Frontal Responses During Learning Predict Vulnerability to the Psychotogenic Effects of Ketamine

Linking Cognition, Brain Activity, and Psychosis

Philip R. Corlett, BA; Garry D. Honey, PhD; Michael R. F. Aitken, PhD; Anthony Dickinson, PhD, FRS; David R. Shanks, PhD; Anthony R. Absalom, MB, ChB, MD, FRCA; Michael Lee, MB, BS, FRCA; Edith Pomarol-Clotet, PhD, MRCPsych; Graham K. Murray, MRCPsych; Peter J. McKenna, PhD, MRCPsych; Trevor W. Robbins, PhD, FRS; Edward T. Bullmore, PhD, MRCPsych; Paul C. Fletcher, PhD, MRCPsych

Context: Establishing a neurobiological account of delusion formation that links cognitive processes, brain activity, and symptoms is important to furthering our understanding of psychosis.

Objective: To explore a theoretical model of delusion formation that implicates prediction error–dependent associative learning processes in a pharmacological functional magnetic resonance imaging study using the psychotomimetic drug ketamine.

Design: Within-subject, randomized, placebo-controlled study.

Setting: Hospital-based clinical research facility, Addenbrooke’s Hospital, Cambridge, England. The work was completed within the Wellcome Trust and Medical Research Council Behavioral and Clinical Neuroscience Institute, Cambridge.

Participants: Fifteen healthy, right-handed volunteers (8 of whom were male) with a mean ± SD age of 29 ± 7 years and a mean ± SD predicted full-scale IQ of 113 ± 4 were recruited from within the local community by advertisement.

Interventions: Subjects were given low-dose ketamine (100 ng/mL of plasma) or placebo while performing a causal associative learning task during functional magnetic resonance imaging. In a separate session outside the scanner, the dose was increased (to 200 ng/mL of plasma) and subjects underwent a structured clinical interview.

Main Outcome Measures: Brain activation, blood plasma levels of ketamine, and scores from psychiatric ratings scales (Brief Psychiatric Ratings Scale, Present State Examination, and Clinician-Administered Dissociative States Scale).

Results: Low-dose ketamine perturbs error-dependent learning activity in the right frontal cortex (P = .03). High-dose ketamine produces perceptual aberrations (P = .01) and delusion-like beliefs (P = .007). Critically, subjects showing the highest degree of frontal activation with placebo show the greatest occurrence of drug-induced perceptual aberrations (P = .03) and ideas or delusions of reference (P = .04).

Conclusions: These findings relate aberrant prediction error–dependent associative learning to referential ideas and delusions via a perturbation of frontal cortical function. They are consistent with a model of delusion formation positing disruptions in error-dependent learning.

Arch Gen Psychiatry. 2006;63:611-621

In this article, we attempt to understand how delusional beliefs may come about by relating them, in the setting of a drug model, to the neurobiology of associative learning. A number of theories of psychosis suggest aberrant learning processes as the basis for delusions. According to such theories, very early in delusion formation, subjects report that their attention is drawn to irrelevant environmental stimuli and that their capacity to form associations (between perceptions, thoughts, stimuli, and events) is markedly heightened. What might be the basis for such experiences and associations? One account is that they arise from inappropriate prediction error signals. Prediction error refers to a mismatch between expectancy and outcome. According to formal learning theories, prediction error guides adaptive behavior by diverting attention to critical environmental stimuli and driving causal associations between those stimuli and significant environmental events.
The neural instantiation of prediction error has been widely explored.\textsuperscript{21-29} The theory posits that noise in the prediction error system leads to inappropriate signaling of a mismatch,\textsuperscript{2,5} resulting in new association formation\textsuperscript{6-8,17,28} and the diversion of attention to potentially salient stimuli in the environment.\textsuperscript{4,8,18,19}

Although it is difficult to track the formation of delusions in patients, we nevertheless believe that an experimental analysis of this theory is tractable. Using the N-methyl-D-aspartate receptor antagonist ketamine, which may provide a useful model for psychosis\textsuperscript{30-32} and has also been increasingly explored in terms of cognitive changes\textsuperscript{30-46} and brain function,\textsuperscript{47-57} we sought to evaluate the suggested link between prediction error–dependent associative learning and the emergence of delusional beliefs.

The aim of this study was therefore to link prediction error–dependent causal associative learning to the very early stages of delusion formation using specific brain responses as a marker for such learning. Frontostriatal engagement, tracked using functional magnetic resonance imaging (fMRI), served as an assay of prediction error in healthy individuals, and we explored the behavior of this system in the setting of a causal learning task.\textsuperscript{29} First, we evaluated the brain response to prediction error–dependent learning with placebo in each subject. This was carried out using an empirically derived, regionally specific analysis.\textsuperscript{27-29} Next, we explored 3 interrelated questions relevant to the learning-based model of delusions. First, does ketamine, which is known to produce delusion-like beliefs in volunteers, act even at subpsychotic doses as a modulator of the brain response to prediction error? Second, with a high (psychotogenic) dose of ketamine, does the occurrence of delusional beliefs correlate with perceptual changes such as heightening of sensations and salience of environmental stimuli experienced by the subjects? Third, can the behavioral phenomena occurring at the high dose be predicted across individuals by brain responses signaling prediction error–dependent learning with placebo and low-dose ketamine?

**METHODS**

**SUBJECTS**

Fifteen healthy, right-handed volunteers (8 of whom were male) with a mean ± SD age of 29 ± 7 years and a mean ± SD predicted full-scale IQ\textsuperscript{48} of 113 ± 4 were recruited from the local community by advertisement, and they were screened using an initial telephone interview and subsequent personal interview. Psychiatric or physical illness, head injury, drug or alcohol dependence, and smoking were excluding factors, as were family history of psychiatric history and alcohol problems. The study was approved by the Cambridge Local Research Ethics Committee, Cambridge, England, and was carried out in accordance with the Declaration of Helsinki. Written informed consent was given by all of the subjects.

**EXPERIMENTAL DESIGN**

A double-blind, placebo-controlled, randomized, within-subjects design was used. On each visit, 2 phases of testing occurred. First, with administration of saline or low-dose ketamine (plasma target, 100 ng/mL), subjects performed the learning task in the fMRI scanner. After scanning, the subject was taken from the scanner and the saline administration was continued or the ketamine dose was increased to 200 ng/mL of plasma. Subjects then underwent a series of clinical interviews exploring the presence, nature, and severity of psychotic phenomena. The order of drug and placebo visits was counterbalanced across subjects and spaced by at least 1 month.

Although we also ran the learning task with high-dose ketamine, performance was inadequate and we will not report this part of the study further. Subjects also performed 3 other cognitive tasks while in the fMRI scanner. These will be reported in a separate article.

**INFUSION PROTOCOL**

Racemic ketamine (1 mg/mL) was administered intravenously by initial bolus and subsequent continuous target-controlled infusion using a computerized pump (Graseby 3500; Graseby Medical Ltd, Watford, United Kingdom) to achieve plasma concentrations of 100 or 200 ng/mL using the pharmacokinetic parameters of a 3-compartment model described by Domino et al.\textsuperscript{50}

**ASSOCIATIVE LEARNING TASK WITH PLACEBO AND LOW-DOSE KETAMINE**

We used the task reported by Corlett et al,\textsuperscript{29} who used a retrospective revaluation paradigm in which engendered expectations are violated to produce a prediction error. Subjects were asked to imagine themselves working as an allergist confronted with a new patient, “Mr X,” who has allergic reactions following some meals but not others. Their task was to work out which foods caused allergic reactions by observing the consequences of eating various foods. The task consisted of a series of trials, each of which had the general structure summarized in Figure 2. Trials comprised presentation of a food picture (representing a meal eaten by Mr X), a predictive response by the subject, and then an outcome. Responses allowed for both prediction and confidence outcome measures for each trial.

The experimental structure was identical to that used previously.\textsuperscript{29} The key manipulations relevant to the question under study are summarized in Figure 2. Each subject was trained concurrently on a number of different contingencies between foods and allergic reactions. Learning occurred over 3 stages: training, unovershadowing, and violation. This design is clarified with examples in Figure 2. In summary, expectancies were set up during a training phase, unovershadowing occurred, and we explored the impact of violation of the consequent expectancies on brain activity at the critical stage. We were then able to determine the impact of ketamine on these brain responses. We used parallel versions of the task so that no subject was exposed to the same foods on separate occasions. The same food was also not repeated across different conditions within the learning session.

**CLINICAL INTERVIEW WITH PLACEBO AND HIGH-DOSE KETAMINE**

Subjects underwent a shortened form of the Present State Examination (PSE)\textsuperscript{60} and the Clinician-Administered Dissociative States Scale (CADSS).\textsuperscript{61} This form comprises 19 self-reported ratings and 4 clinician-rated items (rated from 0 [not at all] to 4 [extremely]). The interviews were videotaped, and the videotapes were rated by 2 of us (E.P.-C. and P.J.M.). The videotapes of subjects’ responses were also rated on the 24-item version of the Brief Psychiatric Ratings Scale.\textsuperscript{62} The CADSS incorporates 5 subscales\textsuperscript{63}: body perception, environmental perception, feelings of unreality, memory impair-
ment, and time perception. Given the specific theoretical predictions regarding perceptual aberrations during the very early stages of psychosis precipitating psychotic phenomena, this investigation made use of the environmental perception subscale of the CADSS. We assessed the strength of correlation between ratings on this subscale and scores for delusional ideation, and time perception. Given the specific theoretical predictions regarding perceptual aberrations during the very early stages of psychosis precipitating psychotic phenomena, this investigation made use of the environmental perception subscale of the CADSS. We assessed the strength of correlation between ratings on this subscale and scores for delusional ideation, and time perception.

**FMRI DATA ACQUISITION AND ANALYSIS**

The fMRI data were acquired using a Bruker MedSpec 30/100 (Bruker Optics, Ettlingen, Germany) operating at 3 T. Gradient-echo echo-planar T2*-weighted images depicting blood oxygenation level–dependent contrast were acquired from 21 noncontiguous near-axial planes with the following parameters: repeat time, 1.1 seconds; echo time, 27.5 milliseconds; flip angle, 66°; in-plane resolution, 3.1 × 3.1 mm; matrix size, 64 × 64; field of view, 20 × 20 cm; bandwidth, 100 kHz. A total of 1608 volumes (21 slices each of 4-mm thickness with an interslice gap of 1 mm) per subject were acquired in 2 scans at each visit. The first 6 volumes from each scan were discarded.

The software SPM2 (available at http://www.fil.ion.ucl.ac.uk; Wellcome Department of Cognitive Neurology, London, England) was used. After realignment, spatial normalization, and spatially smoothing (gaussian kernel, 8 mm), the time series in each session were high-pass filtered (maximum of 1/120 Hz) and serial autocorrelations were estimated using an autoregression 1 model.

The average hemodynamic response to each event type was designated as occurring at the presentation of the outcome stimulus (ie, when subjects were informed whether an allergic reaction had occurred for that trial) and modeled using a canonical, synthetic hemodynamic response function. This function was used as a covariate in a general linear model, generating parameter estimates for each condition. Individuals’ contrast images for prespecified comparisons were produced by comparing parameter estimates across relevant conditions and were entered into group analyses treating intersubject variability as a random effect.

**IDENTIFICATION OF BASIC LEARNING EFFECTS**

Behavioral effects are summarized in Figure 3. A groupwise analysis was carried out to identify key regions responding to basic associative learning processes (all of the learning trials vs all of the fixation trials) during stage 1. In keeping with our regionally specific a priori hypotheses from previous normative data concerning frontostriatal systems, this analysis was confined to anatomically defined regions of interest (Figure 4A) identified by our previous study. These regions of interest were the bilateral substantia nigra, bilateral striatum (dorsal and ventral), and lateral right prefrontal cortex (rPFC). In a subsequent analysis, we extended our exploration to include bilateral hippocampi, particularly in light of the fact that mismatch theories of psychosis suggest that the hippocampus serves as the comparator, computing mismatches between expectation and experience. Brain loci showing learning effects within this predefined system (false discovery rate–corrected P < .05) were recorded and used to generate spheres (radius, 10 mm) that were tested in subsequent analyses. This use of analysis of regions of interest boosts sensitivity while protecting against type I error.

**Figure 1.** Trial structure. Each trial lasted 4 seconds in total and comprised 3 stages. A, Presentation of meal. A picture of a meal (1 or 2 foods) indicated the contents of Mr X’s meal for that trial. B, Predictive response. While the picture displaying the meal was on the screen, subjects predicted whether an allergy would occur using a 2-choice button box. When subjects made each prediction, they were instructed to hold the button down longer with the more confidence they felt. We used the combination of predictive response (positive vs negative) and confidence to obtain a measure of the strength of the subjects’ belief that a cue caused or did not cause the allergic reaction. The predictive response was calculated as R = time length of button push (where R is the predictive response coded as +1 for prediction of an allergy and –1 for prediction of no allergy). C, Outcome presentation. A display depicting the outcome of the meal was presented. A red jagged line encircling the words “allergic reaction” appeared to indicate that the patient had a reaction; if the patient had no reaction, a smooth green box appeared around the words “no reaction.”
Parameter estimates for each of the 4 conditions (violation events and well-learned control items for both drug and placebo) were extracted from the spheres centered on each region of interest. A repeated-measures analysis of variance was used to identify regions showing task and prediction error effects (violation vs control) and drug/task interactions at stage 3.

To assess the relationship between behavior, brain activity, and symptoms, we evaluated correlations (across subjects) between brain responses to prediction error and ratings on perceptual aberration and delusional ideation. The perceptual aberration score was taken from the environmental perception subscale of the CADSS. This included ratings on items pertaining to the qualities of auditory and visual perceptions, particularly their clarity. Referential ideas were rated using the PSE. Each of these scores was entered into a second-order regression model to determine whether psychopathological responses to a high dose of ketamine could be predicted by the extent of prediction error–dependent activation. Analyses were confined to regions of interest identified as described earlier. Regression analyses were thresholded using the false discovery rate at \( P < .05 \).

KETAMINE PLASMA LEVELS

The mean ± SD observed ketamine plasma levels for targets of 100 ng/mL and 200 ng/mL were 88 ± 10 ng/mL and 210 ± 13 ng/mL, respectively. (These data are based on 14 of the 15 subjects since we were unable to draw blood samples from 1 subject.)

Behavioral results are summarized in Figure 3. With both placebo and ketamine, subjects quickly learned to make correct positive and negative predictions during stages 1 and 2. Repeated-measures analyses of variance on the behavioral data from stages 1 and 2 revealed main effects of learning in both cases (stage 1: \( F_{1,110} = 4.04, P < .001 \); stage 2: \( F_{5,49} = 28.39, P < .001 \). There was no main effect of drug on either stage (stage 1: \( F_{1,120} = 0.68, P = .41 \); stage 2: \( F_{1,53} = 2.62, P = .14 \)), and there was no drug × task interaction (stage 1: \( F_{1,110} = 0.97, P = .36 \); stage 2: \( F_{5,20} = 1.12, P = .36 \)), suggesting no measurable impact of ketamine on behavioral performance.

Figure 3C shows the subjects’ first predictive responses to the unovershadowed items with placebo and
ketamine. This can be taken as an index of the extent to which they have indeed revalued the items that were absent at stage 2. There was a nonsignificant trend toward a difference in predictions about the unovershadowed item (banana in Figures 1, 2, and 3) between drug and placebo (2-tailed paired t-test, df = 1, 14, P = .09). This trend was seen as a reduced tendency for subjects to predict a positive link between food and allergy with low-dose ketamine, ie, a reduction in the magnitude of unovershadowing.

fMRI RESULTS

Activations During Stage 1 With Placebo

Acquisition of associative relationships significantly activated rPFC, bilateral caudate, ventral striatum, substantia nigra, and in the subsequent analysis, medial temporal cortex (Figure 4 and Table). As described earlier, these activations defined the loci for planned comparisons pertaining to the violation stage.

Brain Activations Accompanying Violation of Expectancies (Stage 3)

With placebo, prediction error was associated with significant activation in the rPFC as expected. That is, subjects activated this region relatively more when their expectancies about the unovershadowed item were violated than when confirmed. There was a significant drug × condition interaction (P = .03). Post hoc t tests (P < .05) revealed an overactivation of rPFC in response to nonsurprising control items on low-dose ketamine compared with placebo (P = .03) and a strong trend toward underactivation in response to expectancy violation on low-dose ketamine compared with placebo (P = .08) (Figure 4B).

Consistent with previous work, we found a significant main effect of task in the right striatum (caudate) (P = .01). Ketamine did not have a significant impact on the blood oxygenation level–dependent response in this region (P = .30) (Figure 4C). There was increased activity in the left substantia nigra for violation of learned expectancy relative to the well-learned control item (P = .047) (Figure 4D). There was no significant effect of ketamine on prediction error processing in this region (P = .30).

Although bilateral hippocampi showed activation in response to initial learning, only the left medial temporal lobe showed a significant activation in response to prediction error (P = .048). No significant effects of ketamine were found (P = .60).

In addition, the left caudate (P = .06) and right substantia nigra (P = .06) showed a trend toward a main effect of expectancy violation. We will not discuss these trends further.

Figure 3. Behavioral performance at the 3 stages of associative learning. A, Stage 1, training. The average behavioral responses to meals associated with an allergy (thick line) and no allergy (thin lines) are presented. Subjects’ predictive responses are charted (y-axis shows scores based on prediction and confidence as described in the Figure 1 legend). Evolving prediction from trial to trial (x-axis) is shown. An upward deflection reflects a tendency to predict with increasing confidence that a food will be associated with an allergic reaction. A downward deflection indicates negative predictions. B, Stage 2, unovershadowing. Plots show subjects’ predictive responses to single items not paired with an allergy that were previously presented as part of a pair during stage 1. Again, subjects’ predictive responses (y-axis) and evolving prediction from trial to trial (x-axis) are shown. C, Stage 3, predictive response to first presentation during stage 3 of unovershadowed items. This measure is taken as an estimate of the occurrence of unovershadowing during the previous stage. Data on placebo (P) and ketamine (K) are shown. Error bars indicate SEM for confidence ratings.
The higher dose of ketamine precipitated a typical profile of symptoms. Relevant here were changes in the clarity of thoughts and percepts (CADSS perceptual subscale, placebo vs drug, 2-tailed, paired t test, $P = .01$) and referential ideation (PSE simple ideas of reference subscale score, placebo vs drug, 2-tailed, paired t test, $P = .007$). Importantly, there was a significant correlation across subjects between profundity of perceptual aberration and strength of referential ideation (Pearson $r = 0.7$, 2-tailed $P < .001$). This is experimental evidence for a theoretical prediction of the model we are testing, that perceptual aberrations are a prelude to delusional ideas.
amine also induced subjectively inefficient thinking, depressed mood, expansive mood, and anxiety. However, there was no indication of correlation between these scores on the PSE and CADSS (inefficient thinking: Pearson \( r = -0.060, 2\)-tailed \( P = .56; \) depressed mood: Pearson \( r = 0.237, 2\)-tailed \( P = .40; \) expansive mood: Pearson \( r = -0.333, 2\)-tailed \( P = .16; \) and anxiety: Pearson \( r = -0.049, 2\)-tailed \( P = .86). \[8\]

**LINKING BRAIN ACTIVITY TO SUBSEQUENT KETAMINE-INDUCED SYMPTOMS**

Using regression analysis, we observed that both CADSS scores and PSE ideas of references subscale scores were predicted by the magnitude of rPFC response to prediction error with placebo (Figure 5). That is, those subjects who showed the greatest PFC response to violation of their expectancies experienced the greatest perceptual illusions and delusional ideation on the high dose of the drug. The relationship between the rPFC activity and perceptual aberration remained significant \((r_{15}=0.8, 2\text{-tailed } P = .002)\) when plasma levels were included in the analysis. Individual-specific attenuations by ketamine, relative to placebo, of PFC responses to prediction error did not predict CADSS or PSE scores. There was, however, a trend (uncorrected \( P = .002 \) for PSE scores, \(.007 \) for CADSS scores) for subjects showing the least prediction error–related response with low-dose ketamine to show higher CADSS and PSE scores.

**POST HOC ANALYSES**

Interestingly, the PFC responses with placebo during the training stage (stage 1) also predicted scores on the CADSS environmental perception subscale (uncorrected \( P < .001\). Furthermore, subjects’ basic behavioral learning rates during this stage (assessed by individuals’ slopes of change in predictive responses to successive trials) also predicted the rPFC response to prediction error (uncorrected \( P = .004\)).

Finally, other symptoms (negative symptoms and thought disorder, neither of which were theoretically implicated in the model) did not correlate with response to prediction errors in the rPFC system identified.

**COMMENT**

Our results provide behavioral and neurobiological evidence for the existence of a relationship between prediction error–dependent associative learning and the development of delusional beliefs with ketamine. First, perceptual anomalies such as heightened salience correlated with ideas or delusions of reference, consistent with the link between perception of environmental changes and delusions posited by the model. Second, lateral rPFC, as predicted by our data \[^{24,28,29}\] and indeed by related IMRI work, \[^{70}\] index prediction error, and this response is disrupted by low-dose ketamine. The nature of this disruption was consistent with the model in that the region no longer distinguished between predicted and unpredicted occurrences. Finally, subject-specific activity in precisely the same region during placebo administration was predictive of individuals’ likelihood of experiencing perceptual changes and ideas or delusions of reference (Figure 5). Thus, the observed links between brain activity, cognition, and subjective experience are consistent with the model under scrutiny.

We also draw attention to a degree of internal consistency whereby responses with placebo in precisely the same frontal region during the initial stage of basic associative learning were predictive of high-dose–associated symptoms. Moreover, frontal responses to learning at each stage were specific in their predictive power, while reliably predicting CADSS scores and referential beliefs, they did not predict other symptoms engendered by ketamine (negative symptoms, thought disorder, mood changes, or other positive symptoms).

It is interesting to note the nature of frontal modulation produced by low-dose ketamine (Figure 4). The reduction in the extent to which rPFC distinguished between violation and control stimuli was driven both by an attenuation of response to violations and an augmentation in response to stimuli, which should be highly predictable. This disturbance points to a 2-faceted change in prediction error signaling. First, the signal to true expectancy-outcome mismatches is diminished. This is perhaps unsurprising given that low-dose ketamine produces a trend toward a behavioral attenuation of the overshadowing effect. The other phenomenon driving this interaction, and one that is perhaps more directly relevant to the model under study, is that there is a ketamine-related augmentation of the rPFC response to well-learned events (Figure 4B). Taken together, our observations suggest that the perturbation of prediction error signaling may manifest as both decreases in appropriate signaling and increases in inappropriate signaling. The observations are consistent with the idea that delusions arise in the context of a failure to attend to and learn about salient stimuli while attention is increased toward stimuli that should be peripheral, familiar, or predictable.

There are 2 contrasting explanations for correlation between individuals’ “healthy” (placebo) responses to prediction error and subsequent psychopathological scores.

<table>
<thead>
<tr>
<th>Region</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>z Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prefrontal cortex</td>
<td>50</td>
<td>30</td>
<td>38</td>
<td>4.20</td>
</tr>
<tr>
<td></td>
<td>58</td>
<td>26</td>
<td>28</td>
<td>3.19</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>30</td>
<td>20</td>
<td>2.59</td>
</tr>
<tr>
<td>Caudate</td>
<td>-10</td>
<td>6</td>
<td>6</td>
<td>4.27</td>
</tr>
<tr>
<td></td>
<td>-10</td>
<td>14</td>
<td>-6</td>
<td>2.67</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>10</td>
<td>0</td>
<td>4.12</td>
</tr>
<tr>
<td>Midbrain</td>
<td>-10</td>
<td>-22</td>
<td>-10</td>
<td>4.35</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>-20</td>
<td>-10</td>
<td>4.16</td>
</tr>
<tr>
<td>Hippocampus</td>
<td>-16</td>
<td>-30</td>
<td>-6</td>
<td>3.42</td>
</tr>
<tr>
<td></td>
<td>-14</td>
<td>-6</td>
<td>0</td>
<td>3.06</td>
</tr>
<tr>
<td></td>
<td>-26</td>
<td>-26</td>
<td>-8</td>
<td>3.42</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>-32</td>
<td>-6</td>
<td>3.49</td>
</tr>
</tbody>
</table>

![Table. Regions Active During Acquisition of Associations on Placebo](http://archpsyc.jamanetwork.com/pdfaccess.ashx?url=/data/journals/psych/5253/ on 05/28/2017)
One is that greater levels of brain response signal some form of physiological inefficiency and that those subjects showing this are the ones most likely to show a prediction error disturbance when exposed to the ketamine challenge. The other possibility is that the magnitude of right prefrontal response indexes individual sensitivity to prediction error, with more sensitive individuals being most vulnerable to perturbation of the system. Ultimately, our key finding lies in the existence of a relationship, and we are in a position only to speculate on the source of individual vulnerability. However, relevant to this question is our observation that rPFC responses at the violation stage were significantly predicted by subjects’ baseline speed of learning estimated at stage 1. This admittedly post hoc observation counsels in favor of the latter view, that rPFC response indicates subjects’ sensitivity to prediction error–driven associative learning and that this sensitivity may be penalized with high-dose ketamine.

Another question is whether it is prediction error signal or perceptual change that has primacy in accounting for the fundamental change that ultimately engenders delusional beliefs. That is, does disrupted prediction error produce the perceptual change or vice versa? Again, we can only speculate on this, but although ketamine is well known to produce perceptual changes at higher doses, these changes in our study were largely expressed in terms of changes in the ways in which subjects felt about external stimuli rather than changes in perceptions per se. We argue that these changes are engendered by anomalies of prediction error signaling rather than that prediction error arises out of a change in perceptual processing. Further evidence in favor of this view comes from the observation that low-dose ketamine, which produced very little perceptual change, was associated with significant modulation of rPFC prediction error signal.

The idea that delusions may arise from disruptions in prediction error–dependent associative learning would, of course, predict disrupted associative learning in schizophrenia. Indeed, the early development of such models derived from observations of disturbed latent inhibition and blocking in people with schizophrenia. Importantly, reductions in blocking and latent inhibition suggest inappropriate persistence of prediction error firing in response to situations that should be predictable, an observation in keeping with the model. In schizophrenia, there is also evidence of disrupted prediction error...
indexing high-dose ketamine administration. This was for both practical and theoretical reasons. Primarily among the latter was that we do not aim to localize delusions themselves, believing fMRI to be extremely limited in telling us about the functional neuroanatomy of symptoms.

Rather, we use the fMRI signal as a brain marker for the core cognitive processes suspected to underpin the symptoms, and we explore the behavior of this signal both as a predictor of the symptoms and under a drug challenge that produces subbehavioral effect. Thus, we contend that this approach speaks much more directly to the cognitive model in question.

We should also draw attention to the fact that the reported interaction in rPFC between ketamine and condition (violation vs nonviolation control) did not significantly predict changes in the CADSS or PSE scores. We suggest that this is probably owing to the subtlety of such an interaction (i.e., a difference in a difference of activity as a predictor of a behavioral score). Whatever the explanation, we limit ourselves to stating that there is an individual pattern of vulnerability to ketamine predicted by learning-dependent PFC responses with placebo and that precisely the same region shows a groupwise attenuation by low-dose ketamine.

In summary, our study was motivated by the convictions that a satisfactory model of delusion formation must begin with an account of belief formation since delusions are aberrant beliefs and that the development of associations is fundamental to the formation of beliefs. For this reason, we sought to identify links between association formation, brain responses, drug effects, and symptoms. By inducing strong ideas of reference using ketamine, we used the learning of causal beliefs as a setting in which to explore a cognitive model of delusion formation. We demonstrate that a critical process in belief formation (prediction error) and its brain basis (rPFC) are disrupted by low-dose ketamine and that individual variations in task-related activity in this region correlate across individuals with their experience of perceptual aberrations and delusional ideation with a high dose of the drug. We believe that the combination of functional neuroimaging and psychopharmacology may be key in linking the 3 levels of explanation that would be critical to a truer understanding of psychosis: brain, cognition, and symptoms.

Submitted for Publication: September 7, 2005; final revision received October 28, 2005; accepted December 12, 2005.

Author Affiliations: Brain Mapping Unit, Department of Psychiatry, School of Clinical Medicine (Mr Corlett and Drs Honey, Pomarol-Clotet, Murray, McKenna, Bullmore, and Fletcher) and Department of Anaesthesiology (Drs Absalom and Lee), Addenbrooke’s Hospital, and Department of Experimental Psychology (Drs Aitken, Dickinson, and Robbins), University of Cambridge, Cambridge, England; and Department of Psychology, University College London, London, England (Dr Shanks).
REFERENCES


