothesis.

Of course, our observation of the relationship between this putatively prediction error-driven illusion and everyday experiences and beliefs relates to healthy individuals. The relationship in individuals who suffer a mental illness has yet to be elucidated and so we must be cautious in our inferences. Nevertheless, the known relationships between schizotypy and psychotic illness should be emphasized and, moreover, studies such as this offer the invaluable opportunity to study experiences and beliefs that are not contaminated by the effects of medication. To our knowledge, the DRM task has not been studied under conditions of controlled pharmacological manipulation. However, there is evidence that behavioural and brain responses during memory and decision making tasks are modulated by dopamine (Pessiglione et al., 2006; Schott et al., 2006; Gibbs et al., 2007, 2008). Furthermore, the nature of the dopaminergic effect on decision making (whether dopamine agonism improves or impairs perceptual sensitivity) is dependent on personality traits such as belief in paranormal ideas (Krummenacher et al., 2009). We predict that pharmacological studies of the DRM effect will show that performance is modulated by genetic variability at dopaminergic loci and by administration of dopaminergic drugs. These pharmacological and genetic effects may interact with trait level vulnerability to psychosis-like ideas and experiences.

In summary, we have demonstrated an association between illusory memories, perceptual aberrations and odd beliefs. This finding lends support to the theory that inappropriate salience experiences engage an attribution process that leads to formation of aberrant but explanatory beliefs (Kapur, 2003). It is possible that inappropriate engagement of the right prefrontal cortex mediates false memory formation, odd perceptions and unusual beliefs by signaling inappropriate mismatches between expectancy and experience which engage new learning (Corlett et al., 2007a). While the formation of novel associations may have adaptive advantages (Stevens and Price, 2000; Avila et al., 2001), in its extreme form, hyperactivity of association ultimately leads to psychotic symptoms.

ACKNOWLEDGMENTS

PRC is a University of Cambridge Parke-Davis Exchange Fellow and a NARSAD Young Investigator, he is also supported by the University of Cambridge Isaac Newton Trust. JSS is supported by the BBSRC. GKM is supported by the MRC. PCF is supported by the Bernard Wolfe Health Neuroscience fund and the Wellcome Trust. This work was performed within the Behavioural and Clinical Neurosciences Institute, jointly supported by the Medical Research Council and The Wellcome Trust. JK is supported by NIMH, NIAAA and NARSAD, he reports the following: Consulting: AstraZeneca

REFERENCES

- Alloy, L. B., and Tabachnik, N. (1984). Assessment of covariation by humans and animals: the joint influence of prior expectations and current situational information. *Psychol. Rev.* 91, 112–149.
- Arieti, S. (1974). An overview of schizophrenia from a predominantly psychological approach. *Am. J. Psychiatry* 131, 241–249.
- Avila, M., Thaker, G., and Adami, H. (2001). Genetic epidemiology and schizophrenia: a study of reproductive fitness. *Schizophr. Res.* 47, 233–241.
- Bartlett, F. C. (1932). Remembering. A study in experimental and social psychology. Cambridge, Cambridge University Press.
- Benjamin, A. S., Bjork, R. A., and Hirshman, E. (1998). Predicting the future and reconstructing the past: a Bayesian characterization of the utility of subjective fluency. *Acta Psychol. (Amst)* 98, 267–290.
- Bentall, R. P., and Slade, P. D. (1985). Reality testing and auditory hallucinations: a signal detection analysis. *Br. J. Clin. Psychol.* 24(Pt 3), 159–169.
- Conrad, K. (1958). Die Beginnende Schizophrenie. G. Thieme. Stuttgart.
- Corlett, P. R., Aitken, M. R., Dickinson, A., Shanks, D. R., Honey, G. D., Honey, R. A., Robbins, T. W., Bullmore, E. T., and Fletcher, P. C. (2004). Prediction error during retrospective reevaluation of causal associations in humans: fMRI evidence in favor of an associative model of learning. *Neuron* 44, 877–888.
- Corlett, P. R., Frith, C. D., and Fletcher, P. C. (2009a). From drugs to deprivation: a Bayesian framework for understanding models of psychosis. *Psychopharmacology (Berl.)* 206, 515–530.
- Corlett, P. R., Krystal, J. K., Taylor, J. R., and Fletcher, P. C. (2009b). Why do delusions persist? *Front. Hum. Neurosci.* 3, 12.
- 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. (Oxford)* 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.
- Courville, A. C., Daw, N. D., and Touretzky, D. S. (2006). Bayesian theories of conditioning in a changing world. *Trends Cogn. Sci. (Regul. Ed.)* 10, 294–300.
- Crow, T. J. (1980). Positive and negative schizophrenic symptoms and the role of dopamine. *Br. J. Psychiatry* 137, 383–386.
- Dakin, S., Carlin, P., and Hemsley, D. (2005). Weak suppression of visual context in chronic schizophrenia. *Curr. Biol.* 15, R822–R824.
- Dayan, P., and Daw, N. D. (2008). Decision theory, reinforcement learning, and the brain. *Cogn. Affect. Behav. Neurosci.* 8, 429–453.
- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *J. Exp. Psychol.* 58, 17–22.
- Eckblad, M., and Chapman, L. J. (1983). Magical ideation as an indicator of schizotypy. *J. Consult. Clin. Psychol.* 51, 215–225.
- Emrich, H. M. (1989). A three-component-system hypothesis of psychosis. Impairment of binocular depth inversion as an indicator of a functional disequilibrium. *Br. J. Psychiatry Suppl.* 37–39.
- Estes, W. K. (1997). Processes of memory loss, recovery, and distortion. *Psychol. Rev.* 104, 148–169.
- Fleminger, S. (1992). Seeing is believing: the role of 'preconscious' perceptual processing in delusional misidentification. *Br. J. Psychiatry* 160, 293–303.
- Fletcher, P. C., Anderson, J. M., Shanks, D. R., Honey, R., Carpenter, T. A., Donovan, T., Papadakis, N., and Bullmore, E. T. (2001). Responses of human frontal cortex to surprising events are predicted by formal associative learning theory. *Nat. Neurosci.* 4, 1043–1048.
- 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.
- Fox, J. R. (2004). A signal detection analysis of audio/video redundancy effects in television news video. *Communic. Res.* 31, 524–536.
- Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 360, 815–836.
- Gibbs, A., Naudts, K., Spencer, E., and David, A. (2008). Effects of amisulpride on emotional memory using a dual-process model in healthy male volunteers. *J. Psychopharmacol.*
- Gibbs, A. A., Naudts, K. H., Spencer, E. P., and David, A. S. (2007). The role of dopamine in attentional and memory biases for emotional information. *Am. J. Psychiatry* 164, 1603–1609; quiz 1624.
- Gray, J. A., Feldon, J., Rawlins, J. N. P., Hemsley, D., and Smith, A. D. (1991). The neuropsychology of schizophrenia. *Behav. Brain Sci.* 14, 1–84.
- Green, D. M., and Swets, J. A. (1966). Signal Detection Theory and Psychophysics. Wiley: New York.
- Gregory, R. (1996). What are illusions? *Perception* 25, 503–504.
- Helmholtz, H. von (1878/1971). The facts of perception. In Selected Writings of Herman von Helmholtz, R. Kahl, ed. (Connecticut, Wesleyan University Press).
- Hemsley, D. R. (1994). Perceptual and cognitive abnormalities as the basis for schizophrenic symptoms. In The Neuropsychology of Schizophrenia, A. S. David, and J. C. Cutting, eds (Hove, Laurence Erlbaum Associates), pp. 97–118.
- Hemsley, D. R., and Garety, P. A. (1986). The formation of maintenance of delusions: a Bayesian analysis. *Br. J. Psychiatry* 149, 51–56.
- Holt, D. J., Titone, D., Long, L. S., Goff, D. C., Cather, C., Rauch, S. L., Judge, A., and Kuperberg, G. R. (2006). The misattribution of salience in delusional patients with schizophrenia. *Schizophr. Res.* 83, 247–256.
- Jaspers, K. (1963). General Psychopathology. Manchester University Press. Manchester.
- Jensen, J., Willeit, M., Zipursky, R. B., Savina, I., Smith, A. J., Menon, M., Crawley, A. P., and Kapur, S. (2008). The formation of abnormal associations in schizophrenia: neural and behavioral evidence. *Neuropsychopharmacology* 33, 473–479.
- Juckel, G., Schlagenhauf, F., Koslowski, M., Wustenberg, T., Villringer, A., Knutson, B., Wrase, J., and Heinz, A. (2006). Dysfunction of ventral striatal reward prediction in schizophrenia. *Neuroimage* 29, 409–416.
- Kandel, E. R. (1999). Biology and the future of psychoanalysis: a new intellectual framework for psychiatry revisited. *Am. J. Psychiatry* 156, 505–524.
- 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.
- Keefe, R. S., Arnold, M. C., Bayen, U. J., and Harvey, P. D. (1999). Source monitoring deficits in patients with schizophrenia: a multinomial modelling analysis. *Psychol. Med.* 29, 903–914.
- Kim, H., and Cabeza, R. (2007). Trusting our memories: dissociating the neural correlates of confidence in veridical versus illusory memories. *J. Neurosci.* 27, 12190–12197.
- Kishiyama, M. M., Yonelinas, A. P., and Knight, R. T. (2009). Novelty enhancements in memory are dependent on lateral prefrontal cortex. *J. Neurosci.* 29, 8114–8118.
- Krummenacher, P., Mohr, C., Haker, H., and Brugger, P. (2009). Dopamine, paranormal belief, and the detection of meaningful stimuli. *J. Cogn. Neurosci.* [Epub ahead of print].
- Loftus, E. F., and Palmer, J. C. (1974). Reconstruction of automobile destruction: an example of the interaction between language and memory.

- J. Verbal Learn. Verbal Behav.* 13, 585–589.
- Luo, C. R., Luo, C. R., Johnson, R. A., and Gallo, D. A. (1998). Automatic activation of phonological information in reading: evidence from the semantic relatedness decision task. *Mem. Cognit.* 26, 833.
- MacMillan, N., and Creelman, C. (1991). *Detection Theory: A User's Guide*. Cambridge, Cambridge University Press.
- Maher, B. A. (1974). Delusional thinking and perceptual disorder. *J. Individ. Psychol.* 30, 98–113.
- Matussek, P. (1954). Studies in delusional perception. In *The Clinical Roots of the Schizophrenia Concept: Translations of Seminal European Contributions on Schizophrenia*, J. C. Cutting and M. Shepherd, ed. (Cambridge, Cambridge University Press), pp. 89–103.
- McClure, S. M., Berns, G. S., and Montague, P. R. (2003). Temporal prediction errors in a passive learning task activate human striatum. *Neuron* 38, 339–346.
- McDermott, K. B., and Watson, J. M. (2001). The rise and fall of false recall: the impact of presentation duration. *J. Mem. Lang.* 45, 160.
- McGhie, A., and Chapman, J. (1961). Disorders of attention and perception in early schizophrenia. *Br. J. Med. Psychol.* 34, 103–116.
- Menon, M., Woodward, T. S., Pomarol-Clotet, E., McKenna, P. J., and McCarthy, R. (2005). Heightened stimulus salience renders deluded schizophrenics less susceptible to the 'famous names illusion'. *Schizophr. Res.* 80, 369–371.
- Meyer-Lindenberg, A., Miletich, R. S., Kohn, P. D., Esposito, G., Carson, R. E., Quarantelli, M., Weinberger, D. R., and Berman, K. F. (2002). Reduced prefrontal activity predicts exaggerated striatal dopaminergic function in schizophrenia. *Nat. Neurosci.* 5, 267–271.
- Miller, R. (1976). Schizophrenic psychology, associative learning and the role of forebrain dopamine. *Med. Hypotheses* 2, 203–211.
- Moritz, S., and Woodward, T. S. (2002). Memory confidence and false memories in schizophrenia. *J. Nerv. Ment. Dis.* 190, 641–643.
- 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.
- O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., and Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron* 38, 329–337.
- Pally, R. (2007). The predicting brain: unconscious repetition, conscious reflection and therapeutic change. *Int. J. Psychoanal.* 88, 861–881.
- Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., and Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* 442, 1042–1045.
- Peters, E. R., Joseph, S. A., and Garety, P. A. (1999). Measurement of delusional ideation in the normal population: introducing the PDI (Peters et al. Delusions Inventory). *Schizophr. Bull.* 25, 553–576.
- Rao, R. P., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87.
- Reed, P., Wakefield, D., Harris, J., Parry, J., Cella, M., and Tsakanikos, E. (2008). Seeing non-existent events: effects of environmental conditions, schizotypal symptoms, and sub-clinical characteristics. *J. Behav. Ther. Exp. Psychiatry* 39, 276–291.
- Roediger, H. L., and McDermott, K. B. (1995). Creating false memories: remembering words not presented in lists. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 803.
- Roiser, J. P., Stephan, K. E., den Ouden, H. E., Barnes, T. R., Friston, K. J., and Joyce, E. M. (2009). Do patients with schizophrenia exhibit aberrant salience? *Psychol. Med.* 39, 199–209.
- Schacter, D. L., Norman, K. A., and Koutstaal, W. (1998). The cognitive neuroscience of constructive memory. *Annu. Rev. Psychol.* 49, 289–318.
- Schlagenhauf, F., Sterzer, P., Schmack, K., Ballmaier, M., Rapp, M., Wrase, J., Juckel, G., Gallinat, J., and Heinz, A. (2009). Reward feedback alterations in unmedicated schizophrenia patients: relevance for delusions. *Biol. Psychiatry* 65, 1032–1039.
- Schott, B. H., Seidenbecher, C. I., Fenker, D. B., Lauer, C. J., Bunzeck, N., Bernstein, H. G., Tischmeyer, W., Gundelfinger, E. D., Heinz, H. J., and Düzel, E. (2006). The dopaminergic midbrain participates in human episodic memory formation: evidence from genetic imaging. *J. Neurosci.* 26, 1407–1417.
- Shapiro, M. (1994). Signal detection measures of recognition memory. In *Measuring Psychological Responses to Media Messages*, A. Lang, ed. (Hillsdale, NJ, Lawrence Erlbaum), pp. 133–148.
- Simons, J. S., Davis, S. W., Gilbert, S. J., Frith, C. D., and Burgess, P. W. (2006). Discriminating imagined from perceived information engages brain areas implicated in schizophrenia. *Neuroimage* 32, 696–703.
- Simons, J. S., Henson, R. N., Gilbert, S. J., and Fletcher, P. C. (2008). Separable forms of reality monitoring supported by anterior prefrontal cortex. *J. Cogn. Neurosci.* 20, 447–457.
- Stadler, M. A., Roediger, H. L. 3rd., and McDermott, K. B. (1999). Norms for word lists that create false memories. *Mem. Cognit.* 27, 494–500.
- Stevens, A., and Price, J. (2000). *Prophets, Cults and Madness*. Gerald Duckworth. London.
- Strahan, R., and Gerbasi, K. C. (1973). Semantic style variance in personality questionnaires. *J. Psychol.* 85, 109–118.
- Tsakanikos, E., and Reed, P. (2005a). Do positive schizotypal symptoms predict false perceptual experiences in non-clinical populations? *J. Nerv. Ment. Dis.* 193, 809–812.
- Tsakanikos, E., and Reed, P. (2005b). Seeing words that are not there: detection biases in schizotypy. *Br. J. Clin. Psychol.* 44, 295–299.
- Tulving, E. (1987). Multiple memory systems and consciousness. *Hum. Neurobiol.* 6, 67–80.
- Turner, D. C., Aitken, M. R., Shanks, D. R., Sahakian, B. J., Robbins, T. W., Schwarzbauer, C., and Fletcher, P. C. (2004). The role of the lateral frontal cortex in causal associative learning: exploring preventative and super-learning. *Cereb. Cortex* 14, 872–880.
- Whittlesea, B. W. (2002). False memory and the discrepancy-attribution hypothesis: the prototype-familiarity illusion. *J. Exp. Psychol. Gen.* 131, 96–115.
- Whittlesea, B. W., Masson, M. E., and Hughes, A. D. (2005). False memory following rapidly presented lists: the element of surprise. *Psychol. Res.* 69, 420–430.
- Whittlesea, B. W., and Williams, L. D. (1998). Why do strangers feel familiar, but friends don't? A discrepancy-attribution account of feelings of familiarity. *Acta Psychol. (Amst)* 98, 141–165.
- Zink, C. F., Pagnoni, G., Martin, M. E., Dhamala, M., and Berns, G. S. (2003). Human striatal response to salient nonrewarding stimuli. *J. Neurosci.* 23, 8092–8097.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 26 August 2009; paper pending published: 28 October 2009; accepted: 12 November 2009; published online: 24 November 2009.

Citation: Corlett PR, Simons JS, Pigott JS, Gardner JM, Murray GK, Krystal JH and Fletcher PC (2009) Illusions and delusions: relating experimentally-induced false memories to anomalous experiences and ideas. *Front. Behav. Neurosci.* 3:53. doi: 10.3389/fnro.08.053.2009

Copyright © 2009 Corlett, Simons, Pigott, Gardner, Murray, Krystal and Fletcher. This is an open-access article subject to an exclusive license agreement between the authors and the Frontiers Research Foundation, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.