OPEN

Molecular Psychiatry (2015), 1–10

www.nature.com/mp

ORIGINAL ARTICLE

Confidence and psychosis: a neuro-computational account of contingency learning disruption by NMDA blockade

F Vinckier[1,2,3,12], R Gaillard[1,4,5,12], S Palminteri[6,7], L Rigoux[2,3], A Salvador[1,5], A Fornito[8], R Adapa[9,10], MO Krebs[1,5], M Pessiglione[2,3,13] and PC Fletcher[4,11,13]

A state of pathological uncertainty about environmental regularities might represent a key step in the pathway to psychotic illness. Early psychosis can be investigated in healthy volunteers under ketamine, an NMDA receptor antagonist. Here, we explored the effects of ketamine on contingency learning using a placebo-controlled, double-blind, crossover design. During functional magnetic resonance imaging, participants performed an instrumental learning task, in which cue-outcome contingencies were probabilistic and reversed between blocks. Bayesian model comparison indicated that in such an unstable environment, reinforcement learning parameters are downregulated depending on confidence level, an adaptive mechanism that was specifically disrupted by ketamine administration. Drug effects were underpinned by altered neural activity in a fronto-parietal network, which reflected the confidence-based shift to exploitation of learned contingencies. Our findings suggest that an early characteristic of psychosis lies in a persistent doubt that undermines the stabilization of behavioral policy resulting in a failure to exploit regularities in the environment.

Molecular Psychiatry advance online publication, 9 June 2015; doi:10.1038/mp.2015.73

INTRODUCTION

One of the big challenges facing psychiatry is to develop an understanding of psychotic symptoms that goes beyond clinical description to uncover underlying computational and neurobiological mechanisms. A comprehensive account of the bizarre perceptions (hallucinations) and beliefs (delusions) that characterizes psychotic illness would require a mechanistic understanding of how the brain extracts and exploits regularities in the succession of events that occur in its environment. Reinforcement learning theory shows promise in this regard, by offering a framework within which we can consider causative disturbances at both the computational and neurobiological levels.[1–3] Such perspectives might therefore give us the sort of mechanistic understanding that can ultimately shape diagnostic and therapeutic questions.

Insights derived from reinforcement learning models have already proven useful in developing theoretical accounts of how psychotic experiences may arise and how they may relate to disrupted brain processes. Previous empirical studies have focused on how prediction error signaling may be deranged in psychosis.[4–8] Extending this several authors have suggested that the key deficit may reside not in prediction error *per se*, but rather in how prediction errors are used to update representations of the environment.[9,10] Of relevance, probabilistic learning tasks have been widely studied in schizophrenia (see refs. 11–13 for reviews), providing evidence for a complex pattern of deficit depending on the precise nature of the task (for example, complexity, occurrence and number of contingency reversals, explicit vs implicit learning) as well as of the profile of recruited patients (for example, predominantly positive vs negative symptoms, treated vs untreated patients). Interestingly, it has been proposed that the core impairment in schizophrenia might not affect learning ability *per se*, but rather the flexible control required to perform complex tasks and/or the capacity to optimize behavior in order to maintain a high level of performance.[11] In line with such proposals, our hypothesis is that a key feature of early psychosis is a disruption in how confidence is updated and used to drive behavior in a dynamic environment.

In situations of low confidence (or elevated uncertainty), individuals may seek explanations, exploring various possibilities in an effort to identify regularities. Indeed, it has been demonstrated that in such situations, healthy subjects tend to perceive illusory patterns, creating regularities where there are none, and providing superstitious or conspiratorial explanations for ambiguous scenarios.[14] These observations resemble the early features of psychosis, including sense of change and feeling of strangeness,[15–17] search for explanation,[18,19] apophenia[20] and jumping to conclusions.[21,22]

Here, we sought to capture this transitory state in the context of an associative learning task implementing a dynamic environment.

[1]Service de Psychiatrie, Centre Hospitalier Sainte-Anne, Université Paris Descartes, Sorbonne Paris Cité, Faculté de Médecine Paris Descartes, Paris, France; [2]Motivation, Brain, and Behavior Lab, Centre de Neuro-Imagerie de Recherche, Institut du Cerveau et de la Moelle épinière, Groupe Hospitalier Pitié-Salpêtrière, Paris, France; [3]INSERM U975, CNRS UMR 7225, UPMC-P6, UMR S 1127, Paris Cedex 13, France; [4]Department of Psychiatry and Behavioural and Clinical Neuroscience Institute, University of Cambridge, Cambridge, UK; [5]Laboratoire de "Physiopathologie des maladies Psychiatriques", Centre de Psychiatrie et Neurosciences U894, INSERM; Université Paris Descartes, Sorbonne Paris Cité, Paris, France; [6]Laboratoire de Neurosciences Cognitives (LNC), INSERM U960, Ecole Normale Supérieure (ENS), Paris, France; [7]Institute of Cognitive Neurosciences (ICN), University College London (UCL), London, UK; [8]Monash Clinical and Imaging Neuroscience, School of Psychological Sciences and Monash Biomedical Imaging, Monash University, Victoria, Australia; [9]Division of Anaesthesia, University of Cambridge, Cambridge, UK; [10]Addenbrooke's Hospital, Cambridge, UK and [11]Cambridge and Peterborough Foundation Trust, Cambridge, UK. Correspondence: Professor R Gaillard, Service de Psychiatrie, Centre Hospitalier Sainte-Anne, Université Paris Descartes, Sorbonne Paris Cité, Faculté de Médecine Paris Descartes, Paris, Centre Hospitalier Sainte Anne, 1 rue Cabanis, Paris 75014, France.
E-mail: raphael.gaillard@normalesup.org
[12]These authors contributed equally to this work.
[13]These authors co-directed this work.
Received 21 October 2014; revised 28 March 2015; accepted 13 April 2015

We predicted that, during learning of environmental contingencies, lack of confidence could lead to a reduced ability to stabilize an internal model of the world, with an ensuing, persistent sense of surprise. This would eventually result in sub-optimal behavior, characterized by an under-exploitation of true environmental regularities and an accompanying tendency to over-readily update in response to incidental violations of those regularities

Testing our predictions in a clinical setting is challenging given that, by the time psychosis is clearly identified, the expression of altered confidence may have been obfuscated by delusion formation and treatment effects. An established and fruitful solution is to use pharmacological models of early psychosis in healthy volunteers such as ketamine, a noncompetitive N-methyl-D-aspartate (NMDA) receptor antagonist[23,24] that induces subtle dissociative symptoms,[25] perceptual learning alterations and, critically, psychosis-like experiences (see[26] for a review). Here, we examined placebo-controlled, within-subject effects of a single dose of ketamine.

The task was adapted from previous paradigms.[27,28] On each trial, participants made a decision in response to a visual cue. The two options were always betting £1 versus betting 10p. The two options thus differed in risk, defined as the variance of possible outcomes. This does not imply that probability of winning was known, since it had to be learned by trial and error. This probability was 80% given one (positive) cue and 20% for the other (negative) cue. The optimal policy was to select the risky option following the positive cue and the safer option following the negative cue. To introduce instability into the environment, contingencies were reversed three times, such that the positive cue became the negative one and *vice-versa*. This task is close to tasks previously used to examine model learning under volatility (as in Behrens *et al.*[29]), except that transitions in probabilistic contingencies were not smooth but rather abrupt, as we wanted subjects to experience large variations in confidence, from the beginning to the end of learning blocks.

The key challenge posed to participants by our task was to notice unexpected outcomes that signaled a change in contingencies while ignoring those related to the probabilistic nature of these contingencies. Ignoring probabilistic errors requires confidence in the estimates of experimental regularities. Thus, we hypothesized that ketamine would prevent subjects from ignoring probabilistic errors, leading to sub-optimal behavior at the end of learning blocks, where subjects under placebo would fully exploit the learned contingencies. We explored the neural underpinnings of this ketamine-induced dysfunction, with the prediction that activity in confidence-related brain areas would show altered dynamics during the course of learning. Neural responses were concurrently tracked using functional magnetic resonance imaging (fMRI), while subjects performed the probabilistic contingency learning task. Each participant underwent this procedure during both ketamine and placebo infusions.

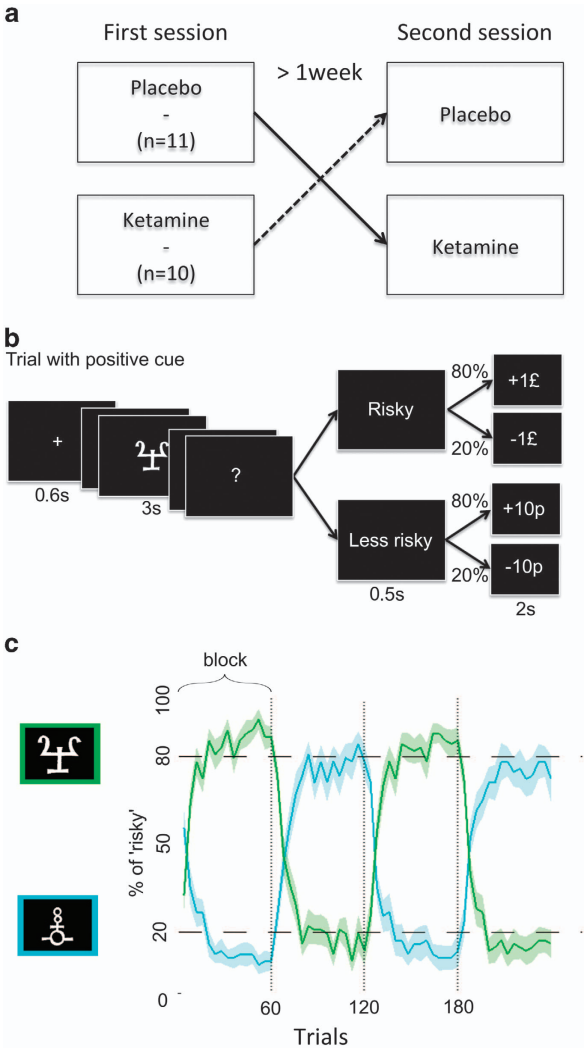

Figure 1. Experimental design. (**a**) A double-blind, placebo-controlled, randomized, within-subject design was used. The order of drug and placebo visits was counterbalanced across subjects and spaced by at least 1 week. (**b**) A typical trial and possible outcomes for a positive cue. Probabilistic contingencies (80 and 20%) would be swapped for a negative cue. (**c**) Percentage of risky responses as a function of trial number (both placebo and ketamine sessions were pooled). The green (respectively, blue) curve represents the choices following the cue that was positive (respectively, negative) in the first block. Bold lines represent means; color-delimited areas represent inter-subject s.e.m (corrected for the variance across subject: the grand mean of each subject was removed from its data before computing sem). Vertical dashed lines indicate reversals.

MATERIALS AND METHODS

Subjects

Twenty-one healthy, right-handed volunteers (11 males), aged 25–37 years (mean 28.7, s.d. 3.2), were recruited from the local community by advertisement, and screened using an initial telephone interview and subsequent personal interview. Exclusion criteria were: personal/familial history of neurological or psychiatric disorders, MRI contra-indications, illicit substance use in the last 12 months or any lifetime substance misuse syndrome or alcoholism, history of cardiac illness or high blood pressure, weight >10% above ideal body mass index. 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.

Ketamine infusion

Racemic ketamine (2 mg ml^{-1}) was administered intravenously by initial bolus and subsequent continuous target-controlled infusion using a computerized pump (Graseby 3500; Graseby Medical, Watford, UK) to achieve plasma concentrations of 100 ng ml^{-1} using the pharmacokinetic parameters of a three-compartment model.[30] One blood sample was drawn prior to the fMRI scan. Blood sample was placed on ice, plasma obtained by centrifugation and plasma samples stored at −70 °C. Plasma ketamine concentration was measured by gas chromatography-mass spectrometry.

Experimental design

A double-blind, placebo-controlled, randomized, within-subjects design was used (see Figure 1a). At each visit, after starting the infusion of saline or low-dose ketamine, subjects underwent a clinical rating of positive psychotic symptoms as assessed by the Rating Scale for Psychotic Symptoms.[31] Seven key items on the Brief Psychiatric Rating Scale[32] representing symptoms of the psychosis prodrome (somatic concerns,

anxiety-depression, elevated mood, grandiosity, hallucination and unusual thought content) were also assessed. Dissociative symptoms were assessed by the Clinician Administered Dissociative States Scale.[33] Subjects then performed the probabilistic learning task in the fMRI scanner. Subjects also performed two other cognitive tasks while in the fMRI scanner. These were perceptual tasks not related to the current task and will not be reported here. Resting state data were also acquired.[34]

Behavioral task

The task (see Figure 1) required participants, on each trial, to make a choice between a more and less risky option, indicating their choice by pressing a key or not. Risk taking was orthogonalized with respect to the motor dimension, so that pressing the key was assigned to the risky response only for half of participants and to the less risky response for the other half.

The risky ('risk' being defined as the variance of the outcome) choice would lead to either the gain or the loss of £1, while the less risky option would lead to either the gain or loss of 10 pence. There were two contextual cues. One was associated with 80% chance of winning £1 (and a corresponding 20% chance of losing £1) following the risky choice and with 80% chance of winning 10 pence (and a 20% chance of losing 10 pence) following the less risky choice. For the other cue the contingencies were the opposite, that is, the risky choice would lead to an 80% chance of losing £1, while the less risky choice gave an 80% chance of losing 10 pence.

An unannounced contingency reversal occurred after each block of 60 trials (for a total of three reversals across the 240 trials). Reversal means that the positive cue (for which the risky choice was optimal) became the negative one and vice-versa. Therefore participants encountered the same contingency set only twice during the experiment.

Two abstract cues randomly taken among 24 letters from the Agathodaimon font were used. After fixation delay and cue display, the response interval was indicated on the computer screen by a question mark. The interval was fixed to 3 s and the response was taken at the end: this response was categorized as 'risky'or 'less risky' and was written on the screen as soon as the delay had elapsed. Monetary outcome was then displayed for 2 s. Participants were explicitly told that they would not receive the virtual money earned during the task. Instead, they were paid a fix amount that compensated for their time and their expenses associated with taking part in the study.

Before performing the task in the scanner, participants were familiarized with the task structure and with the notion that cue-outcome relationships were not necessarily constant. However, they were not warned that contingencies could be reversed.

Model-free behavioral analysis

The overall percentage of risky response and button presses were compared between sessions in order to assess drug effects on choice and motor impulsivity, respectively. To assess drug effects on learning, the percentage of optimal responses (risky choice for the positive cue, less risky choice for the negative cue) were collapsed across the two cues and averaged within six bins of 10 consecutive trials. These data were then submitted to repeated-measure analysis of variance with three experimental factors (bin*block*session) and subjects as random factor. *Post-hoc* comparisons were performed to characterize the learning deficit observed under ketamine.

Model-based behavioral analysis

The whole model space consisted of 27 models (see SOM): three variants of the reinforcement learning level without any confidence monitoring plus 24 variants of the hierarchical model (three reinforcement learning models × two ways to compute confidence × four ways to modulate low-level parameters) (see Figure 3 for a more detailed description of model space).

All models were inverted using a variational Bayes approach under the Laplace approximation,[35–37] http://sites.google.com/site/jeandaunizeaus website/). This algorithm not only inverts nonlinear models but also estimates their evidence, which represents a trade-off between accuracy (goodness of fit) and complexity (degrees of freedom). The log-evidences estimated for each participant and model were submitted to a group-level random-effect analysis separately for placebo and ketamine sessions. To complete model selection, we also performed family analyses.[37]

fMRI data analysis

fMRI data were preprocessed and statistically analyzed using SPM5 toolbox (Wellcome Department of Cognitive Neurology, London, UK) running on Matlab (Mathworks). T1-weighted structural images were coregistered with the mean functional image, segmented, and normalized to a standard T1 template and averaged across all subjects to allow group-level anatomical localization. The first five volumes of each session were discarded to allow for T1 equilibration effects. Preprocessing consisted of spatial realignment, normalization using the same transformation as structural images, and spatial smoothing using a Gaussian kernel with a full-width at half-maximum of 8 mm.

We devised two general linear models (GLM) to account for individual time series. The first GLM included separate categorical regressors for cue and outcome onsets, respectively, modulated by the computational variables, βm and αm. As parametric modulators were applied to different categorical regressors, they were not orthogonalized to each other. Note, however, that their correlation was quite low ($R^2 = 0.1$) In the second GLM, outcome onsets were modulated by two computational variables, outcome category (confirmatory vs contradictory) and αm, that were serially orthogonalized, following on SPM default procedure. This second GLM was exclusively used for the region of interest (ROI) analysis. These variables were computed using subject-specific free parameters of the best fitting computational model (see computational results) and were then *z*-scored. All regressors of interest were convolved with a canonical hemodynamic response function. To correct for motion artifacts, subject-specific realignment parameters were modeled as covariates of no interest. Linear contrasts of regression coefficients were computed at the subject level and then taken to group-level random effect analyses.

Neural correlates of choice temperature and learning rate were identified in placebo sessions using a whole-brain one-sample *t*-test (cluster generating threshold $P < 0.001$ uncorrected, cluster level threshold $P < 0.05$ family-wise error corrected). The impact of ketamine on these networks was assessed using a paired *t*-test between ketamine and placebo sessions (cluster generating threshold $P < 0.01$ uncorrected, cluster level threshold $P < 0.05$ family-wise error corrected). In order to maximize sensitivity and to ensure that drug effects were only assessed within task-relevant networks, this analysis was masked by the parametric modulations (by choice temperature or learning rate) obtained when pooling placebo and ketamine sessions.

For ROI analyses, we extracted the regression estimates (betas) from spheres of 8mm in diameter (corresponding to the full-width at half-maximum of the Gaussian kernel used for spatial smoothing), centered on group-level activation peaks. The ventromedial prefrontal cortex (vmPFC) ROI, that was used to perform a comparison between placebo and ketamine session, was defined from the second-level analysis pooling both placebo and ketamine sessions in order to avoid biasing this comparison in favor of placebo sessions.

Additional GLMs were computed for illustrative purpose only. In these GLMs, trials were sorted in six bins of confidence (as defined in the best computational model) or trial number in a block (as in the model-free analysis: the first ten trials of each block, the following ten and so on). These GLMs were used to plot the hemodynamic response at cue and outcome onsets.

RESULTS

Clinical assessments

The mean blood plasma concentration of ketamine during infusion was 96.01 ± 19.11 ng ml^{-1}. Paired *t*-tests indicated that ketamine caused a significant increase in positive psychotic symptoms as measured by the Rating Scale for Psychotic Symptoms ($t(20) = 5.43$, $P < 0.001$) and the Brief Psychiatric Rating Scale ($t(20) = 2.8$, $P = 0.011$), as well as in dissociative symptoms as measured by the Clinician Administered Dissociative States Scale ($t(20) = 3.72$, $P = 0.0013$).

Behavioral results

Choice and motor impulsivity did not differ between drug conditions (risky choice: 48.2% vs 47.8%, $t(20) = 0.29$, $P = 0.8$; button press: 53.0% vs 51.7%, $t(20) = 1.05$, $P = 0.3$). There was a main effect of learning, with optimal choices increasing across bins ($F(5,100) = 66.77$, $P < 0.001$), a main effect of block ($F(3,60) = 4.57$, $P < 0.01$) with more optimal choices during the first (pre-reversal)

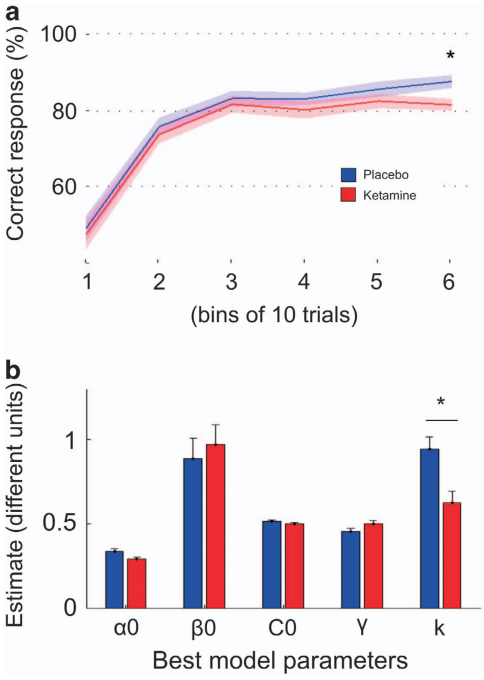

Figure 2. Characterization of the behavioral deficit induced by ketamine (**a**) Learning curves. Curves show percentage of correct response average across blocks, cues and bins of 10 consecutive trials, for the placebo (blue) and ketamine (red) sessions, separately. There was a significant effect of drug status in the last trial bin, with higher performance with placebo. Bold lines represent means; color-delimited areas represent inter-subject s.e.m (corrected for the variance across subject: the grand mean of each subject was removed from its data before computing s.e.m.). (**b**) Parameter estimates for the best computational model. The only parameter that significantly differed between sessions (placebo in blue versus ketamine in red) was κ, the weight that confidence had on learning rate and choice temperature. $\alpha 0$: learning rate value when confidence = 0; $\beta 0$: choice temperature value when confidence = 0; C0: initial confidence value; γ: confidence learning rate. Bars represent means; error bars represent inter-subject s.e.m (corrected for the variance across subject: the grand mean of each subject was removed from its data before computing s.e.m.); $^*P < 0.05$, two-tailed paired *t*-test.

block (80%) compared with others (74%). There was no other main effect and no interaction between factors (all $P > 0.1$). *Post-hoc* analysis showed a significant effect of drug status in the last trial bin (see Figure 2a), with higher performance under placebo ($F(1,20) = 5.641$, $P = 0.028$) without main effect nor interaction with block (both $P > 0.1$). Indeed, during ketamine infusion, participants apportioned their responses in a way that matched or slightly exceeded the 80% probability of positive reinforcement (81.1%, $t(20) = 0.37$, $P = 0.7$ in comparison with 80%). In contrast, they optimized their behavior under placebo (87.2%, $t(20) = 2.52$, $P = 0.02$ compared to 80%). In summary, this preliminary behavioral analysis suggests that ketamine reduced the ability to go beyond probability matching, that is, to stabilize behavior in the face of probabilistic (misleading) unexpected outcomes. This hypothesis was formally assessed by using computational modeling.

Computational modeling results

To explore a comprehensive set of possible strategies, we fitted qualitatively different models to the observed choices (see SOM for details). All models estimate the trial-wise values attached to the two cues, and use these values to predict choices, through a softmax function.

A first series of models were designed to account for low-level reinforcement learning. Following a standard 'delta' rule,[38] these models update after each trial the current cue value in proportion to prediction error, defined as the outcome value minus the expected value.

In a basic version, the outcome was simply the monetary amount (+£1, +0.1£, −0.1£ or -1£). In a second version, we integrated some understanding of the task structure by including the possibility that cue values were coded at a more abstract level, as if subjects figured out that all the information needed was the outcome valence (+ or −). In a third version the two cue values were updated after every outcome, to model the possibility that subjects realized that they always had an opposite valence, that is, information about the status of one cue also gave information about the status of the other.

Reinforcement learning models have constant parameters (learning rate α and choice stochasticity β). This limits the capacity to optimize the behavioral policy around the end of learning blocks, once subjects believe themselves to have a reasonably good estimation of contingencies. At this point, prediction errors should be tempered, and choices tuned to a more deterministic exploitation of learned contingencies.[29,39,40] Conversely, when contingencies suddenly change after reversals, prediction errors should be given more weight, and choices should be more exploratory. This can be implemented in an optimal way using a hierarchical Bayesian architecture.[29,40] Some evidence has been found that human behavior can be accounted for by hierarchical Bayesian models.[41,42] However, Bayesian updates of probability distributions may become computationally cumbersome, and human subjects sometimes follow simpler heuristics, particularly when they are uncertain about the task structure.[43,44] Another way to optimize behavior is to subordinate the reinforcement learning parameters to a higher level of control that monitors performance. This idea has been proposed and formalized in the so-called meta-learning theoretical framework,[45] which addresses the question of how machines can learn how to learn. This principle has been implemented for instance to adjust the exploration rate during the course of learning, and provides a good fit of nearly optimal primate behavior.[46,47]

A second series of models followed this latter principle: they included a meta-cognitive level consisting in updating confidence (the belief that current representations are correct) so as to downregulate contingency learning and choice stochasticity. These hierarchical models allowed us to determine more precisely which level of learning was altered by ketamine infusion. Confidence was monitored using a delta rule in all the following models, which differed in the way outcomes were used to assess performance. A first variant used the absolute value of the prediction error generated in the lower reinforcement learning level, implementing the intuition that subjects should be more confident when prediction errors are reduced.[48,49] A second variant (following Khamassi *et al.*[47]) coded the outcome in terms of optimality: 0 for non-optimal outcomes (losing £1 or winning only 10p) and 1 for optimal outcomes (winning £1 or losing only 10p). In both variants, confidence could be used to modulate learning rate (αm), choice temperature (βm) or both, with different or identical weight. Optimizing choice temperature means favoring exploitation when confidence increases. Optimizing learning rate means increasing sensitivity to confirmatory outcomes and decreasing sensitivity to contradictory outcomes when confidence increases. Confirmatory means that the valence of the outcome is the same as the valence estimated by the model. Thus, when confidence was close to 0, the learning rate was similar for confirmatory and contradictory outcomes, but as confidence increased, it got closer to 1 for confirmatory outcomes and to 0 for contradictory outcomes.

Bayesian model selection was performed separately for placebo and ketamine sessions (see Figure 3 and Figure 4). The best model

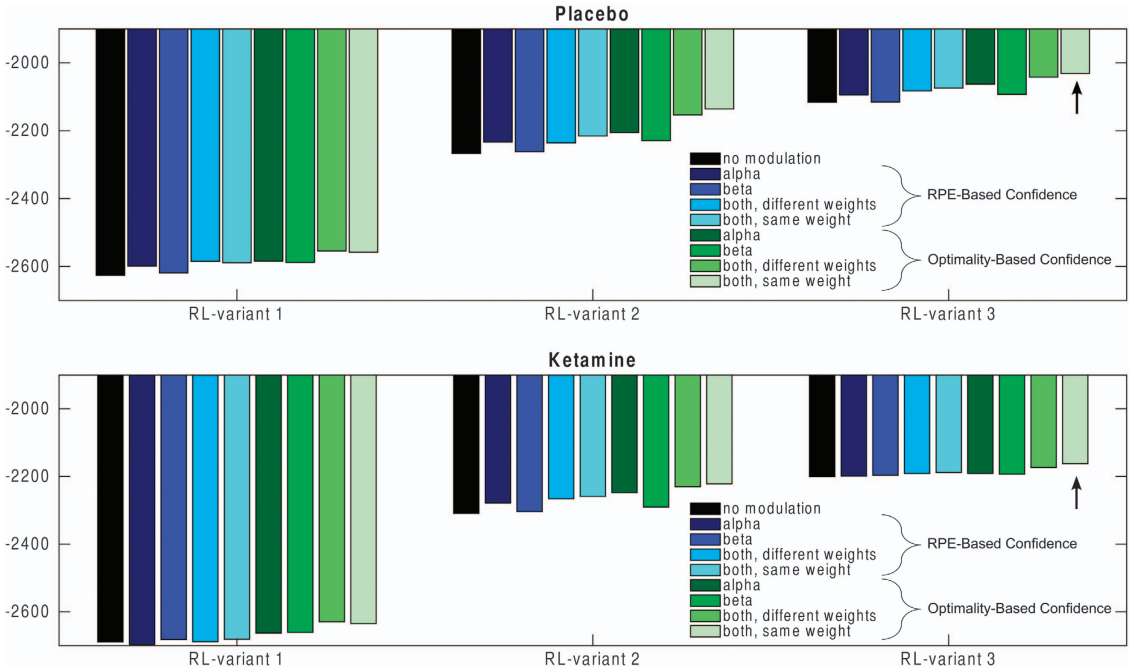

Figure 3. Model evidence (variational Bayesian approximation to marginal likelihood). The structure of the model space can be divided as follows. (i) Which RL variant? In variant 1, the reinforcer was the monetary amount; in variant 2, the reinforcer was the sign of the outcome; in variant 3, the reinforcer was the sign of the outcome and the two cue values were updated after every outcome. (ii) How to compute confidence? In a first variant, it was based on the absolute value of the prediction error. In a second variant, it was based on the optimality of the outcome, that is, 0 for non-optimal outcomes (losing £1 or winning only 10p) or 1 for optimal outcomes (winning £1 or losing only 10p). (iii) How to use confidence? Confidence was used to modulate the learning, choice temperature or both (with same or different weight). The arrow indicates the best model. Note that even the difference between the two rightmost bars is >10, which is considered to be a very strong difference in model evidence.

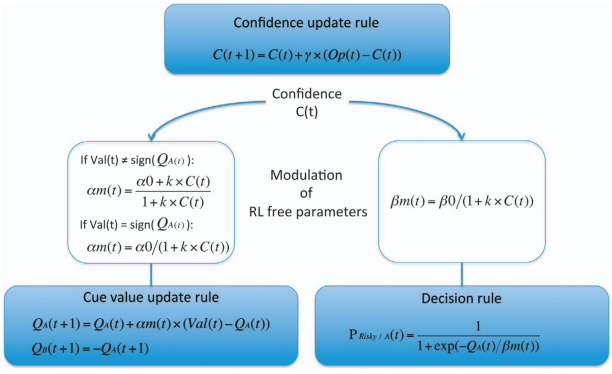

Figure 4. Description of the best model. The best model was selected using a group-level random-effect analysis. It included the third variant of RL (as if subjects figured out that only the outcome valence, and not the monetary amount, was informative about cue value, and that the two cues always had opposite valence such that they could both be updated after every outcome). Confidence was based on outcome optimality and used to modulate both the learning rate and choice temperature, with a same weight. Q is cue value; C is confidence; Op is outcome optimality (1 for winning £1 or losing 10p, -1 otherwise); Val is outcome valence (1 if positive, −1 otherwise); *P* risky/A is the probability of choosing the risky option when cue A is on screen. γ is confidence learning rate; $\alpha 0$ is learning rate value when confidence = 0; $\beta 0$ is choice temperature value when confidence = 0; κ is the weight of confidence on learning rate and choice temperature.

was the same in both sessions but the evidence was higher for placebo (xp = 0.96; Supplementary Table S1) than for ketamine (xp = 0.45; Supplementary Table S2). At the low level, this best model implemented an informed reinforcement learning rule, using the outcome valence (+ or −) to update the two cue values. At the high level, confidence was updated using the outcome optimality, and impacted both learning rate and choice temperature, with identical weights. Family model comparison[37] confirmed that the best model was the same in both sessions though in ketamine sessions there was less clear evidence for the necessity of a meta-cognitive level that monitors confidence and allows confidence to modulate low-level parameters (see SOM for details).

We next compared the free parameters of this best model between placebo and ketamine sessions, with paired-tests (Figure 2b, Supplementary Table S3). The parameter that significantly differed between sessions was the weight that confidence had on learning rate and choice temperature ($t(20) = 2.3$, $P = 0.027$). Thus, ketamine reduced the impact of confidence on low-level parameters. This attenuation could therefore explain the deleterious effect of the drug on ability to optimize behavior when confidence increases, towards the end of learning blocks.

Neuroimaging results

The computational analysis demonstrated that the behavioral effects of ketamine were underpinned by a shift in the dynamics of choice temperature and learning rate (βm and αm), which were insufficiently tuned by the confidence increases within learning blocks. To identify the underlying neural effects, we therefore focused on the neural representation of βm and αm, which, in principle, should be used to make choices at cue onsets and to update values at outcome onsets respectively. For each time point (cue and outcome onsets), we first analyzed the placebo session to identify the neural representation of βm or αm in the normal brain. We then directly compared placebo and ketamine sessions.

At choice onset. Under placebo, βm was correlated with activity in a large fronto-parietal network, including dorsomedial prefrontal cortex (dmPFC), frontopolar cortex and bilateral lateral

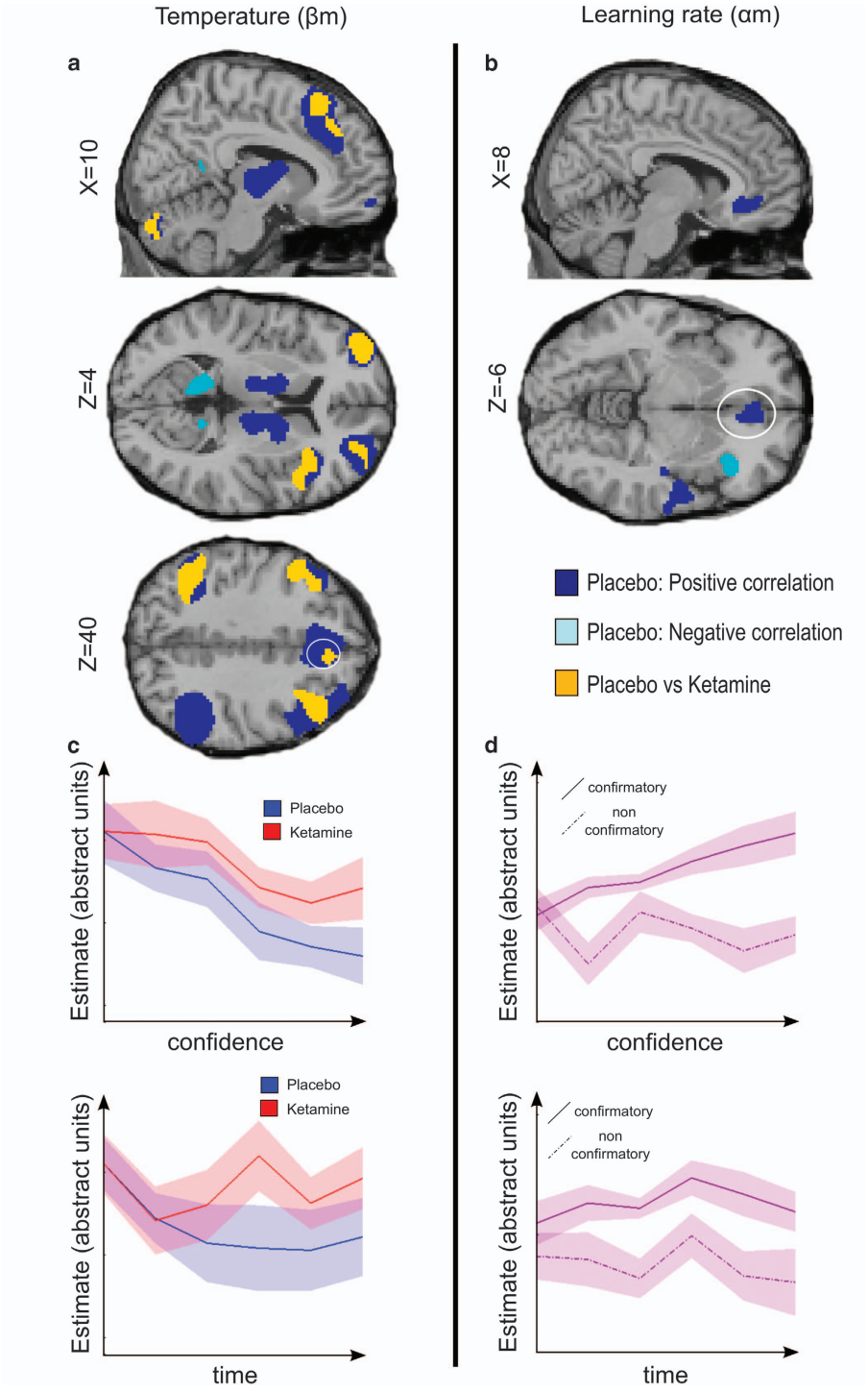

Figure 5. Model-based analysis of ketamine-induced changes in brain activity (**a**) Brain regions reflecting confidence-modulated choice temperature (βm). (**b**) Brain regions reflecting confidence-modulated learning rate (αm). Colored clusters show significant correlation in the placebo session (positive in dark blue, negative in light blue) and significant difference between placebo and ketamine sessions (in orange). All clusters survived a statistical threshold of $P < 0.05$ after family wise error correction for multiple comparisons. Coordinates of anatomical slices are given in Montreal Neurological Institute space. (**c**) Hemodynamic response to cue onset in the dmPFC as a function of confidence bins or as a function of time (trial number, pooled across blocks) (for illustrative purpose). (**d**) Hemodynamic response to outcome onset in the ventromedial prefrontal cortex as a function of confidence bins or as a function of time (trial number, pooled across blocks) (for illustrative purpose), shown separately for confirmatory and contradictory outcomes. Placebo data are in blue, ketamine data are in red, pooled data are in violet. Bold lines represent means; color-delimited areas represent inter-subject s.e.m (corrected for the variance across subject: the grand mean of each subject was removed from its data before computing s.e.m.).

Table 1. Brain regions reflecting confidence-modulated choice temperature (βm)

	Structure	*MNI coordinates (x, y, z)*	*Z-score*
Positive	Dorsomedial prefrontal cortex	−4, 24, 48	4.83
		48, 20, 52	4.78
	Dorsolateral prefrontal cortex	48, 28, 42	5
		36, 52, 24	4.31
		−28, 56, 26	4.18
		−46, 14, 44	4.26
		−38, 30, 42	4.24
		36, 54, −4	3.91
	Frontopolar cortex	14, 60, −10	4.66
		−42, 48, −4	4.05
		−22, 58, −6	4.49
		50, −52, 46	4.01
	(Inferior) parietal cortex	−42, −52, 42	4.73
		34, 22, −12	4.09
	Anterior insula	38, 22, 6	3.99
		−36, 18, −6	3.96
		40, −66, −50	3.74
	Cerebellum	−28, −74, −36	4.72
		16, 2, 14	4.6
	Caudate nucleus	10, −12, 0	4.17
	Thalamus	−8, −14, 0	4.35
Negative	Precuneus	−12, −48, 12	4.31
	Posterior cingulate	22, −52, 20	4.25
		−12, −58, 22	3.93
	Cuneus	22, −88, 26	3.41
	Medial temporal lobe	36, −38, −2	4.55
		−40, −36, −6	3.53
	Post-central gyrus	24, −48, 66	3.49
Placebo vs ketamine	Insula	48, 12, 8	3.77
		36, 20, 8	2.82
	Cerebellum	38, −66, −46	3.87
		−10, −82, −24	3.15
	Middle frontal gyrus	30, 18, 38	3.6
		−42, 16, 42	3.27
	Dorsomedial prefrontal cortex	6, 26, 60	3.07
		12, 26, 52	3.03
		−4, 26, 58	2.96
	Dorsolateral prefrontal cortex	40, 56, 20	2.81
		46, 30, 38	2.62
		−42, 16, 42	3.27
		−36, 12, 52	2.79
	(Inferior) parietal cortex	−54, −44, 42	3.26
	frontopolar cortex	−36, 54, 12	3.1
		42, 58, −4	2.67

prefrontal cortex. Other correlations were observed in the anterior insula, in addition to subcortical regions encompassing bilateral caudate nucleus, thalamus and cerebellum (Figure 5, Table 1). Put simply, elevated temperature was associated with enhanced activity in these regions. Conversely, βm was negatively correlated with activity in a bilateral network including cuneus, precuneus, posterior cingulate and medial temporal lobe.

In the ketamine session, the positive correlation with βm was significantly reduced compared to placebo in a bilateral fronto-parietal network, including the dmPFC, bilateral frontopolar cortex, bilateral lateral prefrontal cortex and left parietal cortex, as well as the anterior insula (Figure 5, Table 1). Thus, trial-to-trial variations in temperature expressed in the fronto-parietal network were diminished under ketamine. There was no significant difference between sessions for the negative correlation with βm.

Table 2. Brain regions reflecting confidence-modulated learning rate (αm)

	Structure	*MNI coordinates (x, y, z)*	*Z-score*
Positive	Posterior insula	58, −8, −2	4.52
		40, −18, −2	3.69
		−36, −18, −2	3.68
	Ventromedial prefrontal cortex	6, 36, −10	4.09
Negative	Anterior insula	38, 26, −6	3.7

At outcome onset. Under placebo, we observed a positive correlation with αm in the vmPFC and bilateral posterior insula extending to the superior temporal cortex (Figure 5, Table 2). These regions therefore increased their responses to confirmatory outcomes, and decreased their responses to contradictory outcomes, as confidence accumulated within learning blocks. Conversely, there was a negative correlation in the right anterior insula.

There was no significant difference in the correlation with αm between placebo and ketamine sessions at the whole-brain level, nor in a ROI analysis focusing on the vmPFC ($P > 0.1$). Correlation with αm corresponds to an interaction between confidence and outcome category (confirmatory or contradictory). We verified that the correlation was not reducible to the main effect of outcome category: when this was regressed out, the correlation with αm was still significant in our vmPFC ROI under placebo ($t(20) = 2.79$; $P = 0.01$) but not under ketamine ($t(20) = 1.41$; $P = 0.18$), though the direct comparison was not significant ($P > 0.1$). In short, under placebo but not ketamine, the difference between confirmatory and contradictory outcomes was amplified following the trial-wise increase in confidence within learning blocks.

DISCUSSION

Our working hypothesis was that early psychosis is characterized by a state in which the ability to acquire a robust and confident model of the world is lost. We tested this hypothesis at both the computational and neural levels, by combining a pharmacological model of early psychosis through NMDA blockade with model-based analysis of behavioral choices and fMRI data. The effects of NMDA blockade manifested in two ways:[1] a decreased ability to optimize contingency learning in conditions of high confidence,[2] a concurrent alteration in the regulation of brain systems reflecting choice stochasticity, notably in a bilateral fronto-parietal network including the dmPFC. Through use of a low dose of ketamine (rather than a higher one which would cause global cognitive difficulties), we have been able to identify a subtle and interpretable effect. Our findings have implications both for our understanding of contingency learning mechanisms and for theoretical perspectives on the emergence of psychosis. Because our experiment was carried out in a limited number of participants, as is common to pharmaco-MRI studies for obvious ethical reasons, we consider the implications below as primarily theoretical suggestions that will guide further investigations.

Contingency learning mechanisms in an unstable environment

Our benchmark computational model was a standard Q-learning algorithm, which has been shown to provide a good account of instrumental learning in a variety of situations.[50] However, the task was expressly designed such that Q-learning would not be optimal. This is because Q-learning gives a constant weight to

outcomes in value updating, and a constant weight to value estimates in decision making. Yet it is adaptive to adjust these weights in unstable environments, where contingencies are stochastic and susceptible to sudden reversals, depending on the confidence in value estimates. Behavioral data suggested that participants did modulate choice and learning parameters as a function of confidence. To analyze this we developed a hierarchical model with a meta-cognitive level that monitors confidence and modulates first-level Q-learning parameters, an approach that has been formalized in the meta-learning framework.[45,46,51] At the meta-cognitive level, Bayesian model selection indicated that an independent delta rule on outcome optimality (similar to that used in Khamassi *et al.*[47]) provided a better fit than a direct accumulation of unsigned prediction errors (as implemented in [48]). Our construct of confidence can therefore be considered as surface monitoring, since it remains blind to the computations driving choices. In the model that best captured the behavioral data, both choice temperature and learning rate were dynamically adjusted as a function of confidence. Moreover, confidence had a differential impact on confirmatory outcomes (whose weight was amplified) and contradictory outcomes (whose weight was reduced). Together, these confidence-based adjustments enabled stabilizing internal representations of environmental contingencies (cue value estimates) and optimizing behavioral policy (exploitation of cue values).

Our concept of confidence can be linked to several recent theoretical propositions, in which higher level representations control lower-level processes. For example, it has been suggested that uncertainty, which quantifies ignorance about true values, drives the trade-off between exploitation and exploration.[52] In the predictive coding framework, the precision of (or confidence in) beliefs determines the weight that prediction errors have in belief updating. Indeed, aberrant encoding of precision has been recently proposed to account for various aspects of psychosis.[10] Some implementations of hierarchical Bayesian modeling can also be seen as very close to our approach, particularly when both the learning and decision rules are modulated by precision estimates.[42] Note however that a new and important feature of our model is the differential impact of confidence on learning depending on the nature of the outcome (confirmatory or not), which allows neglect of contradictory information. We acknowledge that the concept of confidence is used for convenience, and corresponds in fact to a running estimate of performance. Whether this measure matches what participants would report as a feeling of confidence remains to be demonstrated.

Neuroimaging data provided additional support for our hierarchical model. At the time of cues, trial-wise variation in choice temperature was reflected in activation of a fronto-parietal network that has been previously implicated in cognitive control.[53–57] This does not imply that all these regions have the function of representing choice temperature. Their activity might represent an indirect correlate of variations in this computational variable. In particular, regions such as the dmPFC has been involved in monitoring errors,[58,59] detecting conflicts[60,61] and making decisions under uncertainty.[59,62,63] This region might signal the necessity of additional control, or even implement this necessary control, in periods of doubt regarding which choice is the best.[64,65] At the time of outcomes, trial-wise variation in learning rate was positively reflected in regions such as the vmPFC, which has been implicated in encoding the subjective value of stimuli.[66,67] Here, this region increased its response to confirmatory outcomes, and decreased its response to contradictory outcomes, from the beginning to the end of learning blocks. This finding extends a previous report that the vmPFC integrates option value and choice confidence[68] by showing that this integration also applies to outcomes. Interestingly, the reverse pattern of activity was observed in the anterior insula, a region involved in signaling aversive values.[69,70] Thus, these two regions appeared to mediate the influence of meta-cognitive control on proximal reactions to gains and losses, such that they align to the distal goal of optimizing performance.

Emergence of psychosis through NMDA blockade

Model-based analysis of the behavior suggested that NMDA blockade was associated with a reduced capacity to stabilize an internal model in order to capitalize on environmental regularities. This was evidenced by a reduced weight of confidence on choice temperature and learning rate. The performance deficit induced by ketamine infusion was therefore observed at the end of learning blocks, when confidence should be high enough to stabilize cue value estimation and exploitation policy. Our findings thus show that ketamine was associated with diminished ability to stabilize cue value estimates in the presence of probabilistic errors, as if a persistent doubt undermined optimization of behavior and made them more vulnerable to the effects of 'noise' trials.

In a very simple environment as in our task (two cues with opposite values), such an impairment has limited impact and could hardly induce strange beliefs. In a more complex environment, where multiple internal explanatory models can be held at the same time, we would expect this impairment to forge strange beliefs, by combination of existing models or through the emergence of unexpected explanations. Our results therefore extend previous accounts of early psychosis, in which altered prediction errors lead to a sense of strangeness and to abnormalities in belief updating.[9,15,26] Our findings suggest that it is important to take into account not just how prediction errors are used in low-level associative learning, but in how outcome optimality is integrated to modulate low-level parameters, via confidence monitoring.

We note that changes in key behavioral parameters did not correlate with the subtle psychopathology induced by this low dose of ketamine. This is perhaps unsurprising given the lack of statistical power—our experiment was devised with a view to identifying differences between ketamine and placebo rather than across-subject correlations. We see two other reasons that could account for this limitation. First, the neuro-cognitive perturbations that we demonstrated here might have different kinetics from those of psychotic symptoms (the former preceding the latter). Therefore, these two dimensions might remain uncorrelated at a given time. Second, if we assume that psychotic-like symptoms yield from more elementary cognitive dysfunctions, this link could be modulated (and hence blurred) by several factors, such as the existence of baseline (pre-ketamine) bizarre ideas, or the ability to introspect and conscious access to these dysfunctions and therefore to report psychotic-like symptoms.

In line with the behavioral analysis, the fMRI data showed that the confidence-based modulation of Q-learning parameters was significantly altered during ketamine infusion. Specifically, brain activity reflecting choice temperature was significantly less modulated by confidence under ketamine than placebo. This difference was observed in a bilateral fronto-parietal network, including the dmPFC. A detrimental effect of ketamine on dmPFC activation is in line with repeated observations of dorsal cingulate cortex impairment in patients with schizophrenia.[71] Critically, here we offer a computational account of this effect, suggesting that that dmPFC impairment might play a key role in early symptoms of psychosis by compromising belief updating and policy adjustment in unstable environments. This dmPFC dysfunction could either alter confidence level or perturb the impact of confidence on behavioral policy.

The effect of ketamine on fronto-parietal regions might also relate to the well-established changes in consciousness produced by higher doses of ketamine,[72] since the global workspace theory.[73,74] implicates these regions in conscious access by Interestingly, modulation of choice temperature by confidence was initially

proposed to regulate the activity of workspace neurons whose role is to determine the degree of effort invested in decision making[46,47] in keeping with the concept of vigilance.[73] One may speculate that the meta-cognitive component of our model, notably confidence monitoring and down-regulation of choice temperature, requires conscious processing. Thus, dysfunction of this part could be linked to both alteration of consciousness with higher doses of ketamine and to dysfunction of conscious processing in schizophrenic patients,[75,76] who would perform contingency learning in a more implicit way. Evidence for such a speculation would require further experiments manipulating consciousness levels.

The earliest stages of psychotic illness present an intriguing and puzzling set of cognitive changes. Computational psychiatry[1,3] offers new and rich frameworks for considering these changes and linking them to underlying neural alterations. Here we have shown that pharmacological fMRI, employing a well-established drug model of psychosis, presents a powerful tool in developing such frameworks, offering an opportunity to determine how controlled perturbations in glutamate function relate to altered balance in the dynamic control of optimal learning and behavior.

CONFLICT OF INTEREST

RG has received compensation as a member of the scientific advisory board of Janssen, Lundbeck, Roche, Takeda. He has served as consultant and/or speaker for Astra Zeneca, Pierre Fabre, Lilly, Otsuka, SANOFI, Servier and received compensation, and he has received research support from Servier. PCF has consulted for Glaxo SmithKline and Lundbeck and received compensation. MOK has received compensation as a member of the scientific advisory board of Roche. She has served as speaker for Janssen and received unrestricted support for conference organization from Janssen and Otsuka-Lunbeck, and she has been invited to scientific meetings by Lundbeck and Takeda. AS has consulted for Servier and received compensation. FV has served as speaker for Servier and received compensation.

ACKNOWLEDGMENTS

The authors are grateful to Mael Lebreton and Jean Daunizeau for helpful conversations and advices. FV was supported by the Groupe Pasteur Mutualité. RG was supported by the Fondation pour la Recherche Médicale and the Fondation Bettencourt Schueller. SP is supported by a Marie Curie Intra-European fellowship (FP7-PEOPLE-2012-IEF). AF was supported by National Health and Medical Research Council grants (IDs: 1050504 and 1066779) and an Australian Research Council Future Fellowship (ID: FT130100589). This work was supported by the Wellcome Trust and the Bernard Wolfe Health Neuroscience Fund.

REFERENCES

1 Montague PR, Dolan RJ, Friston KJ, Dayan P. Computational psychiatry. *Trends Cogn Sci* 2012; **16**: 72–80.
2 Maia TV, Frank MJ. From reinforcement learning models to psychiatric and neurological disorders. *Nat Neurosci* 2011; **14**: 154–162.
3 Friston KJ, Stephan KE, Montague R, Dolan RJ. Computational psychiatry: the brain as a phantastic organ. *Lancet Psychiatry* 2014; **1**: 148–158.
4 Corlett PR, Murray GK, Honey GD, Aitken MRF, Shanks DR, Robbins TW *et al.* Disrupted prediction-error signal in psychosis: evidence for an associative account of delusions. *Brain* 2007; **130**: 2387–2400.
5 Gradin VB, Kumar P, Waiter G, Ahearn T, Stickle C, Milders M *et al.* Expected value and prediction error abnormalities in depression and schizophrenia. *Brain* 2011; **134**: 1751–1764.
6 Morris RW, Vercammen A, Lenroot R, Moore L, Langton JM, Short B *et al.* Disambiguating ventral striatum fMRI-related BOLD signal during reward prediction in schizophrenia. *Mol Psychiatry* 2012; **17**: 235, 80-9.
7 Murray GK, Corlett PR, Clark L, Pessiglione M, Blackwell AD, Honey G *et al.* Substantia nigra/ventral tegmental reward prediction error disruption in psychosis. *Mol Psychiatry* 2008; **13**: 239, 67-76.
8 Waltz JA, Schweitzer JB, Ross TJ, Kurup PK, Salmeron BJ, Rose EJ *et al.* Abnormal responses to monetary outcomes in cortex, but not in the basal ganglia, in schizophrenia. *Neuropsychopharmacology* 2010; **35**: 2427–2439.
9 Fletcher PC, Frith CD. Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nat Rev Neurosci* 2009; **10**: 48–58.
10 Adams RA, Stephan KE, Brown HR, Frith CD, Friston KJ. The computational anatomy of psychosis. *Front Psychiatry* 2013; **4**: 47.
11 Barch DM, Dowd EC. Goal representations and motivational drive in schizophrenia: the role of prefrontal-striatal interactions. *Schizophr Bull* 2010; **36**: 919–934.
12 Gold JM, Waltz JA, Prentice KJ, Morris SE, Heerey EA. Reward processing in schizophrenia: a deficit in the representation of value. *Schizophr Bull* 2008; **34**: 835–847.
13 Deserno L, Boehme R, Heinz A, Schlagenhauf F. Reinforcement learning and dopamine in schizophrenia: dimensions of symptoms or specific features of a disease group? *Front Psychiatry* 2013; **4**: 172.
14 Whitson JA, Galinsky AD. Lacking control increases illusory pattern perception. *Science* 2008; **322**: 115–117.
15 Kapur S. Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. *Am J Psychiatry* 2003; **160**: 13–23.
16 Corlett P, Fletcher P. The neurobiology of schizotypy: fronto-striatal prediction error signal correlates with delusion-like beliefs in healthy people. *Neuropsychologia* 2012; **50**: 3612–3620.
17 Micoulaud-Franchi JA, Aramaki M, Merer A, Cermolacce M, Ystad S, Kronland-Martinet R *et al.* Toward an exploration of feeling of strangeness in schizophrenia: perspectives on acousmatic and everyday listening. *J Abnorm Psychol* 2012; **121**: 628–640.
18 O'Connor K. Cognitive and meta-cognitive dimensions of psychoses. *Can J Psychiatry* 2009; **54**: 152–159.
19 Coltheart M, Langdon R, McKay R. Delusional belief. *Annu Rev Psychol* 2011; **62**: 271–298.
20 Fyfe S, Williams C, Mason OJ, Pickup GJ. Apophenia, theory of mind and schizotypy: perceiving meaning and intentionality in randomness. *Cortex* 2008; **44**: 1316–1325.
21 Broome MR, Johns LC, Valli I, Woolley JB, Tabraham P, Brett C *et al.* Delusion formation and reasoning biases in those at clinical high risk for psychosis. *Br J Psychiatry Suppl* 2007; **51**: s38–s42.
22 Colbert SM, Peters ER. Need for closure and jumping-to-conclusions in delusion-prone individuals. *J Nerv Ment Dis* 2002; **190**: 27–31.
23 Krystal JH, Karper LP, Seibyl JP, Freeman GK, Delaney R, Bremner JD *et al.* Subanesthetic effects of the noncompetitive NMDA antagonist, ketamine, in humans. Psychotomimetic, perceptual, cognitive, and neuroendocrine responses. *Arch Gen Psychiatry* 1994; **51**: 199–214.
24 Javitt DC, Zukin SR, Heresco-Levy U, Umbricht D. Has an angel shown the way? Etiological and therapeutic implications of the PCP/NMDA model of schizophrenia. *Schizophr Bull* 2012; **38**: 958–966.
25 Pomarol-Clotet E, Honey GD, Murray GK, Corlett PR, Absalom AR, Lee M *et al.* Psychological effects of ketamine in healthy volunteers. Phenomenological study. *Br J Psychiatry* 2006; **189**: 173–179.
26 Corlett PR, Honey GD, Krystal JH, Fletcher PC. Glutamatergic model psychoses: prediction error, learning, and inference. *Neuropsychopharmacology* 2011; **36**: 294–315.
27 Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* 2006; **442**: 1042–1045.
28 Pessiglione M, Petrovic P, Daunizeau J, Palminteri S, Dolan RJ, Frith CD. Subliminal instrumental conditioning demonstrated in the human brain. *Neuron* 2008; **59**: 561–567.
29 Behrens TE, Woolrich MW, Walton ME, Rushworth MF. Learning the value of information in an uncertain world. *Nat Neurosci* 2007; **10**: 1214–1221.
30 Absalom AR, Lee M, Menon DK, Sharar SR, De Smet T, Halliday J *et al.* Predictive performance of the Domino, Hijazi, and Clements models during low-dose target-controlled ketamine infusions in healthy volunteers. *Br J Anaesth.* 2007; **98**: 615–623.
31 Chouinard G, Miller R. A rating scale for psychotic symptoms (RSPS): Part I: theoretical principles and subscale 1: perception symptoms (illusions and hallucinations). *Schizophr Res* 1999; **38**: 101–122.
32 Overall JE, Gorham D. The Brief Psychiatric Rating Scale (BPRS): recent developments in ascertainment and scaling. *Psychopharmacol Bull* 1988; **24**: 97–99.
33 Bremner JD, Krystal JH, Putnam FW, Southwick SM, Marmar C, Charney DS *et al.* Measurement of dissociative states with the clinician-administered dissociative states scale (CADSS). *J Trauma Stress* 1998; **11**: 125–136.
34 Dandash O, Harrison BJ, Adapa R, Gaillard R, Giorlando F, Wood SJ *et al.* Selective augmentation of striatal functional connectivity following NMDA receptor antagonism: implications for psychosis. *Neuropsychopharmacology* 2015; **40**: 622–631.
35 Friston K, Mattout J, Trujillo-Barreto N, Ashburner J, Penny W. Variational free energy and the Laplace approximation. *Neuroimage* 2007; **34**: 220–234.
36 Daunizeau J, Adam V, Rigoux L. VBA: a probabilistic treatment of nonlinear models for neurobiological and behavioural data. *PLoS Comput Biol* 2014; **10**: e1003441.

37 Rigoux L, Stephan KE, Friston KJ, Daunizeau J. Bayesian model selection for group studies – revisited. *Neuroimage* 2014; **84**: 971–985.
38 Sutton RS, Barto AG. *Reinforcement Learning, a Bradford book*. MIT Press: Cambridge, MACambridge, MA, 1998.
39 Rushworth MF, Behrens TE. Choice, Uncertainty and value in prefrontal and cingulate cortex. *Nat Neurosci* 2008; **11**: 389–397.
40 Mathys C, Daunizeau J, Friston KJ, Stephan KE. A bayesian foundation for individual learning under uncertainty. *Front Human Neurosci* 2011; **5**: 39.
41 Iglesias S, Mathys C, Brodersen KH, Kasper L, Piccirelli M, den Ouden HE *et al.* Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. *Neuron* 2013; **80**: 519–530.
42 Diaconescu AO, Mathys C, Weber LA, Daunizeau J, Kasper L, Lomakina EI *et al.* Inferring on the intentions of others by hierarchical Bayesian learning. *PLoS Comput Biol* 2014; **10**: e1003810.
43 Collins A, Koechlin E. Reasoning, learning, and creativity: frontal lobe function and human decision-making. *PLoS Biol* 2012; **10**: e1001293.
44 Payzan-LeNestour E, Bossaerts P. Risk, unexpected uncertainty, and estimation uncertainty: Bayesian learning in unstable settings. *PLoS Comput Biol* 2011; **7**: e1001048.
45 Doya K. Metalearning and neuromodulation. *Neural Netw.* 2002; **15**: 495–506.
46 Khamassi M, Enel P, Dominey PF, Procyk E. Medial prefrontal cortex and the adaptive regulation of reinforcement learning parameters. *Prog Brain Res.* 2013; **202**: 441–464.
47 Khamassi M, Lallee S, Enel P, Procyk E, Dominey PF. Robot cognitive control with a neurophysiologically inspired reinforcement learning model. *Front Neurorobot* 2011; **5**: 1.
48 Krugel LK, Biele G, Mohr PN, Li SC, Heekeren HR. Genetic variation in dopaminergic neuromodulation influences the ability to rapidly and flexibly adapt decisions. *Proc Natl Acad Sci USA* 2009; **106**: 17951–17956.
49 Lee SW, Shimojo S, O'Doherty JP. Neural computations underlying arbitration between model-based and model-free learning. *Neuron* 2014; **81**: 687–699.
50 Rangel A, Camerer C. Montague PR. A framework for studying the neurobiology of value-based decision making. *Nat Rev Neurosci.* 2008; **9**: 545–556.
51 Doya K. Modulators of decision making. *Nat Neurosci* 2008; **11**: 410–416.
52 Daw ND, Niv Y, Dayan P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat Neurosci* 2005; **8**: 1704–1711.
53 Zanto TP, Gazzaley A. Fronto-parietal network: flexible hub of cognitive control. *Trends Cogn Sci* 2013; **17**: 602–603.
54 Cole MW, Reynolds JR, Power JD, Repovs G, Anticevic A, Braver TS. Multi-task connectivity reveals flexible hubs for adaptive task control. *Nat Neurosci* 2013; **16**: 1348–1355.
55 Nee DE, Wager TD, Jonides J. Interference resolution: insights from a meta-analysis of neuroimaging tasks. *Cogn Affect Behav Neurosci.* 2007; **7**: 1–17.
56 Glascher J, Adolphs R, Damasio H, Bechara A, Rudrauf D, Calamia M *et al.* Lesion mapping of cognitive control and value-based decision making in the prefrontal cortex. *Proc Natl Acad Sci USA* 2012; **109**: 14681–14686.
57 Niendam TA, Laird AR, Ray KL, Dean YM, Glahn DC, Carter CS. Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cogn Affect Behav Neurosci* 2012; **12**: 241–268.
58 Carter CS, Braver TS, Barch D, Botvinick MM, Noll D, Cohen JD. Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science* 1998; **280**: 747–749.
59 Brown JW, Braver TS. Learned predictions of error likelihood in the anterior cingulate cortex. *Science* 2005; **307**: 1118–1121.
60 Botvinick M, Nystrom LE, Fissell K, Carter CS, Cohen JD. Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature* 1999; **402**: 179–181.
61 Kerns JG, Cohen JD, MacDonald AW 3rd, Cho RY, Stenger VA, Carter CS. Anterior cingulate conflict monitoring and adjustments in control. *Science* 2004; **303**: 1023–1026.
62 Rushworth MF, Walton ME, Kennerley SW, Bannerman DM. Action sets and decisions in the medial frontal cortex. *Trends Cogn Sci* 2004; **8**: 410–417.
63 Venkatraman V, Huettel SA. Strategic control in decision-making under uncertainty. *Eur J Neurosci.* 2012; **35**: 1075–1082.
64 Shenhav A, Botvinick MM, Cohen JD. The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron* 2013; **79**: 217–240.
65 Cohen JD, McClure SM, Yu AJ. Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Phil Trans R Soc Lond B Biol Sci* 2007; **362**: 933–942.
66 Lebreton M, Jorge S, Michel V, Thirion B, Pessiglione M. An automatic valuation system in the human brain: evidence from functional neuroimaging. *Neuron* 2009; **64**: 431–439.
67 Levy DJ, Glimcher PW. The root of all value: a neural common currency for choice. *Curr Opin Neurobiol* 2012; **22**: 1027–1038.
68 De Martino B, Fleming SM, Garrett N, Dolan RJ. Confidence in value-based choice. *Nat Neurosci* 2013; **16**: 105–110.
69 Palminteri S, Justo D, Jauffret C, Pavlicek B, Dauta A, Delmaire C *et al.* Critical roles for anterior insula and dorsal striatum in punishment-based avoidance learning. *Neuron* 2012; **76**: 998–1009.
70 Büchel C, Morris J, Dolan RJ, Friston KJ. Brain systems mediating aversive conditionning: an event-related fonctional magnetic resonance imaging. *Neuron* 1998; **20**: 947–957.
71 Fornito A, Yucel M, Dean B, Wood SJ, Pantelis C. Anatomical abnormalities of the anterior cingulate cortex in schizophrenia: bridging the gap between neuroimaging and neuropathology. *Schizophr Bull* 2009; **35**: 973–993.
72 Marland S, Ellerton J, Andolfatto G, Strapazzon G, Thomassen O, Brandner B *et al.* Ketamine: use in anesthesia. *CNS Neurosci Therapeut* 2013; **19**: 381–389.
73 Dehaene S, Kerszberg M, Changeux JP. A neuronal model of a global workspace in effortful cognitive tasks. *Proc Natl Acad Sci USA* 1998; **95**: 14529–14534.
74 Dehaene S, Naccache L. Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition* 2001; **79**: 1–37.
75 Dehaene S, Artiges E, Naccache L, Martelli C, Viard A, Schurhoff F *et al.* Conscious and subliminal conflicts in normal subjects and patients with schizophrenia: the role of the anterior cingulate. *Proc Natl Acad Sci USA* 2003; **100**: 13722–13727.
76 Del Cul A, Dehaene S, Leboyer M. Preserved subliminal processing and impaired conscious access in schizophrenia. *Arch Gen Psychiatry* 2006; **63**: 1313–1323.

Supplementary Information accompanies the paper on the Molecular Psychiatry website (http://www.nature.com/mp)

Supplemental Experimental Procedures

Description of model space

- Reinforcement learning level

We started with a basic, model-free reinforcement learning algorithm.(1) For each cue (say A or B), the model estimated the expected value, based on the individual outcome history, and made a choice between risky and less risky options. The expected values were set at zero before learning, and after each trial t the value of the ongoing cue (say A) was updated in proportion to prediction error, according to the 'delta' rule (2, 3):

$Q_A\,(t+1) = Q_A\,(t) + \alpha \times \delta\,(t)$

where δ(t) is the prediction error, defined as the difference between actual and expected outcome: $\delta(t)=R_Q(t)- Q_A(t)$.

Then the probability, or likelihood, of a "Risky" choice was estimated from the expected value according to the softmax rule:

$P_{A\ Risky}\,(t) = 1/\,(1 + \exp^{(-Q_A(t)/\beta)})$

The learning rate α and the choice temperature β are free parameters, with the constraints $0\leq\alpha\leq1$ and $\beta>0$. The learning rate adjusts the weight assigned to prediction error in value updating, and the choice temperature the degree of exploration (as opposed to exploitation of the learned value).

We devised three variants of this reinforcement learning level, following step-by-step increments from model-free to model-based strategy, i.e. adding pieces of information about task structure.

In a first variant, the reinforcer R_Q was the monetary value of the outcome (1, 0.1, -0.1 or -1). This variant can be considered as a model-free strategy, in line with the law of effect, meaning that outcomes increased the probability of repeating the same choice, depending on their sign and magnitude.

In a second variant, the reinforcer R_Q was defined according to outcome valence (Val) and not its magnitude (i.e. 1 when winning £1 or 10p; -1 when loosing £1 or 10p). This variant implies that subjects understood that cues determined the outcome valence (positive or negative) and not its magnitude, which depended on the choice.

In a third variant, the reinforcer R_Q was defined according to outcome valence (Val) and the update of the current cue (say cue A) was transferred to the alternative cue (cue B).

$$Q_B(t+1) = -Q_A(t+1)$$

This variant implies that subjects understood that there were only two cues, with opposite valence. In other words, the two cue values summed up to zero.

- *Meta-learning level*

Reinforcement learning models have constant parameters (learning rate and choice stochasticity). This limits the capacity to optimize the behavioral policy around the end of learning blocks, once subjects believe themselves to have a reasonably good estimation of contingencies. At this point, prediction errors should be tempered, and choices tuned to a more deterministic exploitation of learned contingencies.(4-6) Conversely, when contingencies suddenly change after reversals, prediction errors should be given more weight, and choices should be more exploratory. One way to optimize the behavior is to subordinate the reinforcement learning parameters to a higher level of control that monitors performance. A second series of models therefore included a meta-cognitive level consisting in updating confidence so as to down-regulate contingency learning and choice stochasticity. We compared two ways to monitor confidence and four ways to use it.

- *Confidence monitoring level*

In both variants, confidence was monitored using a delta rule. The confidence learning rate γ was a free parameter, with $0 \leq \gamma \leq 1$. The initial value of confidence, $C(0)$ was also fitted as a free parameter, with $0 \leq C(0) \leq 1$.

In a first variant, we used the absolute value of the prediction error computed at the reinforcement learning level to update confidence (7):

$$C(t+1) = C(t) + \gamma \times ((2-|\delta(t)|)/2 - C(t))$$

In a second variant, we used outcome optimality (Op) to update confidence (i.e. 1 for winning £1 or losing 10p, -1 otherwise):

$$C(t+1) = C(t) + \gamma \times (Op(t) - C(t))$$

- Modulation of low-level free parameters

Confidence was used to modulate the free parameters in the reinforcement learning models. This was done after each outcome, which brought information about how accurate the reinforcement learning model was, in terms of value estimates or behavioral policy. We considered four possibilities: modulation of learning rate or choice temperature, or both with the same weight, or both with a different weight.

The learning rate was modulated on the basis of not only confidence but also the outcome category. The idea is that to stabilize a representation of learned contingencies, subjects should increase their sensitivity to confirmation and decrease their sensitivity to contradiction. The impact of confidence on the learning rate α therefore depended on whether the outcome was confirmatory (outcome and cue value have the same sign; Val(t) = sign(Q(t))) or not.

For confirmatory outcomes, α was modulated as follows:

$\alpha m(t) = (\alpha 0 + k_\alpha * C(t))/(1+k_\alpha * C(t))$ where $\alpha 0$ and k_α are free parameter

And for contradictory outcomes:

$\alpha m(t) = \alpha 0/(1+ k_\alpha * C(t))$

Therefore, when confidence increased the modified learning rate αm got closer to 1 for confirmatory outcomes and closer to zero for contradictory outcomes.

The choice temperature β was modulated such that exploration was reduced when confidence increased:

$\beta m(t) = \beta 0/(1+k_\beta * C(t))$ where $\beta 0$ and k_β are free parameter

This modulation enables increasing exploitation above matching behavior, i.e. choosing the risky option more than 80% of the time following a cue that associated to a reward 80% of the time.

To test whether these modulations improved the fit of observed choices, we compared between models that included or not the free parameters (k_α and k_β), which could have or not identical values.

Other hierarchical models

Other hierarchical models have been developed to implement a form of second-level confidence that modulates first-level estimates. For instance, one hierarchical Bayesian architecture models the behavior in a probabilistic reversal learning task, with a second-level inference that tracks the occurrence of contingency reversals.(8) However, reversals were numerous in this task and participants were extensively trained, so they had built an internal model of the task (including the possibility of a reversal) before entering the scanner. In our paradigm, participants were not informed about the presence of reversals and they encountered only three of them, which was not enough to build and integrate an explicit notion of reversal at the meta-cognitive level. Unsurprisingly, we found no evidence that the behavior was more easily reversed the last time compared to the first one: if anything, performance in the last block was worst. For similar reasons, we did not include the possibility of re-using contingencies that were learned in previous blocks, as was done in another hierarchical model with task-set monitoring on top of Q-learning.(9) Indeed, there was no evidence that the second and third reversals, after which subjects could have returned to previous contingency sets, were learned faster than the first one. In addition, none of the existing models implemented the differential impact of the meta-cognitive level on the first-level learning rule, which enables participants to specifically ignore contradictory outcomes (probabilistic errors), a key way to stabilize behavior at the end of blocks. It is important to keep in mind that our paradigm was not designed to investigate reversal processes per se, but to examine how behavior is optimized between reversals.

Supplemental data

Family analyses

Family model comparison (10) was used to test whether each level of complexity added to the basic reinforcement learning model was necessary for explaining choice data. In a first comparison, the model space was divided into three families, depending on RL-variant (i.e. whether the monetary value or the outcome valence was integrated in the delta-rule and

whether the two cues or only the current cue was updated). Results confirmed that participants integrated the two aspects of task structure (xp =0.97 and 0.93 for placebo and ketamine sessions, respectively): first that only the outcome valence, and not monetary amount, was informative about cue value, and second that the two cues always had opposite valence such that they could both be updated after every outcome. In a second comparison, we divided the model space into three families, according to the way confidence was updated (see Fig 3: no confidence monitoring (and therefore no modulation), prediction error-based and optimality-based confidence updating. Evidence was higher for optimality-based confidence updating (xp=1 and 0.74 for placebo and ketamine sessions, respectively). In a third comparison, we divided the model space into five families, according to the way confidence was used (see Fig 3): no modulation (and therefore no confidence monitoring), modulation of learning rate, choice temperature, both with the same weight, or both with different weights. Results indicated that confidence was used to modulate both learning rate and choice temperature, with the same weight (xp=1 and 0.76 for placebo and ketamine sessions, respectively).

Correlation with psychotic-like symptoms

As our ultimate goal is to model delusion formation, we looked for correlations between psychotic-like symptoms induced by ketamine and our behavioural and imaging findings. For each of the 3 scales (RSPS, CADS, BPRS), the difference between placebo and ketamine sessions was computed for each subject and correlated to our behavioural markers of ketamine effect (decreased performance in the last bin and decreased weight of confidence on learning rate and choice temperature). No significant correlation was found. As far as imaging results are concerned, we also searched to correlate the increase of psychological symptoms to the reduction of the positive correlation with βm induced by ketamine. Scores were entered as a covariant in second-level GLM. No significant cluster was found.

Supplementary tables

Table S1

Exceedance probabilities table for Placebo sessions.

Table S2

Exceedance probabilities table for Ketamine sessions.

Table S3

Parameter estimates for the best computational model

Table S1

Placebo				
		RL - Variant		
Confidence – Monitoring Variant	Modulation of free parameters - Variant	1	2	3
0	0	0.00	0.00	0.00
1	1	0.00	0.00	0.00
1	2	0.00	0.00	0.00
1	3	0.00	0.00	0.00
1	4	0.00	0.00	0.00
2	1	0.00	0.00	0.00
2	2	0.00	0.00	0.00
2	3	0.00	0.00	0.00
2	4	0.00	0.02	0.96

Table S2

Ketamine				
		RL - Variant		
Confidence – Monitoring Variant	Modulation of free parameters – Variant	1	2	3
0	0	0.01	0.03	0.03
1	1	0.01	0.01	0.10
1	2	0.01	0.02	0.04
1	3	0.01	0.01	0.01
1	4	0.01	0.01	0.10
2	1	0.01	0.01	0.01
2	2	0.00	0.03	0.02
2	3	0.01	0.01	0.01
2	4	0.01	0.05	0.45

Table S3

	$\alpha 0$	$\beta 0$	C0	γ	κ
Placebo	0.34 (0.12)	0.89 (0.74)	0.51 (0.04)	0.45 (0.15)	0.95 (0.39)
Ketamine	0.29 (0.16)	0.97 (1.17)	0.50 (0.04)	0.50 (0.16)	0.62 (0.53)

Supplementary references

1. Watkins CJ, Dayan P. Q-learning. Machine learning. 1992;8(3-4):279-92.
2. Rescorla RA, Wagner AR. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. Classical conditioning II: Current research and theory. 1972;2:64-99.
3. Sutton RS, Barto AG. Reinforcement learning, a Bradford book. MIT Press, Cambridge, MA; 1998.
4. Rushworth MF, Behrens TE. Choice, uncertainty and value in prefrontal and cingulate cortex. Nature neuroscience. 2008;11(4):389-97.
5. Behrens TE, Woolrich MW, Walton ME, Rushworth MF. Learning the value of information in an uncertain world. Nature neuroscience. 2007;10(9):1214-21.
6. Mathys C, Daunizeau J, Friston KJ, Stephan KE. A bayesian foundation for individual learning under uncertainty. Frontiers in human neuroscience. 2011;5:39.
7. Krugel LK, Biele G, Mohr PN, Li SC, Heekeren HR. Genetic variation in dopaminergic neuromodulation influences the ability to rapidly and flexibly adapt decisions. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(42):17951-6.
8. Hampton AN, Bossaerts P, O'Doherty JP. The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2006;26(32):8360-7.
9. Collins A, Koechlin E. Reasoning, Learning, and Creativity: Frontal Lobe Function and Human Decision-Making. PLoS biology. 2012;10:e1001293.
10. Rigoux L, Stephan KE, Friston KJ, Daunizeau J. Bayesian model selection for group studies - revisited. NeuroImage. 2014;84:971-85.