BRAIN
A JOURNAL OF NEUROLOGY

Estimating changing contexts in schizophrenia

Claire M. Kaplan,[1,2] Debjani Saha,[1] Juan L. Molina,[2] William D. Hockeimer,[1] Elizabeth M. Postell,[1] Jose A. Apud,[2] Daniel R. Weinberger[1,2,3,4] and Hao Yang Tan[1,2,3]

See Stephan *et al.* (doi:10.1093/aww120) for a scientific commentary on this work.

Real world information is often abstract, dynamic and imprecise. Deciding if changes represent random fluctuations, or alterations in underlying contexts involve challenging probability estimations. Dysfunction may contribute to erroneous beliefs, such as delusions. Here we examined brain function during inferences about context change from noisy information. We examined cortical-subcortical circuitry engaging anterior and dorsolateral prefrontal cortex, and midbrain. We hypothesized that schizophrenia-related deficits in prefrontal function might overestimate context change probabilities, and that this more chaotic worldview may subsequently gain familiarity and be over-reinforced, with implications for delusions. We then examined these opposing information processing biases against less expected versus familiar information patterns in relation to genetic risk for schizophrenia in unaffected siblings. In one experiment, 17 patients with schizophrenia and 24 normal control subjects were presented in 3 T magnetic resonance imaging with numerical information varying noisily about a context integer, which occasionally shifted up or down. Subjects were to indicate when the inferred numerical context had changed. We fitted Bayesian models to estimate probabilities associated with change inferences. Dynamic causal models examined cortical–subcortical circuitry interactions at context change inference, and at subsequent reduced uncertainty. In a second experiment, genetic risk for schizophrenia associated with similar cortical–subcortical findings were explored in an independent sample of 36 normal control subjects and 35 unaffected siblings during processing of intuitive number sequences along the number line, or during the inverse, less familiar, sequence. In the first experiment, reduced Bayesian models fitting subject behaviour suggest that patients with schizophrenia overestimated context change probabilities. Here, patients engaged anterior prefrontal cortex relatively less than healthy controls, in part driven by reduced effective connectivity from dorsolateral prefrontal cortex to anterior prefrontal cortex. In processing subsequent information indicating reduced uncertainty of their predictions, patients engaged relatively increased mid-brain activation, driven in part by increased dorsolateral prefrontal cortex to midbrain connectivity. These dissociable reduced and exaggerated prefrontal and subcortical circuit functions were accentuated in patients with delusions. In the second experiment, analogous dissociable reduced anterior prefrontal cortex and exaggerated midbrain engagement occurred in unaffected siblings when processing less expected versus more familiar number sequences. In conclusion, patients overestimated ambiguous context change probabilities with relatively reduced anterior frontal engagement. Subsequent reduced uncertainty about contextual state appeared over-reinforced, potentially contributing to confirmation bias and a cascade of aberrant belief processing about a more chaotic world relevant to delusions. These opposing cortical–subcortical effects relate in part to genetic risk for schizophrenia, with analogous imbalances in neural processing of less expected versus familiar information patterns.

1 Lieber Institute for Brain Development, Baltimore, Maryland, USA
2 Clinical Brain Disorders Branch, National Institute of Mental Health, Bethesda, Maryland, USA
3 Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
4 Neurology, Neuroscience, Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

Correspondence to: Dr Hao Yang Tan,
Lieber Institute for Brain Development,
855 North Wolfe Street,
Baltimore, Maryland 21205, USA

E-mail: Haoyang.tan@libd.org

Introduction

Fluctuations in real world information, from financial indices to social interactions, may arise from random variation, or from alterations in underlying contexts. Timely and accurate inferences about updating contexts from ambiguous information is important, but can be challenging and engage uniquely human elaborations of cortex (Koechlin and Hyafil, 2007; Fleming *et al.*, 2010). Models of belief formation posit the importance of prediction error signals and Bayesian inference in guiding appropriate learning of new contextual contingencies (Fletcher and Frith, 2009; Friston, 2012). Prediction error arises from the mismatch between expected and observed information. This signal was originally identified in the dopaminergic midbrain neurons (Schultz, 1992; Mirenowicz and Schultz, 1996) but later found to exist in other brain regions, including in its connections with striatum (Kim *et al.*, 2009) and prefrontal cortex (PFC) (Asaad and Eskandar, 2011). Thus the computation of prediction error signals to guide learning and updating of new contingencies involves a distributed network of frontal–subcortical brain regions.

Aberrant processing of prediction error signals has been noted in delusions. Here, salience may be misattributed to less relevant stimuli, with exaggerated midbrain prediction error signalling, and reinforcement of entrenched beliefs in delusions (Corlett *et al.*, 2007; Menon *et al.*, 2011). However, less is known about the distributed cortical–subcortical engagement of the anterior prefrontal functions in processing probabilistic contextual inferences from noisy information, their roles in learning from subsequent feedback, and associated vulnerabilities in delusions. Indeed, we might expect that processing ambiguous information would be particularly challenging for patients with delusions, characterized by deficits and biases in inferential thinking (Garety and Freeman, 1999; Coltheart *et al.*, 2011; Balzan *et al.*, 2013).

We therefore examined the neural correlates of context inference from noisy dynamic information. We examined cortical–subcortical networks as subjects inferred and updated beliefs about context change (C), evaluated subsequent feedback (F) information that decreased uncertainty about their earlier inference, and processed information suggesting a stable context that is to be maintained (M). We focused on hierarchical anterior PFC and dorsolateral PFC functions. We posited that the anterior PFC role in processing contextual episodes (Koechlin *et al.*, 2003; Badre and D'Esposito, 2009) would include inferring and updating context changes, with less anterior PFC engagement for maintenance of stable context. We also examined the coordination of PFC function with parietal cortex known to be engaged in number processing and executive function (Dehaene *et al.*, 2003), and with midbrain (dopaminergic) regions engaged in prediction error processing and information updating (Schultz, 1992; Mirenowicz and Schultz, 1996; O'Reilly and Frank, 2006).

We then examined the implications for delusions in schizophrenia, and more basic information processing biases with relevance to genetic risk for schizophrenia. As prefrontal deficits, as well as inappropriate (increased) salience processing are implicated in delusions and schizophrenia (Corlett *et al.*, 2007; Fletcher and Frith, 2009; Menon *et al.*, 2011), we examined hypotheses of dissociable, reduced and exaggerated prefrontal connectivity at contextual belief inference and feedback salience processing in relation to delusions. In particular, we hypothesized that schizophrenia-related deficits in prefrontal function might miscalculate (overestimate) likelihoods of deviation from expectations representing context change, and that this dysfunctional (more chaotic) worldview may subsequently be over-reinforced with resultant influence on delusion severity. More fundamentally, we examined, in a second independent experiment with unaffected siblings of patients, if these effects relate to information processing biases against less expected information patterns in favour of preconceived or more familiar patterns, with implications on the genetic risk for schizophrenia.

Materials and methods

Subjects

Seventeen patients with schizophrenia and 24 normal control subjects, and a second independent sample of 36 healthy subjects and 35 unaffected siblings of patients with schizophrenia with high quality MRI data were included. All participants for this study were previously ascertained as part of the NIMH Clinical Brain Disorders Branch 'Sibling Study' (Egan *et al.*, 2001), for which they were given a Structured Clinical Interview for DSM-IV, a neurological examination, a neuropsychological assessment, and a screening MRI examination. They then underwent further functional MRI experiments detailed below. Exclusion criteria included an IQ <70, a history of prolonged substance abuse or significant medical problems, and any abnormalities found by EEG or MRI. The healthy subjects had no previous history of psychiatric or neurological disorders. All participants gave written informed consent and were reimbursed for their time. The Institutional Review Board of NIMH approved the study.

Experiment 1

Cognitive paradigm

In the first of two independent experiments, 24 normal control subjects and 21 patients with schizophrenia were trained on the dynamic numerical inference task (for ~10 min) prior to performing it in the MRI scanner. The event-related functional MRI task required subjects to infer changes in context integers driving a noisy number series varying around that context integer, which changed from time to time (Fig. 1). At the beginning of each set of trials, participants were presented with the context integer in green. They were then presented with a series of numbers varying stochastically [standard deviation (SD) 0.5] about the underlying context with stimulus-onset intervals jittered between 4 and 8 s. At a probabilistically defined point (hazard rate of 0.1), the underlying integer shifted up or down by one unit, with resultant changes in the noisy number series. All participants viewed an identical sequence of numbers for this task. Participants were to respond whenever they inferred that a context change had occurred by making an 'up' or 'down' button press, and updated their mental representation of the new context accordingly. 'Feedback' (F) trials were identified for each participant and defined as the trial following an indicated context change that led to a reduced uncertainty about the underlying context (i.e. prediction error < 0.5, corresponding to <1 SD of the generative random number series). Consistent with sharp decreases in prediction error observed following adaptation to context changes (Nassar *et al.*, 2010), each F trial represented the first (and generally largest) reduction in prediction error associated with the new context, thus signalling decreased uncertainty about the particular context change. Each set of 8 to 20 trials comprised one to four context changes. At the end of each set, subjects rated their confidence on a scale of 1 to 4

regarding their timely detection of context changes. Each participant performed a total of 29 number sets (320 trials) divided over two runs (total ~24 min).

Behavioural modelling

We adapted a reduced Bayesian model of belief updating in a changing environment (Nassar *et al.*, 2010) to model subject behaviour as they inferred changes in context information, and estimated associated context change probabilities driving their behaviour. Here, probability of a context change-point (cp) given the stimulus (X) is calculated at each time point, t, using Bayes' rule:

$$P(cp|X_t) = P(X_t|cp)\ P(cp)/P(X_t)$$

$$= P(X_t|cp)\ P(cp)/P(X_t|cp)\ P(cp) + P(X_t|\sim cp)\ P(\sim cp)$$

$$= N(X_t|\mu_t = \mu_{t-1} \pm 1, \sigma_t)\ H/N(X_t|\mu_t = \mu_{t-1} \pm 1, \sigma_t)\ H$$

$$+ N(X_t|\mu_t = \mu_{t-1}, \sigma_t)(1 - H)$$

where $N(X_t|\mu_t = \mu_{t-1} \pm 1, \sigma_t)$ is a normal distribution from which stimulus X_t is generated from a context integer (u_t, that may increase or decrease by one unit at an expected hazard rate H) and standard deviation σ_t. The variance (σ_t^2) of the predictive distribution at time t is a function of both the expected run length (r_t) and expectation of noise (η) from the generative distribution (Nassar *et al.*, 2010), as follows: $\sigma_t^2 = \eta^2 + \eta^2/r_t$. This follows the intuition that the variance at each time point is a function of expected noise as well as uncertainty whether the variance is due to a context change point. The latter is inversely related to run length, where the shorter the present run length or closer to a putative change point the more uncertainty, whereas if there was a longer run with more information gathered, there would be less uncertainty. At each time t, expected run length is weighted by
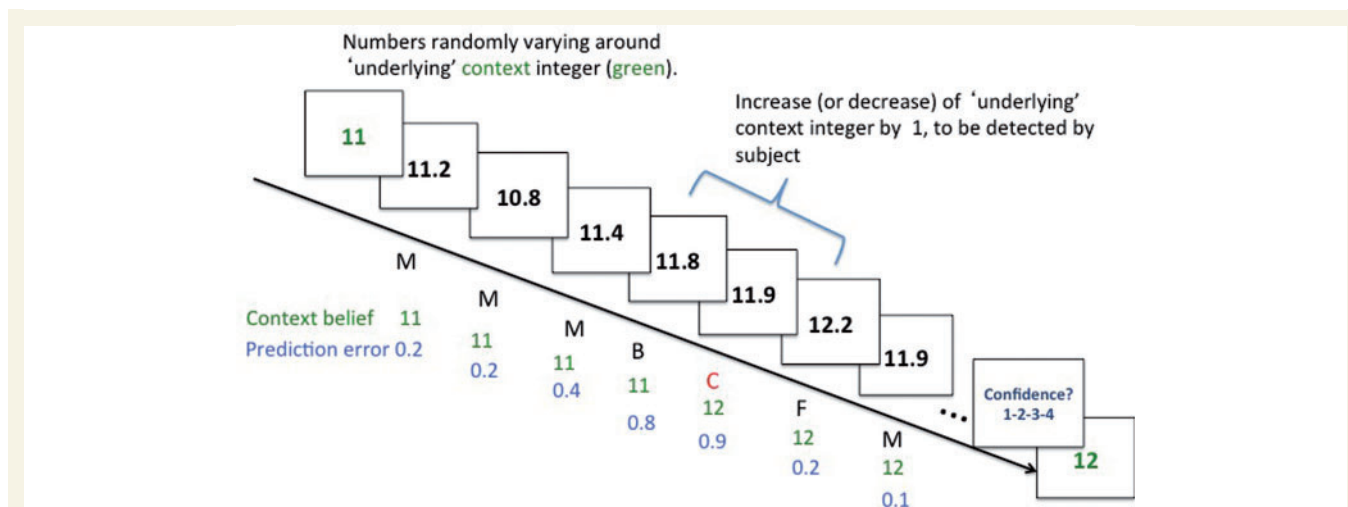

Figure 1 Dynamic numerical inference task in Experiment 1. Subjects were presented with numbers varying stochastically (SD = 0.5) about an underlying integer (green) that was shown once at the beginning of each set. At a probabilistically defined point (hazard rate = 0.1), this underlying integer shifted up or down by one unit. Based on the deviation of the presented information from their belief about the underlying integer (prediction error), subjects indicated when they inferred a context change (C) has occurred. Subsequent reduction in prediction error (designated 'feedback', F) would further reduce uncertainty about this change. M = small prediction errors (<0.5) that support maintaining the belief of a relatively stable context; B = larger prediction errors (>0.5).

the probability of a context change (Nassar *et al.*, 2010): $r_t = (r_{t-1} + 1) P(\sim cp|X_t) + P(cp|X_t)$. We defined $P(cp|X_t) > 0.5$ as giving rise to a model prediction that context change had occurred. Values of ŋ and H for each subject can then be optimized using a maximum likelihood approach to account for the individual's behaviour by minimizing the total squared difference between model predictions and subject behaviour. Associated posterior probabilities driving each subject's behavioural responses to context changes may then be estimated.

Functional imaging

Whole-brain blood oxygen level-dependent functional MRI data were collected on a 3 T scanner (General Electric Systems) using a gradient-echo echo-planar imaging pulse sequence acquisition of 24 contiguous slices (repetition time, 2000 ms; echo time, 30 ms; flip angle, 90°; field of view, 24 cm; matrix, 64 × 64; voxel size, 3.75 × 3.75 × 5 mm). The first four volumes of each functional time series were discarded to allow for signal saturation. Stimuli were presented via a back-projection system, and responses were recorded through a fibre optic response box, allowing for the measurement of reaction time for each trial.

The functional MRI data were preprocessed and analysed with SPM8 software (Wellcome Department of Cognitive Neurology, London, UK; http://www.fil.ion.ucl.ac.uk/spm) implemented in Matlab (Mathworks, Natick, MA). For each participant, images were slice timing corrected, realigned to the first volume in the time series, and corrected for head motion. Images were then spatially normalized to the standard echo planar image template provided by the Montreal Neurological Institute (MNI) using a 12-parameter affine model, and resampled to voxel size 2 × 2 × 2 mm. Spatial smoothing was applied with a Gaussian filter set at 8 mm full-width at half-maximum. After realignment, datasets were individually examined for motion correction <3 mm translation and $<2°$ rotation. This led to the exclusion of four patients from the original group of 21 patients.

Univariate analyses were conducted using the general linear model implemented in SPM8. Each task-evoked stimulus was modelled as a separate delta function and convolved with a canonical haemodynamic response function, and temporally filtered using a high-pass filter of 128 s. The differing task-evoked stimulus events were modelled as follows: context encoding (green integer presented at the start of each set); inferred context change-point (C) as indicated by a button-press response by the participant; feedback (F), defined as the stimulus following C associated with prediction error <0.5 from the new inferred context (corresponding to the generative distribution of the stimuli, which had an SD 0.5). F represented the first and generally steepest decrease in prediction error (and modelled context change probability) following a particular C response. We also modelled subsequent numerical stimuli with prediction error within 0.5 and thus indicating a stable context to be maintained (M); numerical stimuli with prediction error >0.5 from context without an inference of change-point (B); and confidence ratings. Incorrect responses were those that resulted in consecutive increased prediction error >1 in the opposite direction, and modelled as regressors of no interest, as were residual movement parameters. Here, we report only planned contrasts of interest at C, F and M events. See the 'Results' section for further justification for definition of C, F and M events. These contrasts

were subsequently taken to a second-level group analysis in which inter-subject variability was treated as a random effect. We examined the main effects at the C, F and M task phases at thresholds of $P < 0.05$ corrected for whole brain family-wise error (FWE) across all subjects. Contrasts across patient and control groups at C, F and M were evaluated within regions of interest in the anterior PFC, dorsolateral PFC, parietal cortex and midbrain, at $P < 0.001$ uncorrected and $P < 0.05$ small volume corrected. These orthogonal regions of interest were defined as significantly activated voxels ($P < 0.05$ FWE whole brain corrected) in the respective task across all subjects, that were also within a 15-mm radius sphere from the peak activated voxel in Wake Forest University Pickatlas (www.fmri.wfubmc.edu/download.htm) defined Brodmann area (BA) 10 for anterior PFC, BA 9/46 for dorsolateral PFC, BA7/40 for parietal cortex, or all significantly activated voxels in smaller atlas-defined midbrain region of interest.

Dynamic Causal Modelling

We used dynamic causal models (DCM version 10, as implemented in SPM8; Friston *et al.*, 2003) to investigate how prefrontal, parietal and midbrain regions interacted during C and F task phases. Dynamic causal models enable the estimation of the strength and direction of these regional interactions (Friston *et al.*, 2003). Using Bayesian parameter estimation, the observed blood oxygen level-dependent responses were compared with that predicted by a neurobiologically plausible model. This model comprised parameters describing how regional neural activity and their interactions were influenced by the external inputs (at C and F), as well as describing how these neuronal effects were biophysically linked to form the blood oxygen level-dependent responses in the imaging data acquired.

In constructing the dynamic causal models, we first selected time series from 10-mm radius regions of interest in the left anterior PFC, dorsolateral PFC, parietal cortex and midbrain regions of interest that were robustly engaged across the sample ($P < 0.05$ FWE corrected, at each at C and F task phase), and for the anterior PFC and midbrain (see 'Results' section) also differentially engaged between patients and controls ($P < 0.05$ small-volume corrected). This was in order to study nodes that were potentially vulnerable in schizophrenia and of disease relevance. From within these regions of interest, we then extracted time-series data from each individual subjects' task-specific *t*-contrast map that met an activation threshold of $P < 0.01$ uncorrected, as well as within 10 mm of the group level peak. This resulted in 0 to 4 runs ($<10\%$) being excluded for each task phase per group of healthy or patient subjects because of lack of activation in a particular subject, but with no significant differences in distribution of excluded runs across groups.

While there is evidence in neuroanatomical and lesion studies that dorsolateral PFC and parietal cortex receive information from sensory areas, and that efferent and afferent pathways link dorsolateral PFC to parietal and midbrain regions (Leh *et al.*, 2010), connectivity of the anterior PFC is less understood. Thus, for each C and F task phase, we examined four plausible models. Task-related modulation was set to occur bidirectionally between dorsolateral PFC, parietal cortex and midbrain, but differed on whether anterior PFC communicated to midbrain directly or through dorsolateral

PFC, and whether external inputs to anterior PFC occurred directly or through dorsolateral PFC (Badre and D'Esposito, 2009). Models were defined and estimated separately for the C and F task phases. Bayesian model selection (Penny *et al.*, 2004; Stephan *et al.*, 2009) determined the winning model of the observed blood oxygen level-dependent data based on a balance of fit and parsimony. Bayesian inversion of each model approximated the logarithm of the model evidence (i.e. the probability of the observed data given the model) through the negative variational free-energy (Stephan *et al.*, 2009). This approximation served as a lower bound on the model evidence, and represented the relative fit of the model based on model complexity and the number of parameters. We used random-effects Bayesian model selection at the group level for optimal model selection (Stephan *et al.*, 2009). Although similar results were obtained within each healthy and schizophrenia group, we reported Bayesian model selection with all healthy subjects and patients together. This allowed for subsequent analyses of between-group differences on the same model parameters using classical inference. For this, we identified significantly task-modulated connectivity using a one-sample *t*-test ($P < 0.05$ Bonferroni corrected for the number of tests), before testing for group differences in each of these connections. Finally, we examined for differences in connectivity in these illness-related prefrontal networks in patients with schizophrenia with or without significant delusions (the former defined as scoring moderate and above on the delusions scale of the Positive and Negative Symptom Scale) (Kay *et al.*, 1989).

Experiment 2

We then examined relationships of the same prefrontal and midbrain findings with more basic number pattern processing differences and genetic risk for schizophrenia in an independent sample of controls ($n = 35$) and unaffected siblings of patients with schizophrenia ($n = 36$). Here, we extended hypotheses from Experiment 1 that there were schizophrenia-related prefrontal deficits in processing deviations from expectations (at context change-point), and opposing changes as uncertainty is reduced (increased familiarity, at feedback). In Experiment 2, we tested if analogous processing of relatively less expected number sequences versus more natural, familiar sequences might have similarly dissociable findings at the same regions of interest in relation to genetic risk for schizophrenia. We used a well-established spatial number-line effect, where a smaller to larger number sequence in a left to right configuration (e.g. 2 7) would be more pre-potent than the converse sequence (e.g. 7 2) (Dehaene *et al.*, 1993; de Hevia *et al.*, 2014; Rugani *et al.*, 2015). We examined randomly balanced trials of each sequence configuration during encoding in an event-related working memory paradigm (Tan *et al.*, 2007*a*). The two numbers at encoding were also balanced in terms of distance with each other, as well as subsequent events of maintenance or mental arithmetic during working memory, though we will not be reporting on these other task phases here. By studying unaffected siblings, we thereby extended our inferences to putative genetic risk for schizophrenia, free of potential confounds of treatment and other lifestyle differences in patients (Egan *et al.*, 2001).

The MRI scanner, sequence and preprocessing in SPM8 were the same as in Experiment 1. Similar event-related modelling in SPM8 was used to examine the encoding of the two differing number-sequence configurations, as well as the other working memory events (Tan *et al.*, 2007*a*). We examined only the healthy control subjects and unaffected sibling effects across the two number sequence configurations and only at regions of interest centred on the exact same anterior PFC and midbrain peaks differentially engaged between the independent healthy controls and patients with schizophrenia sample in Experiment 1. These orthogonally-defined region of interests were then extracted from regions that were also robustly engaged at the group level in Experiment 2 ($P < 0.05$ FWE corrected) and within 10-mm radius spheres centred at the same corresponding coordinates differentially engaged from Experiment 1.

Results

Experiment 1

Demographics and behavioural performance

In the first of two experiments, we acquired high quality functional MRI data from 24 healthy subjects and 17 patients with schizophrenia. Subjects were not significantly different across age, gender and measures of premorbid IQ (wide range achievement test). However, as expected, patients had significantly reduced years of education and present IQ (estimated from the Wechsler Adult Intelligence Scale – Revised); they had a mean ($\pm$SD) duration of illness of 8.02 ± 7.02 years with a mean age of onset 19.4 ± 3.9 years, mean total positive and negative symptom score of 58 ± 22 and were on 295 ± 391 mg/day chlorpromazine-equivalent dose of antipsychotics.

Both normal control and schizophrenia groups performed significantly better than chance (controlled for number of button presses, $P < 0.0001$). Individual subjects varied in their responses as to when they perceived context changed with respect to the same stimuli stream in the MRI scanner (Supplementary Fig. 1A–E). However, there were no significant differences between normal control subjects and patients with schizophrenia for total number of context changed responses or reaction times. There were no significant differences between normal control subjects and patients with schizophrenia in the number of feedback events; normal control subjects and patients with schizophrenia experienced similar numbers of events where stimuli had prediction errors < 1 SD (0.5) from the subject's belief (M), i.e. where context was perceived to be stable; they also experienced similar numbers of events where prediction errors were >1 SD from their underlying context belief. Normal control subjects and patients with schizophrenia had similar confidence ratings during the task ($P > 0.2$).

However, patients with schizophrenia ended up with larger summed prediction error across the entire task (320 trials, $P = 0.026$, Table 1). This less optimal outcome resulted from patients responding with a higher learning

rate when they perceived context change points ($P < 0.01$, Fig. 2A). As learning rate is inversely proportional to prediction error according to the delta rule (Williams, 1992) used in the reduced Bayesian model (Nassar *et al.*, 2010) we implemented, and context change was fixed at 1 unit in our task, this corresponded with patients responding to smaller prediction errors eschewed by controls as noise. Patients also perceived context stability (the number of trials before a context change) to be shorter than controls ($P < 0.02$, Fig. 2B).

Table 1 Demographic and behavioural task performance of controls and patients with schizophrenia

	Total	Normal control subjects	Patients with schizophrenia
n	41	24	17
Gender, males, *n*	22	10	12
Age, years	27.4 (6.6)	27 (6.2)	27.9 (7.3)
Mean response reaction time, s	1.54 (0.43)	1.43 (0.29)	1.69 (0.55)
Total responses, *n*	34 (11)	31 (8)	37 (14)
Prediction error sum	199.06 (12.62)	195.42 (9.45)	204.20 (14.89)

Prediction error sum is calculated across all trials. Standard deviation is in parentheses.

The reduced Bayesian model (Nassar *et al.*, 2010) fit normal control and patient behaviour similarly as they inferred context change points (Supplementary Fig. 2). These models were associated with higher estimated posterior probabilities of change at C in patients with schizophrenia relative to control subjects ($P < 0.001$, Fig. 2C). Correspondingly, the subsequent decrease in estimated change point probabilities between the C and F phases, as more information at feedback reduced the likelihood of a context change, was greater in patients with schizophrenia ($P < 0.001$). Estimates of change-point probabilities were, however, not different between patients with schizophrenia and control subjects across the F to M transitions, and during the F, M and B task phases. Given the modelled behavioural differences in patients with schizophrenia at C and F but not M, we posited that differences in brain function at C and F might emerge between patients and normal control subjects at C and F, but not M.

Task-related activations in functional MRI

Healthy control subjects and patients with schizophrenia robustly engaged dorsolateral PFC, anterior PFC, parietal and midbrain activity as they inferred that an underlying context change (C) had occurred ($P < 0.05$ FWE corrected,
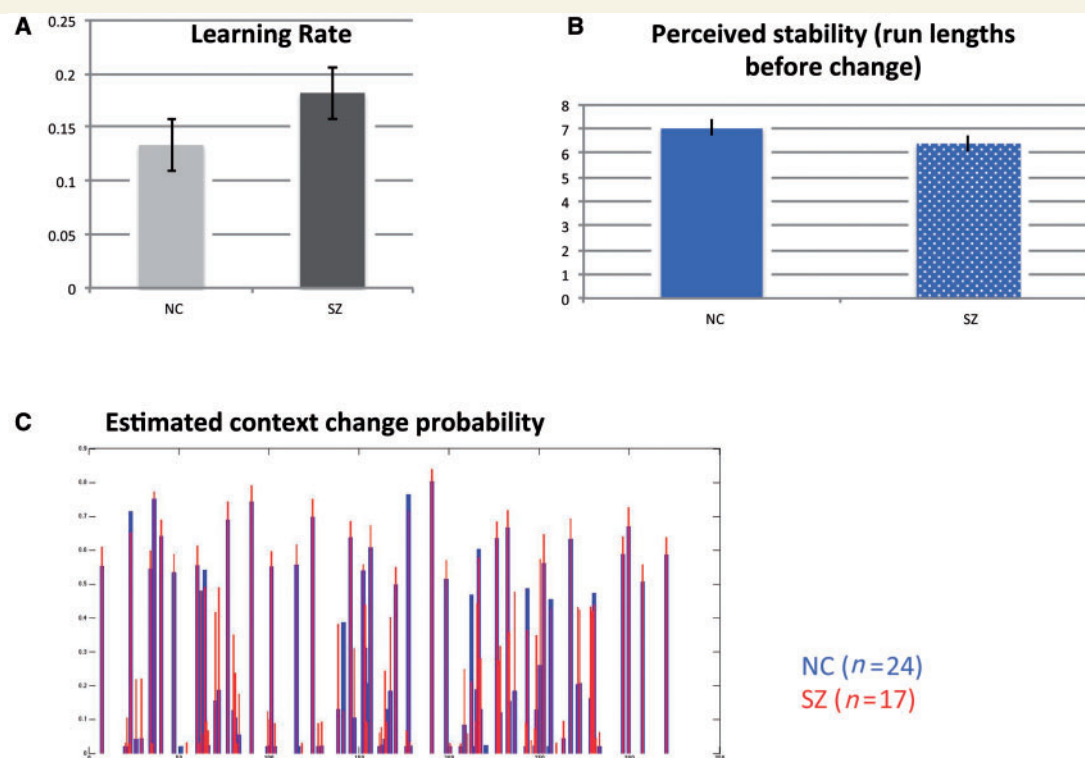

Figure 2 Behavioural performance in Experiment 1. (**A**) Patients with schizophrenia (SZ) responded with higher learning rates ($P < 0.01$) than controls. (**B**) Patients also perceived the average number of trials before a context change to be shorter ($P < 0.02$). (**C**) Mean Bayesian estimates of context change probabilities at C in patients and normal control subjects (NC) across trial events (320 trials on horizontal axis). Relative to normal control subjects, there was overestimation of modelled posterior probabilities of change in patients with schizophrenia ($P < 0.001$). Error bars are ± 1 standard error.

Fig. 3 and Supplementary Table 1). As they processed subsequent data representing the largest decrease in uncertainty about their new belief (feedback, F), similar brain regions were engaged (Supplementary Table 2). When context was stable (M), however, dorsolateral PFC and parietal brain regions were engaged, but not the anterior PFC or midbrain (Fig. 3 and Supplementary Table 1).

Contrasting patients with schizophrenia and healthy control subjects in the regions of interests (Fig. 4) during C, control subjects engaged anterior PFC region of interest activity (peak −46 44 12, *t* = 3.55, *P* < 0.05 small volume corrected) to a greater extent than patients with schizophrenia. On the other hand, at F, patients with schizophrenia engaged the midbrain region of interest to a greater extent (peak −4 −18 −8, *t* = 3.13, *P* < 0.05 small volume corrected, Fig. 4). There were no significant dorsolateral PFC and parietal cortex activation differences between healthy control subjects and patients in processing M.

Dynamic Causal Modelling

We then used dynamic causal models to investigate how prefrontal, parietal and midbrain regions interacted during each C and F task phase. We examined effective connectivity across the following region of interests: anterior PFC region differentially engaged between patients with schizophrenia and control subjects (−46 44 12) that was also robustly engaged in the combined sample (*t* > 5 at C or

F, *P* < 0.05 FWE corrected), dorsolateral PFC (− 34 30 16, *t* > 8 at C or F, *P* < 0.05 FWE corrected), parietal cortex (−58 −58 38, *t* > 9 at C or F, *P* < 0.05 FWE corrected) and midbrain (0 −16 −18, *t* > 8 at C or *t* > 3.4 at F, *P* < 0.05 small-volume corrected). Of the four models tested, Bayesian model selection favoured the connectivity model where anterior PFC interacted through dorsolateral PFC, and with dorsolateral PFC engaging parietal cortex and midbrain connectivity (Fig. 5). This was the winning model whether control subjects and patients with schizophrenia were grouped together, or estimated separately.

Significant task-modulated effective connectivity for C and F are shown in Fig. 6A and D. At C, task-related information entered at dorsolateral PFC and parietal cortex with subsequent reciprocal task-modulated effective connectivity between dorsolateral PFC and parietal cortex; there was also dorsolateral PFC to anterior PFC connectivity, and dorsolateral PFC to midbrain and parietal cortex to midbrain connectivity (*P* < 0.01). At F, task-related information entered at dorsolateral PFC and parietal cortex; there was dorsolateral PFC to parietal cortex task-modulated effective connectivity, reciprocal dorsolateral PFC to anterior PFC connectivity, as well as dorsolateral PFC to midbrain and parietal cortex to midbrain connectivity (*P* < 0.01).

During information processing at C, patients with schizophrenia had deficits in task-modulated cortical effective
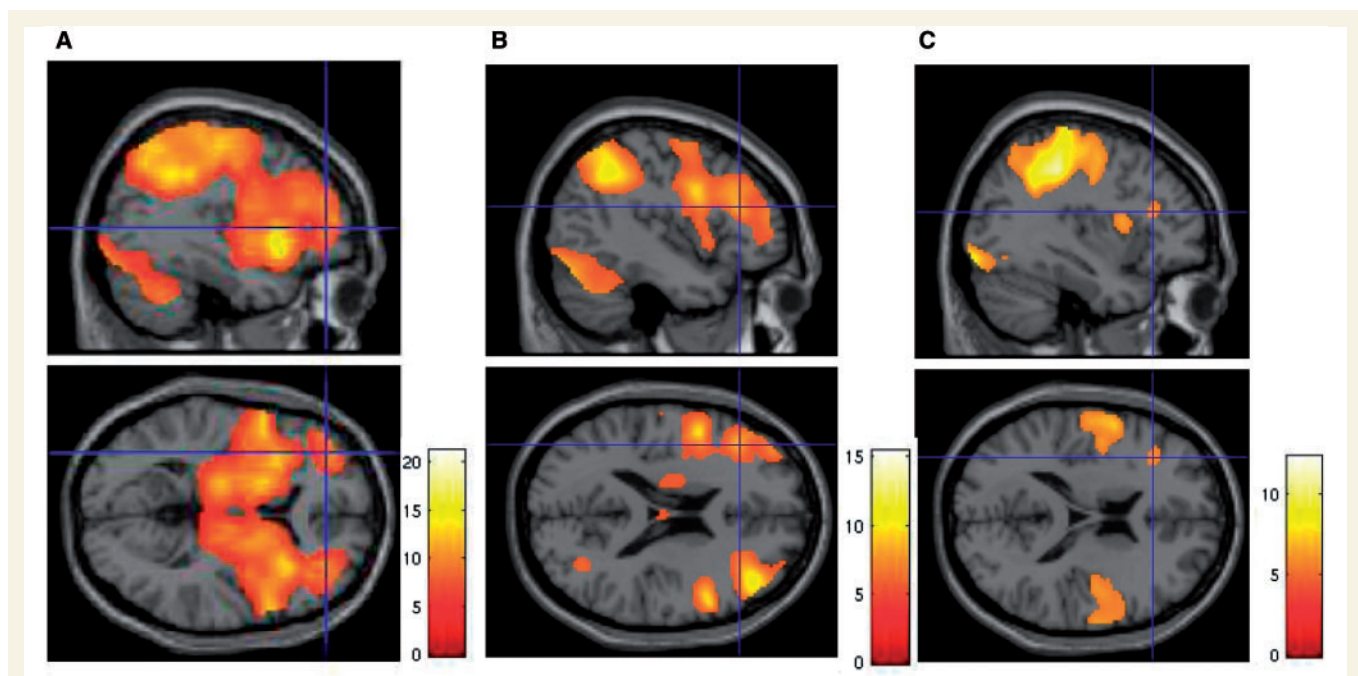
Figure 3 Task-related activation in Experiment 1. (**A**) When processing information perceived to infer a change in its underlying contextual structure (C), regions in the anterior PFC, dorsolateral PFC, parietal cortex, striatum and were engaged (*n* = 17 patients with schizophrenia and *n* = 24 normal control subjects, *P* < 0.05 FWE corrected). (**B**) Anterior PFC, DLFPC, parietal and striatal regions were engaged when processing subsequent information supporting the change decision (F) (*P* < 0.05 FWE corrected). (**C**) Dorsolateral PFC and parietal regions were engaged but relatively less robustly when processing information perceived to be consistent with maintenance of a stable context (M, *P* < 0.05 FWE corrected).
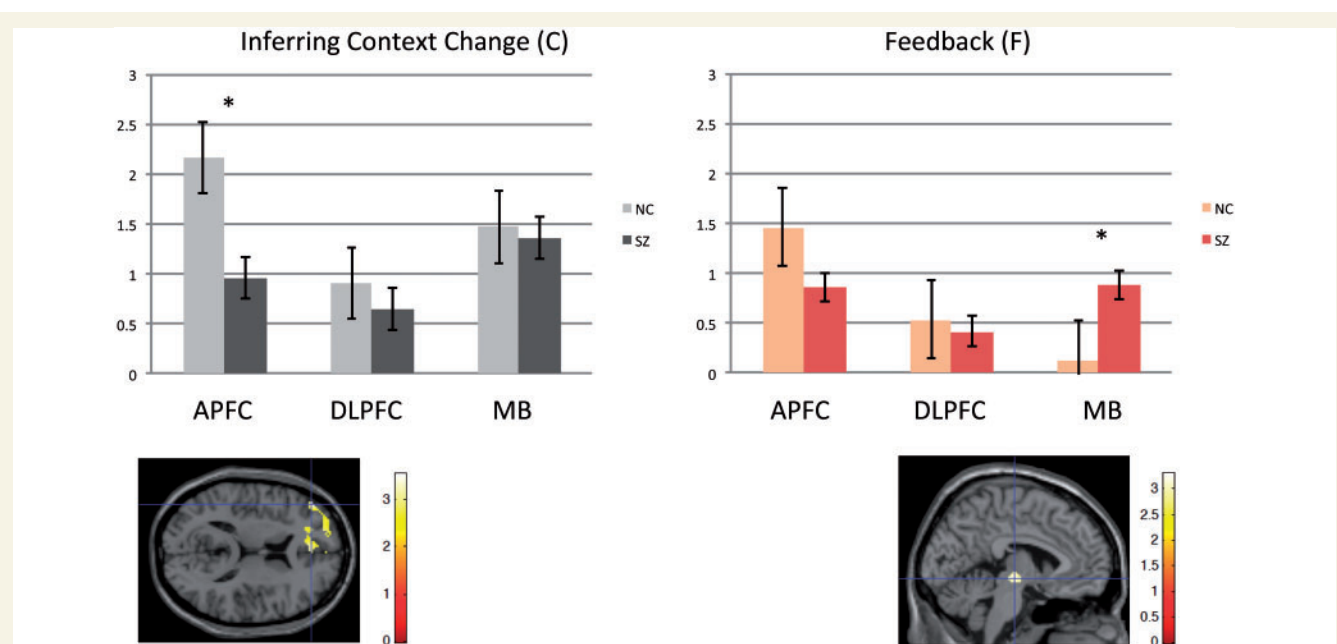
Figure 4 Activation at anterior PFC (APFC), dorsolateral PFC (DLPFC), and midbrain (MB) regions of interest across C and F task phases. Schizophrenia patients (SZ) had reduced anterior PFC engagement at C, but increased midbrain engagement at F (*$t > 3$, $P < 0.05$ small volume corrected). Error bars are ± 1 standard error. NC = normal control subjects.
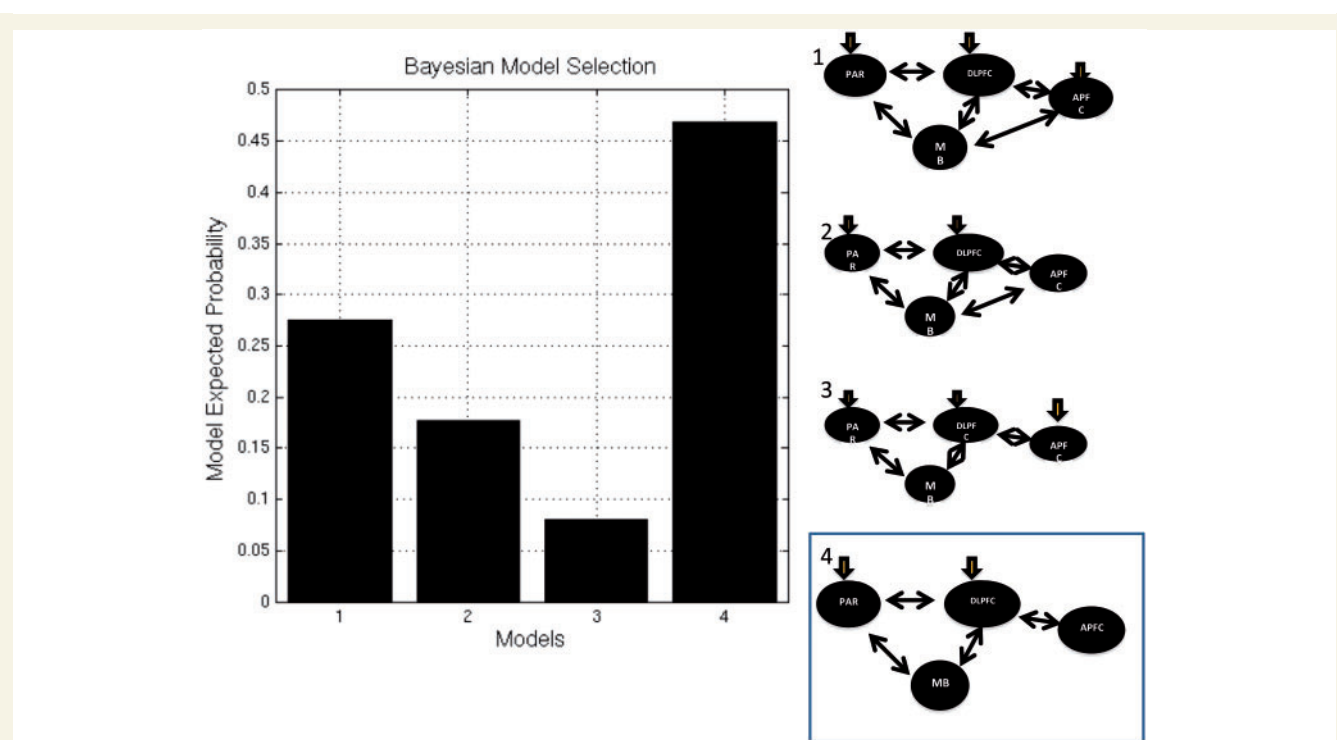

Figure 5 Dynamic causal models. Bayesian model selection favouring an effective connectivity model (#4), where anterior PFC communicates with dorsolateral PFC; and dorsolateral PFC with parietal cortex and midbrain regions of interest. Task information inputs were at dorsolateral PFC and parietal cortex.

connectivity at anterior PFC, dorsolateral PFC and parietal cortex ($P < 0.05$, Fig. 6A). There was reduced effective connectivity from dorsolateral PFC to anterior PFC, and between dorsolateral PFC and parietal cortex (Fig. 6B).

During information processing in F, patients had increased task-modulated dorsolateral PFC to midbrain and parietal cortex to midbrain effective connectivity ($P < 0.05$, Fig 6D and E).
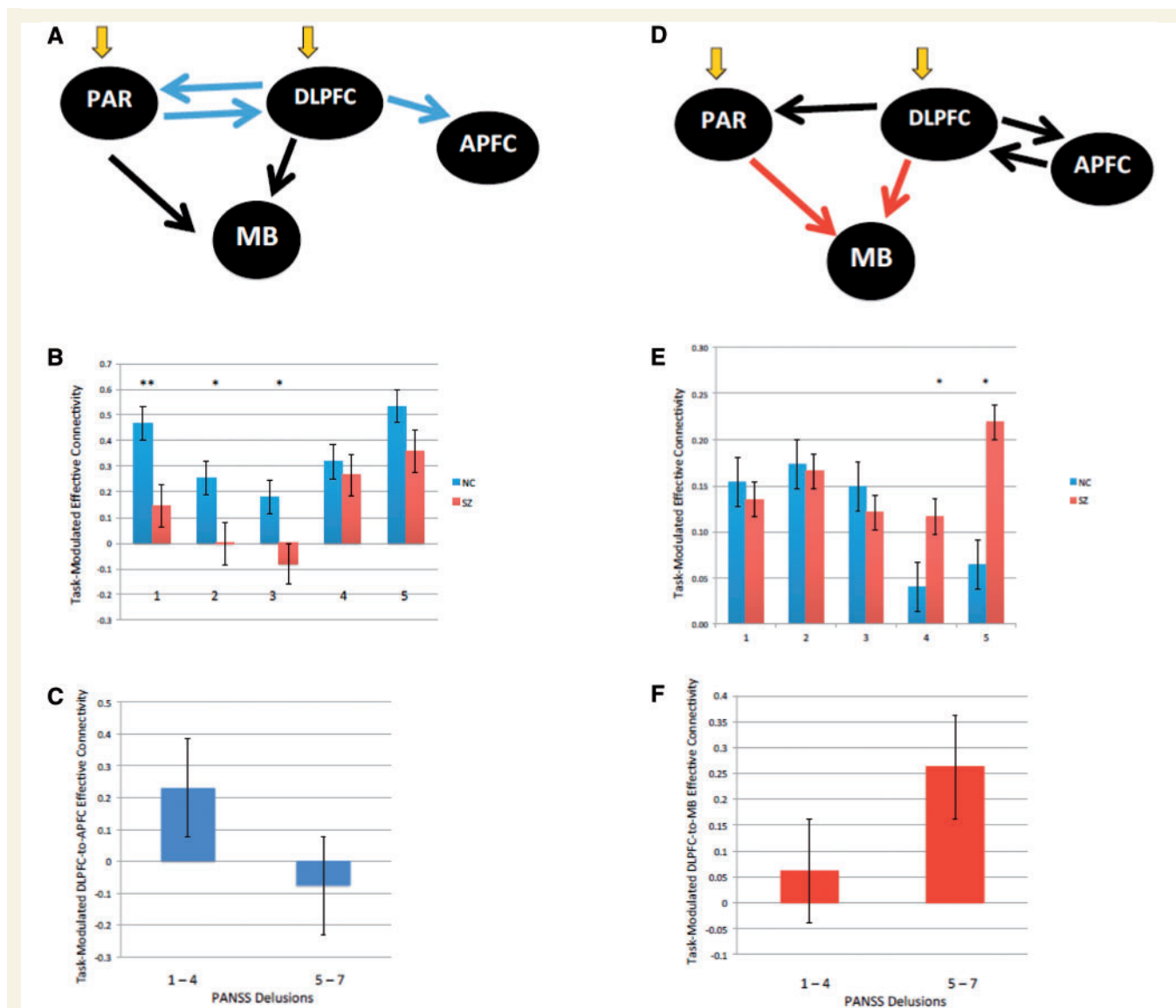
Figure 6 Prefrontal network effective connectivity in Experiment 1. (**A**) Processing information perceived to infer a change in its underlying contextual structure (C), patients with schizophrenia (SZ) had deficits in task-modulated cortical effective connectivity at parietal cortex (PAR), dorsolateral PFC (DLPFC) and anterior PFC (APFC) ($P < 0.05$, blue arrows). Orange arrows denote task information input (C) into the system. Black arrows denote significant task-modulated effective connectivity ($n = 24$ healthy control subjects and $n = 17$ patients with schizophrenia, $P < 0.05$). (**B**) Task-modulated connectivity during C in normal control subjects and patients with schizophrenia from (1) dorsolateral PFC to anterior PFC; (2) dorsolateral PFC to parietal cortex; (3) parietal cortex to dorsolateral PFC; (4) dorsolateral PFC to midbrain; and (5) parietal cortex to midbrain. *$P < 0.05$, **$P < 0.005$. (**C**) Reduced task-modulated connectivity during C in patients with schizophrenia with significant delusions ($n = 7$ versus 10, $P < 0.05$). (**D**) When processing feedback favouring the change decision (F), patients had increased task-modulated dorsolateral PFC–midbrain and parietal cortex–midbrain effective connectivity ($P < 0.05$, red arrows). (**E**) Task-modulated connectivity during F in normal control subjects and patients with schizophrenia from (1) dorsolateral PFC to anterior PFC; (2) anterior PFC to dorsolateral PFC; (3) dorsolateral PFC to parietal cortex; (4) dorsolateral PFC to midbrain; and (5) parietal cortex to midbrain. (**F**) Relatively increased task-modulated dorsolateral PFC to midbrain connectivity during F in patients with schizophrenia with significant delusions ($n = 7$ versus 10, $P < 0.05$). NC = normal control subjects.

To evaluate the relationship with delusions, we examined for differences in connectivity of the prefrontal illness-related networks in patients with schizophrenia with or without significant delusions. There were no age, gender or task performance differences across these patient groups. However, we found that patients with delusions engaged dorsolateral PFC to anterior PFC connectivity relatively less at C

($P < 0.05$, Fig. 6C), but engaged dorsolateral PFC to midbrain connectivity more at F ($P < 0.05$, Fig. 6F)

Experiment 2

In Experiment 2, we examined if basic processing (encoding) of relatively less expected number sequences versus more

naturally expected (familiar) number sequences might have similarly dissociable findings at the same anterior PFC (−46 44 12) and midbrain (−4 −18 −8) regions of interest differentially engaged in Experiment 1, in an independent sample of healthy control subjects ($n = 37$) and unaffected siblings ($n = 36$). There were no behavioural or demographic differences across healthy control subjects and unaffected siblings (Supplementary Table 3). Regions of interest from the same anterior PFC and midbrain regions were robustly engaged during encoding of each of the two configurations of number sequences at the group level ($P < 0.05$ FWE corrected, Fig. 7 and Supplementary Table 4). Across groups at the regions of interest, unaffected siblings engaged relatively increased midbrain activation when processing numbers in the more familiar number line sequence (Fig. 7A, $P < 0.05$), but engaged relatively reduced anterior PFC processing in the opposing less pre-potent number sequence (Fig. 7B, $P < 0.05$).

Discussion

We studied the neural correlates of patients with schizophrenia and healthy controls as they made inferences about changing contexts in a dynamic numerical task. We extended these findings to neural information processing biases against deviations from expectation in favour of preconceived or familiar information patterns, and their relevance to genetic risk for schizophrenia. Patients apparently overestimated noisy context change probabilities, and engaged anterior PFC relatively less than healthy controls, with reduced effective connectivity from dorsolateral PFC to anterior PFC. In processing subsequent feedback (F) reducing uncertainty about context, patients engaged relatively increased mid-brain activation, associated with increased dorsolateral PFC to midbrain connectivity. These dissociable reduced and exaggerated prefrontal and subcortical circuit functions were also accentuated in relation to delusions in patients. Analogous dissociable reduced anterior PFC and exaggerated midbrain engagement occurred in unaffected siblings of patients when processing less expected versus more familiar number sequences. Thus unaffected siblings also appeared to tackle relative deviations from expectations with relatively reduced anterior frontal engagement, but the converse with increased midbrain engagement.

In Experiment 1, we examined distributed anterior PFC, dorsolateral PFC, parietal and subcortical networks in
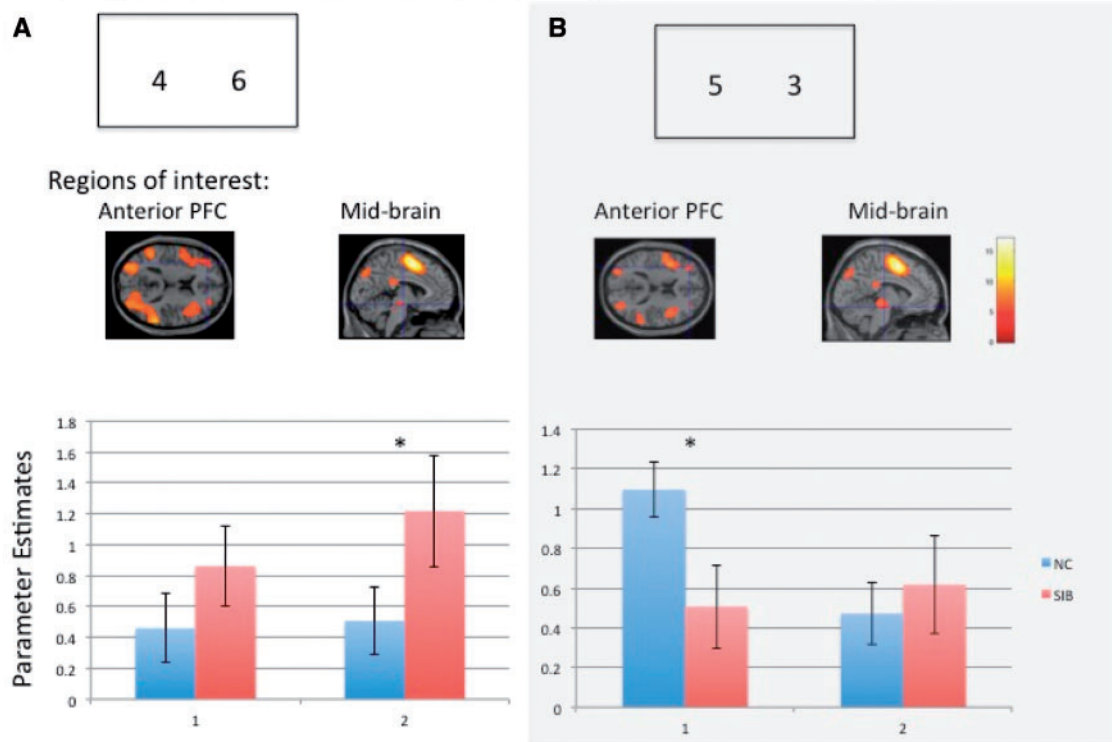

Figure 7 Controls and unaffected siblings' activations in regions-of-interest during Experiment 2. Encoding in working memory of two numbers in the familiar number-line sequence (**A**) versus the opposing less expected sequence (**B**) in healthy control subjects (NC; $n = 37$) and unaffected siblings (SIB, $n = 36$). Regions of interest were extracted from robustly activated regions ($P < 0.05$ FWE corrected) centred at the same anterior PFC (1: −46 44 12) and midbrain (2: −4 −18 −8) regions differentially engaged in Experiment 1. Analogous to the findings in Experiment 1, unaffected siblings engaged relatively increased midbrain activation ($P < 0.05$) while processing numbers in the expected number line sequence (**A**), but engaged relatively reduced anterior PFC processing in the less pre-potent number sequence (**B**).

inferring context change from dynamic information streams. This information was symbolic, abstract and noisy. Inferences about change in information structure (C) engaged anterior PFC, which at least in part occurs through dorsolateral PFC. This is consistent with the role of anterior PFC in updating changes in higher-order contextual episodes (Donoso *et al.*, 2014) and with the conceptualization that the anterior PFC processing is hierarchically engaged through dorsolateral PFC (Koechlin *et al.*, 2003; Badre and D'Esposito, 2009). When there was perceived instability in context driving the inference that it has changed (C), cortical (anterior PFC, dorsolateral PFC and parietal cortex) information processing was also integrated with subcortical regions (e.g. midbrain). This is reflective of dopaminergic prediction error signalling (Schultz *et al.*, 1997), it's distributed cortical processing (Asaad and Eskandar, 2011), and their combined roles in information updating. These dopaminergic processes have been suggested to engage D2-mediated gating of new information into cortex (Seamans and Yang, 2004; O'Reilly and Frank, 2006).

As uncertainty surrounding the contextual prediction was subsequently reduced at F, anterior PFC, dorsolateral PFC and subcortical networks were engaged, but these were reduced at M when context was putatively stable. This suggests that the anterior PFC was engaged in the control of changes to contextual information processing (Donoso *et al.*, 2014), rather than in the stable maintenance of this information. The concurrent engagement of midbrain function through dorsolateral PFC at C but not M would also be consistent with the specific computational role of dopamine (D2) systems in the updating and gating of new information into cortical processing (Seamans and Yang, 2004; O'Reilly and Frank, 2006).

In patients with schizophrenia, however, there was aberrant neural processing of contextual beliefs, which showed an association with delusions. Overall, patients accumulated more prediction errors than healthy controls. Patients engaged relatively dysfunctional belief inference and updating, characterized by an overestimation of posterior probabilities of context change. This occurred at relatively smaller prediction errors eschewed by healthy control subjects as noise, and was associated with reduced anterior PFC activity and effective connectivity. These findings are consistent with conceptualizations of schizophrenia as a disorder of connectivity (Weinberger *et al.*, 1992; Friston and Frith, 1995; Meyer-Lindenberg *et al.*, 2001), and with information processing deficits that misattributed significance to noise in the sensory environment (Fletcher and Frith, 2009; Jardri and Denève, 2013). Earlier work also suggested patients have reduced prefrontal-parietal connectivity in higher cognitive tasks that relate to dysfunction in dopaminergic and glutamatergic brain systems (Farber, 2003; Egan *et al.*, 2004; Honey *et al.*, 2004; Tan *et al.*, 2007*b*, 2012). Here, our data further suggest that dysfunction involving anterior PFC, and reduced dorsolateral PFC to anterior PFC connectivity, may be associated with

aberrant computations about expectations and belief inference, with relevance to delusions. The processing of beliefs is consistent with anterior PFC being phylogenetically and ontogenetically later developing (Koechlin, 2011), and at the apex of a cascade of posterior to anterior prefrontal cortical circuits engaged in increasingly complex contextual processing (Koechlin *et al.*, 2003). Thus, anterior PFC's coordinated information processing with dorsolateral PFC would ostensibly contribute to human probabilistic inferences about environmental information, and when dysfunctional, be implicated in delusional beliefs in psychosis.

Further, our data suggest that dysfunctional subcortical processing also contributes to abnormal contextual belief processing. The misattribution of salience has been implicated in delusion formation (Corlett *et al.*, 2007; Fletcher and Frith, 2009; Menon *et al.*, 2011). Our data suggest that neural processing of information that first decreased uncertainty about the prior context change prediction (F) in patients engaged relatively increased dopamine-associated midbrain activation and dorsolateral PFC to midbrain information processing. Patients with significant delusions had relatively accentuated engagement of this midbrain function. Similar opposing cortical and subcortical relationships have been modelled in relation to deficits in cortical dopaminergic D1 (and associated glutamatergic) signalling, and reciprocal exaggerated subcortical D2 signalling (Akil *et al.*, 2003; Meyer-Lindenberg *et al.*, 2005; Tan *et al.*, 2012). We suggest these basic predictions (Akil *et al.*, 2003) may extend to complex human belief processing at anterior PFC-related cortical circuits at C, and subsequently dorsolateral PFC-midbrain circuitry at F. Dysfunction in these neural circuit mechanisms may have relevance to delusion formation in schizophrenia, potentially through biased salience about one's miscalculated probabilistic inferences and their over-reinforcement. Indeed, converging evidence supports a Bayesian model of disequilibrium between prior beliefs and sensory evidence, with a bias towards overestimation of the precision of sensory information and their reinforcement in delusions (Adams *et al.*, 2013; Jardri and Denève, 2013).

These dissociable anterior PFC and midbrain neural processing biases against deviations from expectation versus more familiar information patterns were extended to putative genetic risk for schizophrenia. The tasks in Experiments 1 and 2 have overlapping similarities. They had similar timings and involved thinking about number sequences. In Experiment 1, patients with schizophrenia and healthy control subjects detected larger than expected differences at context change, and processed sequences with increased familiarity. In Experiment 2, unaffected siblings and controls processed sequences that were in a less expected reverse number-line sequence, which are less familiar than those in a number-line sequence (Dehaene *et al.*, 1993; de Hevia *et al.*, 2014; Rugani *et al.*, 2015). If indeed dysfunction in anterior PFC affects processing of probabilistic context change, we might also expect that the dysfunctional encoding of less expected context information to be at least

partly implicated, as we have found in Experiment 2 in unaffected siblings and controls. Relative dysfunction in processing less familiar or less expected information, particularly at higher cortical anterior PFC regions, could thus affect the unbiased evaluation and integration of this information into current beliefs. This would be consistent with data from Experiment 1 where probabilities associated with changing contexts were putatively overestimated in patients with schizophrenia. This also builds on recent computational models of delusion formation which suggest failure in the top–down control of information integration to update precise representations of abstract contexts relative to prior dysfunctional beliefs (Jardri and Denève, 2013). Our data further support the notion that vulnerabilities in context processing at anterior PFC is related to genetic risk for schizophrenia.

On the other hand, the processing of relatively pre-potent information appears over-reinforced. In Experiment 1, we found that information reducing uncertainty about current beliefs (F) in patients with schizophrenia over-engaged dopaminergic midbrain. In Experiment 2, analogous increased midbrain engagement occurred in unaffected siblings during encoding of more pre-potent number sequences along the mental number line. These findings are consistent with conceptualizations about the over-representation of prior beliefs in recent computational models of delusions (Jardri and Denève, 2013). Our data suggest the potential role of the midbrain in mediating these effects. This model is also an elaboration of a number of earlier representations of the cortical regulation of subcortical dopamine activity, and its dysfunction in psychosis (Weinberger, 1987; Saunders et al., 1998; Meyer-Lindenberg et al., 2005). Our data suggest these opposing cortical-subcortical mechanisms may relate to neural vulnerabilities for false beliefs and delusions associated with genetic risk for schizophrenia.

We note that alternative interpretations of our findings are possible, as behavioural physiology is invariably fraught with unknowns in underlying mechanisms. For instance, it is conceivable that our patients have less confidence in their performance on tests. The under-activation of cortical networks and their reduced connectivity might reflect this psychological state. By analogous reasoning, their receiving encouraging feedback about the correctness of their responses might enlist an exaggerated positive reward response, as they were less confident than controls that their choices would be correct. Nevertheless, the lack of differences in patients' and controls' confidence ratings obtained throughout the task despite differences in performance indices suggests this may be less likely. On the other hand, if indeed patients overestimated context change probabilities, they would have, as we found, at least similar if not higher confidence ratings as controls. Nevertheless, future work may be needed to better define these relationships, for example using explicit measurements of confidence windows (Nassar et al., 2010).

In our dynamic causal modelling analyses we have focused on a limited set of hypothesized regions of interest engaging higher cognitive functions vulnerable in schizophrenia. The winning model, while consistently so for both healthy control and schizophrenia groups modelled together and separately, was not estimated by Bayesian model selection as having much greater expected probability than the next best model tested. However, the model coincides with conceptualizations about the functional organization of anterior frontal cortex (Koechlin et al., 2003; Badre and D'Esposito, 2009). As model selection arguably should be informed by the significant weight of prior evidence, we have elected to report on connectivity estimates based on this model. Nevertheless, similar results were obtained using Bayesian Model Averaging (Penny et al., 2010) across the four models tested.

The extent to which antipsychotic medications confound these findings is a limitation. However, this applies only to Experiment 1, but not in Experiment 2 where we extend inferences to genetic risk for schizophrenia in unaffected siblings of patients. There was also no relationship between severity of delusions or associated connectivity findings with medication dose in Experiment 1. While delusions often co-occur with hallucinations, we did not find significant imaging effects with the latter. We are, however, unable to safely conclude about the specificity of our findings to only delusions because of a lack of power to exclude type II errors. Our cognitive task might also be better able to detect effects associated with delusions, while more auditory or visual-based tasks might better relate with auditory or visual processing disturbances (Teufel et al., 2015). Future studies should also investigate the neural correlates of subacute symptoms and cognitive changes in siblings or individuals at risk for psychosis.

In conclusion, we posit that the dysfunctional cognitive repertoire in delusions is associated with deficient engagement of anterior PFC at inferring changes in context beliefs, and subsequent exaggerated midbrain engagement. Patients with schizophrenia apparently overestimated the significance of noise in inferring change in environmental context, and did so with reduced anterior PFC engagement, that may at least in part represent dysfunction in integrating less pre-potent information. Subsequent reduced uncertainty about contextual state appeared over-reinforced, potentially contributing to confirmation bias, and the cascade of aberrant strengthening of beliefs about a more chaotic world relevant to delusions. These information processing vulnerabilities relate to the genetic risk for schizophrenia, and may be associated with excitatory and inhibitory imbalances in glutamatergic and dopaminergic function.

Acknowledgements

Funding

Supplementary material

Supplementary material is available at *Brain* online.

References

Adams RA, Stephan KE, Brown HR, Frith CD, Friston KJ. The computational anatomy of psychosis. Front Psychiatry 2013; 4: 47.

Akil M, Kolachana BS, Rothmond DA, Hyde TM, Weinberger DR, Kleinman JE. Catechol-O-Methyltransferase genotype and dopamine regulation in the human brain. J Neurosci 2003; 23: 2008–13.

Asaad WF, Eskandar EN. Encoding of both positive and negative reward prediction errors by neurons of the primate lateral prefrontal cortex and caudate nucleus. J Neurosci 2011; 31: 17772–87.

Badre D, D'Esposito M. Is the rostro-caudal axis of the frontal lobe hierarchical? Nat Rev Neurosci 2009; 10: 659.

Balzan R, Delfabbro P, Galletly C, Woodward T. Confirmation biases across the psychosis continuum: the contribution of hypersalient evidence-hypothesis matches. Br J Clin Psychol 2013; 52: 53–69.

Coltheart M, Langdon R, McKay R. Delusional belief. Annu Rev Psychol 2011; 62: 271–98.

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: 2387–400.

de Hevia MD, Girelli L, Addabbo M, Macchi Cassia V. Human infants' preference for left-to-right oriented increasing numerical sequences. PLoS ONE 2014; 9: e96412.

Dehaene S, Bossini S, Giraux, P. The mental representation of parity and number magnitude. J Exp Psychol 1993; 122: 371–96.

Dehaene S, Piazza M, Pinel P, Cohen L. Three parietal circuits for number processing. Cogn Neuropsychol 2003; 20: 487–506.

Donoso M, Collins AGE, Koechlin E. Foundations of human reasoning in the prefrontal cortex. Science 2014; 344: 1481–6.

Egan MF, Goldberg TE, Gscheidle T, Weirich M, Rawlings R, Hyde TM, et al. Relative risk for cognitive impairments in siblings of patients with schizophrenia. Biol Psychiatry 2001; 50: 98–107.

Egan MF, Straub RE, Goldberg TE, Yakub I, Callicott JH, Hariri AR, et al. Variation in GRM3 affects cognition, prefrontal glutamate, and risk for schizophrenia. Proc Natl Acad Sci USA 2004; 101: 12604–9.

Farber NB. The NMDA receptor hypofunction model of psychosis. Ann N Y Acad Sci 2003; 1003: 119–30.

Fleming SM, Weil RS, Nagy Z, Dolan RJ, Rees G. Relating introspective accuracy to individual differences in brain structure. Science 2010; 329: 1541–3.

Fletcher PC, Frith CD. Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. Nat Rev Neurosci 2009; 10: 48–58.

Friston K. The history of the future of the Bayesian brain. Neuroimage 2012; 62: 1230–3.

Friston KJ, Frith CD. Schizophrenia: a disconnection syndrome? Clin Neurosci (New York, NY) 1995; 3: 89–97.

Friston KJ, Harrison L, Penny W. Dynamic causal modelling. NeuroImage 2003; 19: 1273.

Garety PA, Freeman D. Cognitive approaches to delusions: a critical review of theories and evidence. Br J Clin Psychol 1999; 38 (Pt 2): 113–54.

Honey RAE, Honey GD, O'Loughlin C, Sharar SR, Kumaran D, Bullmore ET, et al. Acute ketamine administration alters the brain responses to executive demands in a verbal working memory task: an fMRI study. Neuropsychopharmacology 2004; 29: 1203–14.

Jardri R, Denève S. Circular inferences in schizophrenia. Brain 2013; 136: 3227–41.

Kay SR, Opler LA, Lindenmayer JP. The positive and negative syndrome scale (PANSS): rationale and standardisation. Br J Psychiat 1989; 155 (Suppl 7): 59–67.

Kim H, Sul JH, Huh N, Lee D, Jung MW. Role of striatum in updating values of chosen actions. J Neurosci 2009; 29: 14701–12.

Koechlin E. Frontal pole function: what is specifically human? Trends Cogn Sci 2011; 15: 241; author reply 243.

Koechlin E, Hyafil A. Anterior prefrontal function and the limits of human decision-making. Science 2007; 318: 594–8.

Koechlin E, Ody C, Kouneiher FT. The architecture of cognitive control in the human prefrontal cortex. Science 2003; 302: 1181–5.

Leh SE, Petrides M, Strafella AP. The neural circuitry of executive functions in healthy subjects and Parkinson's disease. Neuropsychopharmacology 2010; 35: 70–85.

Menon M, Schmitz TW, Anderson AK, Graff A, Korostil M, Mamo D, et al. Exploring the neural correlates of delusions of reference. Biol Psychiatry 2011; 70: 1127–33.

Meyer-Lindenberg A, Kohn PD, Kolachana B, Kippenhan S, McInerney-Leo A, Nussbaum R, et al. Midbrain dopamine and prefrontal function in humans: interaction and modulation by COMT genotype. Nat Neurosci 2005; 8: 594–6.

Meyer-Lindenberg A, Poline JB, Kohn PD, Holt JL, Egan MF, Weinberger DR, et al. Evidence for abnormal cortical functional connectivity during working memory in schizophrenia. Am J Psychiatry 2001; 158: 1809–17.

Mirenowicz J, Schultz W. Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. Nature 1996; 379: 449–51.

Nassar MR, Wilson RC, Heasly B, Gold JI. An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. J Neurosci 2010; 30: 12366–78.

O'Reilly RC, Frank MJ. Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. Neural Comput 2006; 18: 283–328.

Penny WD, Stephan KE, Daunizeau J, Rosa MJ, Friston KJ, Schofield TM, et al. Comparing families of dynamic causal models. PLoS Comput Biol 2010; 6: e1000709.

Penny WD, Stephan KE, Mechelli A, Friston KJ. Comparing dynamic causal models. NeuroImage 2004; 22: 1157.

Rugani R, Vallortigara G, Priftis K, Regolin L. Number-space mapping in the newborn chick resembles humans' mental number line. Science 2015; 347: 534–6.

Saunders RC, Kolachana BS, Bachevalier J, Weinberger DR. Neonatal lesions of the medial temporal lobe disrupt prefrontal cortical regulation of striatal dopamine. Nature 1998; 393: 169.

Schultz W. Activity of dopamine neurons in the behaving primate. Semin Neurosci 1992; 4: 129–38.

Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. Science 1997; 275: 1593–9.

Seamans JK, Yang CR. The principal features and mechanisms of dopamine modulation in the prefrontal cortex. Prog Neurobiol 2004; 74: 1–57.

Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ. Bayesian model selection for group studies. NeuroImage 2009; 46: 1004.

Tan HY, Chen AG, Kolachana B, Apud JA, Mattay VS, Callicott JH, et al. Effective connectivity of AKT1-mediated dopaminergic working memory networks and its relationship to the pharmacogenetics of cognition in schizophrenia Brain 2012; 135: 1436–45.

Tan HY, Chen Q, Goldberg TE, Mattay VS, Meyer-Lindenberg A, Weinberger DR, et al. Catechol-O-methyltransferase Val158Met modulation of prefrontal-parietal-striatal brian systems during arithmetic and temporal transformations in working memory. J Neurosci 2007a; 27: 13393–401.

Tan HY, Chen Q, Sust S, Buckholtz JW, Meyers JD, Egan MF, et al. Epistasis between catechol-O-methyltransferase and type II metabotropic glutamate receptor 3 genes in working memory brain function. Proc Natl Acad Sci USA 2007b; 104: 12536–41.

Teufel C, Subramaniam N, Dobler V, Perez J, Finnemann J, Mehta PR, et al. Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. Proc Natl Acad Sci USA 2015; 112: 13401–6.

Weinberger DR. Implications of normal brain development for the pathogenesis of schizophrenia. Arch Gen Psychiatry 1987; 44: 660–9.

Weinberger DR, Berman KF, Suddath R, Torrey EF. Evidence of dysfunction of a prefrontal-limbic network in schizophrenia: a magnetic resonance imaging and regional cerebral blood flow study of discordant monozygotic twins. Am J Psychiatry 1992; 149: 890–7.

Williams RJ. Simple statistical gradient-following algorithms for connectionist reinforcement learning. Mach Learn 1992; 8: 229–56.