Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Treatment resistance in first-episode psychosis is characterised by altered effective connectivity between anterior cingulate and amygdala

Anjali Bhat

a.bhat@exeter.ac.uk

University of Exeter

Charlotte Crisp

Population Health Sciences, Bristol Medical School, Universtiy of Bristol https://orcid.org/0000-0001-8174-3436

Sukhi Shergill

King's College

Rosalyn Moran

King's College London

Article

Keywords: schizophrenia, treatment resistance, dynamic causal modelling, reward learning

Posted Date: August 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10316694/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Treatment resistance in first-episode psychosis is characterised by altered effective connectivity between anterior cingulate and amygdala

Anjali Bhat[1,2], Charlotte M. Crisp, Sukhwinder S. Shergill[4,5*], Rosalyn J. Moran[1*]

Affiliations:
[1] Centre for Neuroimaging Sciences, King's College London
[2] Department of Psychology, University of Exeter
[3] School of Psychological Science, University of Bristol
[4] Kent and Medway Medical School, Canterbury Kent
[5] Department of Psychosis Sudies, Institute of Psychiatry, Psychology and Neuroscience, King's College London
[*] These senior authors contributed equally to this work

Corresponding author:

Anjali Bhat (a.bhat@exeter.ac.uk)

Abstract

Antipsychotic medications are ineffective for approximately one-third of patients with psychosis. These medications universally target dopamine D2 receptors, so it is thought that psychosis in 'treatment resistant' patients is underwritten by non-dopaminergic neural mechanisms. Aberrant glutamatergic cortical control beyond functions that rely on dopamine signalling – such as reward learning and threat perception – offers an alternative aetiology. Differential connectivity from executive prefrontal regions such as the anterior cingulate cortex (ACC) to threat and reward processing regions such as the amygdala and striatum has been observed in

treatment-resistant patients with chronic schizophrenia, but it is not clear if this is an effect of treatment and illness chronicity or a genuine characteristic of treatment response. We therefore examined differences in effective connectivity within a cortico-striatal-limbic network between 53 patients with first-episode psychosis (FEP), separated into 18 treatment-resistant and 35 treatment-responsive, with 16 healthy controls. We used Dynamic Causal Modelling of fMRI data from a reward learning task in which participants selected emotional (happy, angry) or neutral faces associated with a contingent reward. Our analyses revealed reduced (stimulus-driven) top-down connectivity between the ACC and amygdala in treatment-responsive patients compared to controls. This difference was not seen in treatment-resistant FEP patients. These findings contrast with previous findings in patients with chronic illness showing *enhanced* effective connectivity from ACC to amygdala in treatment-responsive patients compared to HCs, indicating a potential disease trajectory from FEP to chronic schizophrenia in processing of social threat and emotion. Our results also further support the relevance of glutamatergic top-down connectivity in the aetiology of treatment-resistant schizophrenia, potentially guiding development of more effective medications or treatment strategies for this population.

Keywords: *schizophrenia, treatment resistance, dynamic causal modelling, reward learning*

Introduction

Antipsychotic medications that are currently available for people with schizophrenia are ineffective for approximately one third of patients (Demjaha et al 2017, Lindenmayer 2000, Siskind et al 2022). The primary drug target of traditional antipsychotics is the dopaminergic D2 receptor (Coppens et al 1991, Sakurai et al 2013). Blocking D2 alleviates some of the 'positive' symptoms in treatment responsive patients, such as hearing voices and visual hallucinations (although primary negative symptoms – such as apathy and flat affect – and cognitive impairment associated with psychosis are not improved by dopaminergic antipsychotics (Leucht et al

2017)). Indeed, antipsychotic drugs block D2 receptors in resistant patients to an equivalent degree to responsive patients (Wolkin et al 1989), so differences in uptake are not the reason for resistance. A plausible inference is that the symptoms of 'treatment resistant' schizophrenia (TRS) patients may be associated with a different (non-dopaminergic) neurobiological pathway. If so, understanding this alternative pathway would be essential in developing new medications that are effective in treating positive symptoms in treatment-resistant patients, as well as potentially addressing more of the primary negative and cognitive symptoms of schizophrenia – which have a greater impact on quality of life (Kirkpatrick et al 2006). In a recent review, McCutcheon et al (2020) suggest that dopaminergic dysfunction in schizophrenia may be downstream of glutamatergic dysfunctions, such that aberrant dopamine signalling may be more relevant to some patients than others.

One emerging theory is that 'top-down' cortical control of subcortical brain regions, which relies on glutamate neurotransmission, distinguishes resistant from responsive patients (Demjaha et al 2014, Horne et al 2021, Egerton et al 2018, McCutcheon et al 2018). In other words, treatment resistance may be underpinned by glutamatergic mechanisms in the cortex, which are not targeted by dopaminergic drugs. Indeed, glutamate levels in the anterior cingulate cortex (ACC) are increased in treatment resistant (compared to treatment responsive) patients (Demjaha et al 2014). A prevailing computational account of glutamatergic cortical control over subcortical dopaminergic mechanisms frames the brain as a statistical model of its environment, which compares sensory information flowing 'bottom-up' (ascending a neuronal hierarchy) from sensory receptors to the cortex, against predicted sensations flowing 'top-down' from 'higher' (i.e., deeper) cortical layers. These predictions are based on prior experience and, crucially, are continuously updated to minimise 'prediction errors' (Friston 2009, Knill & Pouget 2004). In treatment resistant patients, this top-down cognitive control appears to be disrupted (Horne et al 2022, Horne et al 2021), annulling its compensatory effect against a volatile sensorium.

Reward learning is a process that is particularly dependent on dopamine signalling in the striatum, which is densely packed with dopaminergic neurons (Deserno et al 2016, FitzGerald et al 2015, Schultz 1998, Vanes et al 2018). Prediction error in this specific context – 'reward prediction error' (RPE) – is the difference

between expected and actual reward outcomes. Vanes et al (2018) showed that striatal activity due to RPE is reduced compared to healthy controls in treatment responsive patients but not treatment resistant schizophrenia patients. This difference was further examined by Horne et al (2021), who used Dynamic Causal Modelling (DCM) of fMRI data from a reward learning task in which participants selected emotional (happy, angry) or neutral face stimuli associated with a contingent monetary reward and found that treatment-responsive patients (compared to healthy controls) had increased top-down connectivity (from ACC to amygdala and fusiform), and decreased connectivity from cortical regions into the striatum. Conversely, this potentially compensatory pathways of enhanced top-down connectivity was not evidenced in treatment-resistant patients compared to controls. They concluded that this impaired top-down control from the ACC in treatment-resistant patients potentially allows misevaluation of task-relevant information, such that the caudate is updated with erroneous reward predictions.

Schizophrenia is a progressive illness, with significant differences in symptoms and neural structure between patients with first-episode psychosis (FEP) and patients with chronic schizophrenia (Bartholomeusz et al 2017, Griffa et al 2019, Rund 2009, Whitford et al 2007). For example, there are more pronounced reductions in white matter tracts in the brains of chronic patients than FEP patients (Griffa et al 2019, Whitford et al 2007), suggesting whole-brain structural connectivity tends to degenerate over the course of the illness. A recent study (Adams et al, 2022) found that pyramidal cell gain (sensitivity to action potentials) is reduced in frontal regions of patients with psychosis and that interneuron disinhibition emerges later as an allostatic compensatory mechanism to rebalance excitatory-inhibitory processing (Adams et al, 2022). The patients in the study by Horne et al (2021) had been diagnosed with schizophrenia and treated for approximately 15 years. In the current study, we aim to test whether there are similar alterations in top-down control in a cohort of first-episode psychosis (FEP) patients, who are earlier in the progression of the disease and consequently medicated for a shorter duration (approximately a year or less). Here we examine the same task employed by Horne et al (2021) in a new cohort of first-episode patients who have been followed longitudinally over 1 year to provide prognostic information on subsequent treatment-responsive or treatment-resistant status (Figure 1C). We consider both standard the Positive and Negative Syndrome Scale (PANSS) scores as indicators of disease

severity, as well as the 'psychosis configuration' principal component (PC3), which was recently proposed by Jie et al (2021) as a variable that captures relevant features of schizophrenia symptoms and maps more directly onto neural data than the standard composite scores calculated using the PANSS scale. Our earlier work suggests that the extent to which cognitive control related mechanisms are associated with symptoms differs as a function of illness chronicity (Vanes et al, 2019). The analyses in the current paper will be conducted similarly to that in Horne et al (2021) to allow comparison between their findings in chronic schizophrenia patients and the results from our first-episode psychosis cohort. We hypothesise that there will be differences in the cortico-striatal-limbic network between TRS and NTR patients in this cohort.

Methods

Participants

There were 53 first-episode medicated early psychosis patients (within one year of their first psychotic episode) and 16 healthy control (HC) subjects included in the current study, recruited from the South London and

Maudsley (SLaM) NHS foundation trust, and from the Oxleas NHS Foundation Trust and North East London NHS Foundation Trust (NELFT). Patients attended three study visits, at baseline (visit A), and at 6-month (visit B) and 1-year (visit C) follow-ups (fMRI data was collected at visits A and C, and clinical assessments were done at all three visits). Of the initial 53, 35 were subsequently classified as treatment responsive/non-treatment-resistant (NTR) and 18 Treatment resistance was defined at visit C as meeting a symptom threshold of at least one positive symptom item scored ≥5 (moderately severe) or at least two positive symptom items scored ≥4 (moderate) on the Positive and Negative Syndrome Scale (PANSS; Kay et al, 1987), alongside evidence of at least two adequate antipsychotic treatment trials (4–6 weeks at 400–600 mg chlorpromazine equivalent), adequate medication adherence, and impaired social and occupational functioning (Social and Occupational Functioning Assessment Scale (SOFAS) score <60). Treatment responsiveness was defined as adequate medication adherence and the absence of treatment-resistant symptom thresholds at both visits B and C, indicating sustained symptom stability over a six-month period. Participants who did not meet criteria for either treatment resistance or treatment responsiveness (e.g., inadequate medication adherence, insufficient treatment trials, treatment-resistant symptoms with preserved social functioning) were not classified and were excluded from subsequent analyses. Analyses in the current paper are conducted on fMRI scans from visit A (first-episode psychosis). All participants provided written informed consent prior to participation and ethical approval was obtained from the London Camberwell St Giles Research and Ethics Committee.

Procedure

Clinical assessments

PANSS and SOFAS assessments were conducted by trained clinicians, while treatment history and medication adherence were determined from clinical records and consultation with the treating clinical team. Exclusion criteria included a history of neurological illness, current major physical illness, diagnosed substance dependence within the preceding six months, or any contraindication to MRI. Treatment-resistant and

treatment-responsive groups did not differ significantly in chlorpromazine-equivalent antipsychotic dose. Any participant with a history of comorbid psychiatric or neurological disorder, substance addiction in the six months prior to participation and contraindication to MRI were excluded. Any healthy controls with a first-degree relative with a history of psychosis were also excluded.

Reward learning task

Participants undertook a well-established reward-learning task (Evans et al., 2011), in which they were asked to choose one of two faces presented side-by-side – one choice being associated with a monetary 'win' (of 10p), and the other being associated with reward omission. Over the course of each block, one face was associated with a higher contingent reward (60% likelihood of winning), and one was associated with a lower contingent reward (40% probability of winning) – effective reward learning was demonstrated by learning to choose the face associated with higher reward contingency. This task was administered in the fMRI scanner, over four testing blocks of 30 trials each. Two of the four blocks were 'emotional' blocks, in which participants chose between faces with angry and happy facial expressions. The other two testing blocks were 'neutral' blocks, in which participants chose between two neutral faces of different identities (Figure 1A). Please refer to Supplementary Methods for details of fMRI data collection. (See Horne et al (2022), who studied an overlapping cohort, for further details of the task).

First-level fMRI analysis

Statistical parametric mapping version 12 (SPM12, www.fil.ion.ucl.ac.uk/spm/software/spm12) was used for fMRI pre-processing and analysis of previously published (Horne et al 2022) fMRI data (please see Supplementary Methods for details of fMRI pre-processing). For first-level (individual) analysis, six linear model regressors modelled task conditions (face presentation, feedback from 'win' and 'loss' trials) in each block, with an additional four feedback regressors which were parametrically modulated by reward prediction

error (RPE) values. The resulting ten regressors were: 1) face onset (emotional blocks); 2) face onset (neutral blocks); 3) feedback in 'win' trials (emotional blocks); 4) feedback in 'loss' trials (emotional blocks); 5) feedback in 'win' trials (neutral blocks); 6) feedback in 'loss' trials (neutral blocks); 7) RPE x 'win' feedback (emotional blocks); 8) RPE x 'loss' feedback (emotional blocks); 9) RPE x 'win' feedback (neutral blocks); and 10) RPE x 'loss' feedback (neutral blocks). The RPE values were estimated trial by trial, as in (Vanes et al 2018), using a double-update reinforcement learning model (Schlagenhauf et al 2014) that implements a *Q*-learning algorithm in which expected outcome (*Q)* values for face stimuli were updated on every trial. Regressors were modelled using a delta function with a duration of zero and convolved using a canonical haemodynamic response function (and its temporal derivative). Six standard motion parameters as well as an individual-specific motion artifact confound matrix were added as regressors of no interest (please see Supplementary Methods for more detail on the reward learning algorithm and fMRI motion correction).

Neural network of interest

Effective connectivity (directed influences between neural regions) was assessed within a cortico-striatal-limbic network comprised of four regions of interest (ROIs) associated with task-related activity in the same cohort, as reported in Horne et al (2022). Here, we apply DCM to examine connectivity among these regions. The striatal region chosen spanned the caudate, and the 'cognitive control' region chosen was the anterior cingulate cortex (ACC). The sensory processing regions chosen were the amygdala, which responds to emotion-related sensory information (Horne et al 2022, McFadyen et al 2019, Wang et al 2017), and the fusiform gyrus, which responds to visual information – in particular, to faces (Weiner & Zilles 2016). Second-level analysis of fMRI data from the subjects included in the current study confirmed that the chosen regions of interest were activated during the task on presentation of faces (emotional and neutral blocks), at a height threshold of $F(1, 136) = 11.313$ ($p < 0.001$, uncorrected) and an extent threshold of 20 voxels (Figure 1B).

Network Time Series Extraction

The timeseries for the caudate, amygdala and fusiform were extracted from the left hemisphere only, as in Horne et al (2021). For the fusiform and amygdala, an 'All Faces' *T*-contrast was used, which contrasted volumes from emotional and neutral face presentation periods with all other volumes. For the caudate and ACC, an 'All RPE' *T*-contrast, which included all RPE-modulated feedback regressors, was used. Peak activation was identified using an automated approach (see associated MATLAB code) within an 8-12mm sphere from a set centre (selected based on the peak voxel at each region in an initial second-level analysis on HCs). The significance threshold for identifying peak activation was set deliberately low ($p<0.2$), as in Horne et al (2021), as noise is accounted for at the effective connectivity analysis stage. In 5 of 360 (72 subjects x 4 regions) instances, the algorithm did not identify activity within the 12mm sphere around the peak voxel, so peak activation was identified manually at a lower threshold of $p<0.4$. The eigenvariate was then extracted from the peak sphere for the selected contrast. The resulting timeseries were then used for dynamic causal modelling.

Dynamic causal modelling

Dynamic Causal Modelling (DCM) (Friston et al 2003) was used to examine effective connectivity within the network of interest (fusiform, caudate, amygdala, ACC). DCM is a Bayesian inferential approach which compares several possible network connectivity models that could have caused the fMRI timeseries generated while performing the task. DCM therefore allows the direction of influence between regions to be inferred (standard functional connectivity approaches do not account for directionality, as these simply assess statistical dependencies between regional activity). This is done by implementing a dynamic model to estimate changes in brain states as a function of 3 types of parameters: 1) endogenous/fixed connections between ROIs; 2) modulatory effects of task-related cues upon these connections; and 3) 'driving inputs' from the task that directly stimulate activity in specific regions. Additionally, haemodynamic parameters are also estimated with the neural model, mapping the hidden neural states underlying the BOLD signal to the observed timeseries (Zeidman et al 2019a). The 'full' neural model of all parameters to be compared was specified (Figure 1D, Table 1) such that bidirectional fixed connections were 'switched on' (expected connection strengths had a non-zero expected variance/were not fixed to a prior value) between all regions, including self-inhibitory

connections within each region. These endogenous connections were set to be modulated by task-related regressors, with face cues modulating forward connections from the fusiform and amygdala, and RPE-modulated regressors as well as feedback cues modulating connections from and between the caudate and ACC. Face presentation cues were set as 'driving inputs' from task stimuli inducing activity in the fusiform and amygdala. Reward feedback cues, as well as RPE-modulated regressors, were set as driving inputs into the caudate and ACC. Connection strengths were estimated for each subject after model inversion (performed in SPM12). A representative 'model fit' from the amygdala of one participant, showing actual timeseries data alongside timeseries data predicted using DCM, is shown in Supplementary Figure 1A.

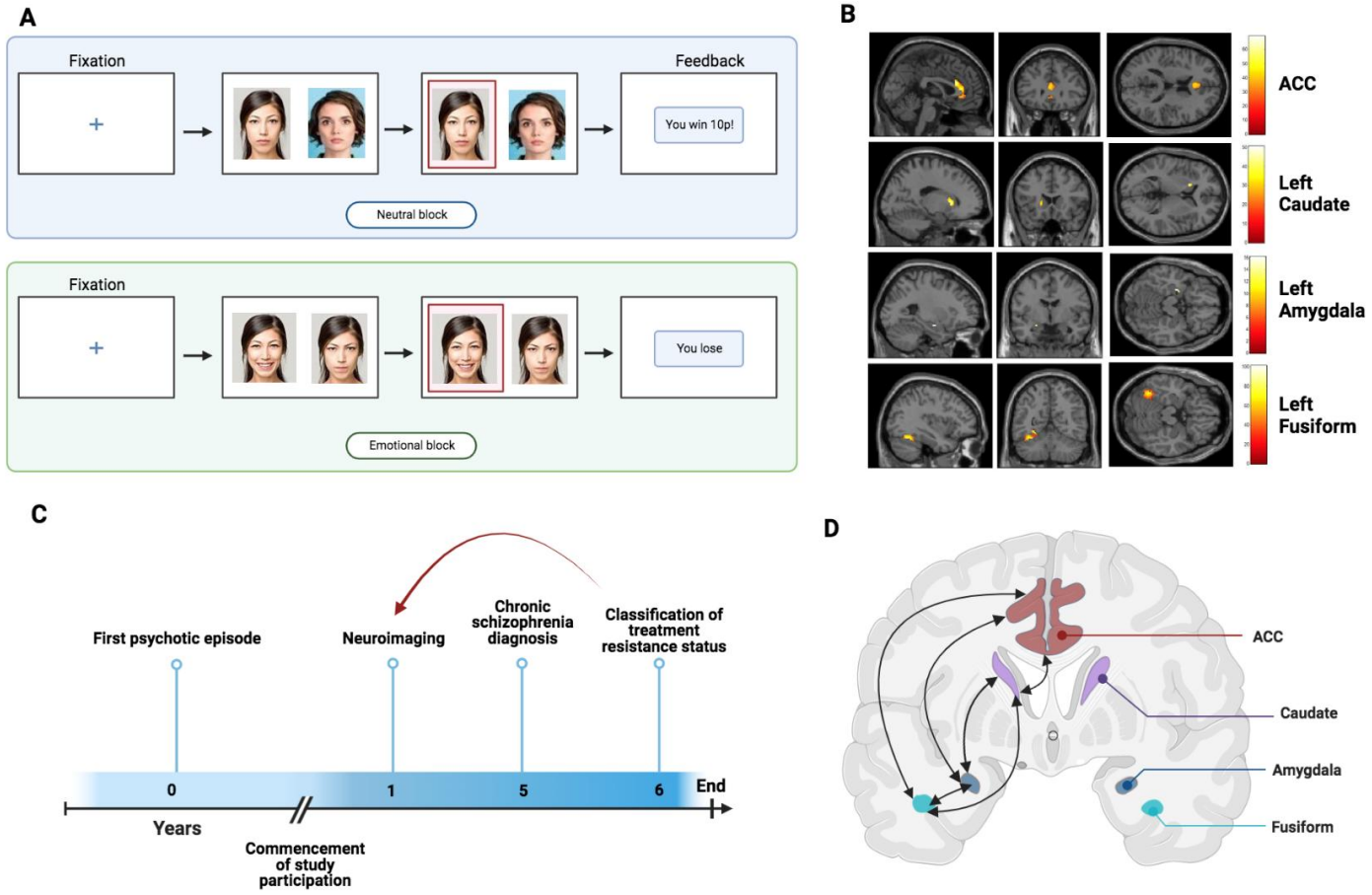


Figure 1. *Data collection and analysis.* **A. Representation of task presented in the MRI scanner.** Participants learnt which facial expression was associated with a higher probability of reward (60/40% contingencies). There were 30 trials per block, with 2 emotional blocks (in which participants chose between

happy and angry faces) and 2 neutral blocks (in which participants chose between faces of different identities with a neutral expression). **B. Significant activity in regions of interest (ROIs) at the group level, across all subjects.** Significant ROI activity for the 'All Faces' contrast for the four selected regions of interest (the ACC, left caudate, left amygdala. and left fusiform face area). Activity within the fusiform, caudate and ACC are depicted at a family-wise error corrected threshold of $p < 0.05$. Activity within the amygdala is depicted at $p < 0.001$ uncorrected. Activity in each ROI was visualised using the standard Neuromorphometrics masks in SPM12. Second level results are presented here primarily for illustrative purposes, as DCM does not rely on statistically significant results of second-level fMRI analysis (Zeidman et al, 2019). **C. Timeline of data collection.** Patients were recruited and scanned at one year following their first psychotic episode. They were classified as treatment resistant (TRS) or responsive (NTR) at 1 year. We used this information to retrospectively classify patients at the first-episode psychosis stage, allowing us to compare between groups at this early stage. **D. DCM neural model.** The full model of possible interactions underlying fMRI timeseries in the left fusiform gyrus, left caudate, left amygdala and anterior cingulate cortex (ACC) of each participant. Forward and backward connections were considered between all regions, with face cues driving activity in the amygdala and fusiform; and feedback cues and RPE-modulated regressors driving activity in the ACC and caudate. Modulation of connection strengths by task-related regressors are shown in Table 1. All self-inhibitory connections within regions were also 'switched on' to accommodate inhibitory control.

Table 1. Neural model task and modulatory inputs.

	(From) **Fusiform**	*(From)* **Caudate**	*(From)* **Amygdala**	*(From)* **ACC**
Driving inputs	Emotional + Neutral Faces (Reg* 1 & 2)	Feedback + RPEs (Reg 3-10)	Emotional + Neutral Faces (Reg 1 & 2)	Feedback + RPEs (Reg 3-10)
Modulatory inputs				
(To) Fusiform	-	Feedback + RPEs	Emotional + Neutral Faces	Feedback + RPEs
(To) Caudate	Emotional + Neutral Faces	-	Emotional + Neutral Faces	Feedback + RPEs
(To) Amygdala	Emotional + Neutral Faces	Feedback + RPEs	-	Feedback + RPEs
(To) ACC	Emotional + Neutral Faces	Feedback + RPEs	Emotional + Neutral Faces	-

**Reg* = regressor

First-level general linear model regressors: 1) Emotional block face onset; 2) Neutral block face onset; 3) Emotional block feedback cue 'win'; 4) Emotional block 'win' x reward prediction error (RPE); 5) Neutral block feedback cue 'win'; 6) Neutral block 'win' x RPE; 7) Emotional block feedback cue 'loss'; 8) Emotional block 'loss' x reward prediction error (RPE); 9) Neutral block feedback cue 'loss'; 10) Neutral block 'loss' x RPE.

Group-level effects

Group-level effects were identified using standard frequentist statistics applied to connection strengths estimated at the individual level using DCM. First – as connectivity to and from the ACC is of particular relevance to hierarchical (bottom-up/top-down) information processing –connections to and from the ACC that were significantly activated by the task were identified. This was done using a standard one-sample two-tailed *t-test* of estimated connections strengths (for each connection in the neural model) across all subjects. Subsequently, comparisons were conducted between the three groups (HC, NTR and TRS) using a one-way analysis of variance (ANOVA). All of these comparisons were implemented in MATLAB.

Downstream analyses

We subsequently assessed the influence of positive, negative, and general symptoms on network connectivity in patients by examining relationships between specific sub-components of the PANSS and estimated connection strengths, using linear regressions. We additionally conducted canonical covariance analysis to assess whether any linear combinations of PANSS scores significantly influenced connection strengths. We also explored the relevance of the principal component 'PC3', which Jie *et al* (2021) recently reported as being a potentially more neurobiologically robust predictor of psychosis than standard PANSS composite scores. We calculated PC3 for each of our subjects by taking the inner product of PC3 loadings from Jie *et al* (2021). This was regressed against estimated connection strengths to assess whether PC3 was an effective predictor of connectivity in our sample. We finally conducted Pearson correlations between PANSS sub-components and PC3 scores within our sample to examine whether PC3 was related more strongly to any of the three main subcategories of the PANSS (positive, negative and general symptoms).

Results

Demographics

All three groups were matched for age and sex, and the two patient groups were matched for age of illness onset and illness duration (Table 1). There were significantly more smokers in the TRS group (n = 18) than the NTR (n = 35) or HC (n = 16) groups.

Table 2. Means (M) and standard deviations (SD) of clinical and demographic variables by group

	HC		**NTR**		**TRS**		**Group Statistics**		
							χ^2	df	p
Female (%)	50		44		22		2.88	2	0.24
Smokers %	7		50		72		14.53	2	< 0.001
	M	SD	M	SD	M	SD	F	df	p
Age	27.25	3.30	25.88	5.29	24.67	4.60	1.233	2	0.298
							t	df	p

Onset age (years)	25.56	5.44	23.24	4.27	0.92	37	0.36
Illness duration (years)	1.6	1.4	1.5	1.0	-0.33	41	0.75
PANSS score							
Positive symptoms	11.97	4.46	14.17	4.01	-1.55	36	0.13
Negative symptoms	11.19	4.35	13.61	4.61	-1.58	31	0.12
General symptoms	28.22	7.28	30.89	7.15	-1.07	33	0.29
Total score	51.38	14.28	58.67	13.47	-1.55	34	0.13

Abbreviations: HC, heathy controls; NTR, non-treatment resistant; TRS, treatment resistant schizophrenia; PANSS, Positive and Negative Symptom Scale

Network connectivity

Dominant connections

The dominant connections within the network (Figure 2A) to or from the ACC were from the fusiform to the ACC (M = 0.3426 Hz; SD = 0.9383; $t(68)$ = 3.0327; FDR = 0.0034) and from the ACC to the amygdala (M = 0.3426 Hz; SD = 1.3103; $t(68)$ = 3.1921; FDR = 0.0021), after false discovery rate (FDR) correction (Benjamini & Yekutieli, 2001) for multiple tests (using a FDR of 0.05). The means here reflect the estimated connection strengths from the individual DCMs across all subjects: these 'dominant connections' were those significantly activated by task inputs.

Between-group comparisons

Of the two dominant connections, a one-way ANOVA comparing mean estimated connection strengths between the diagnostic groups – HCs (N = 16), NTRs (N = 35) and TRS (N = 18) – revealed a statistically significant difference between groups for the ACC to amygdala connection ($F(2, 66)$ = 4.06, p = 0.0218). Tukey's HSD Test for multiple comparisons found that the mean estimated connection strength of the ACC to amygdala

connection was significantly different between NTR patients and HCs (p = 0.0192, 95% *C.I.* = 0.1447, 1.9610), with the NTR patients showing significantly *reduced* top-down influence (M = 0.2289 Hz) between ACC and amygdala compared to HCs (M = 1.2817 Hz). No significant difference was observed between TRS patients (M = 0.3459 Hz) and HCs, or between TRS and NTR patients (Figure 2B). No significant difference was seen between groups for the fusiform to ACC connection ($F(2, 66) = 0.93$, p = 0.3984).

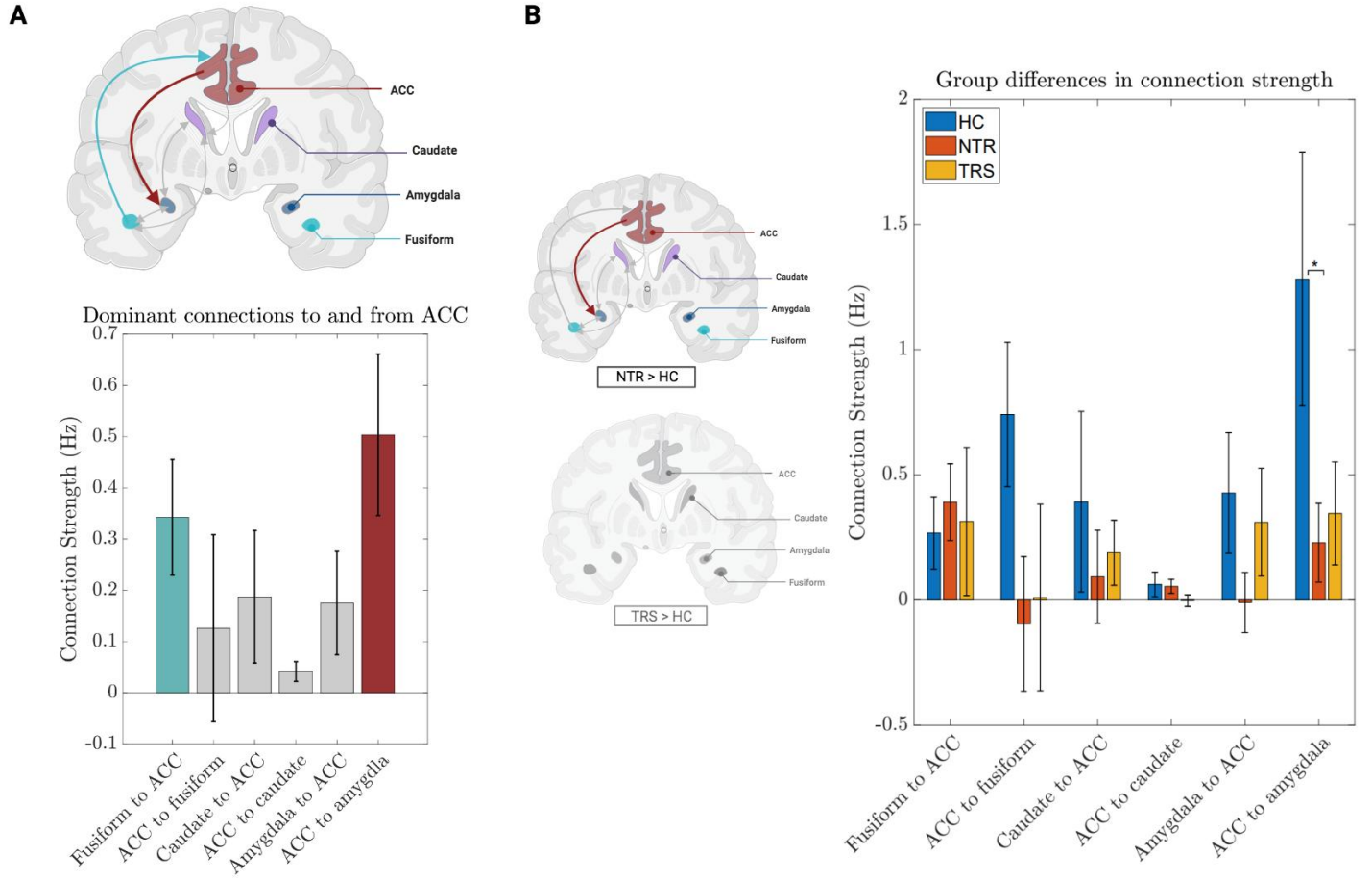


Figure 2. *Group-level effects in connectivity.* **A. Dominant connections in network model.** ***Upper panel:*** The two connections that were significantly activated by task inputs are highlighted within the neural network of interest: the Fusiform to ACC connection and the ACC to Amygdala connection. ***Lower panel:*** Average estimated connection strength (in Hz) across all subjects for each connection to and from the ACC. **B. Between-group comparisons of connection strengths.** ***Left panel:*** No significant connections were observed when comparing TRS to HC groups. ***Right panel:*** Group differences by connection strengths for each connection to and from the ACC within the neural network of interest. A significantly reduced influence from ACC to amygdala is seen in NTR patients (in red) compared to HCs (in blue). No difference is seen between TRS patients (in yellow) and either of the other groups.

Downstream analyses

PANSS subscales and network connectivity

Linear regressions examining the influence of positive, negative, and general symptoms on estimated connection strengths in patients revealed no significant relationships (Supplementary Figure 2B-D). Canonical covariance analysis assessing whether any linear combinations of PANSS scores significantly influenced connection strengths also showed no significant relationships (Supplementary Figure 3).

Principal component PC3, network connectivity and PANSS scores

Principal component 'PC3' values (Jie *et al*, 2021) calculated for the patients in our sample was regressed against estimated connection strengths. These regressions showed no significant relationships between PC3 and connection strengths (Supplementary Figure 2A). Pearson correlations between PANSS sub-components and PC3 scores (Figure 3) interestingly showed that PC3 was significantly related to negative (r = -0.64709 , p < 0.05) symptoms, but not to positive (r = 0.2476, p < 0.05) or general (r = -.022977, p < 0.05) symptoms.

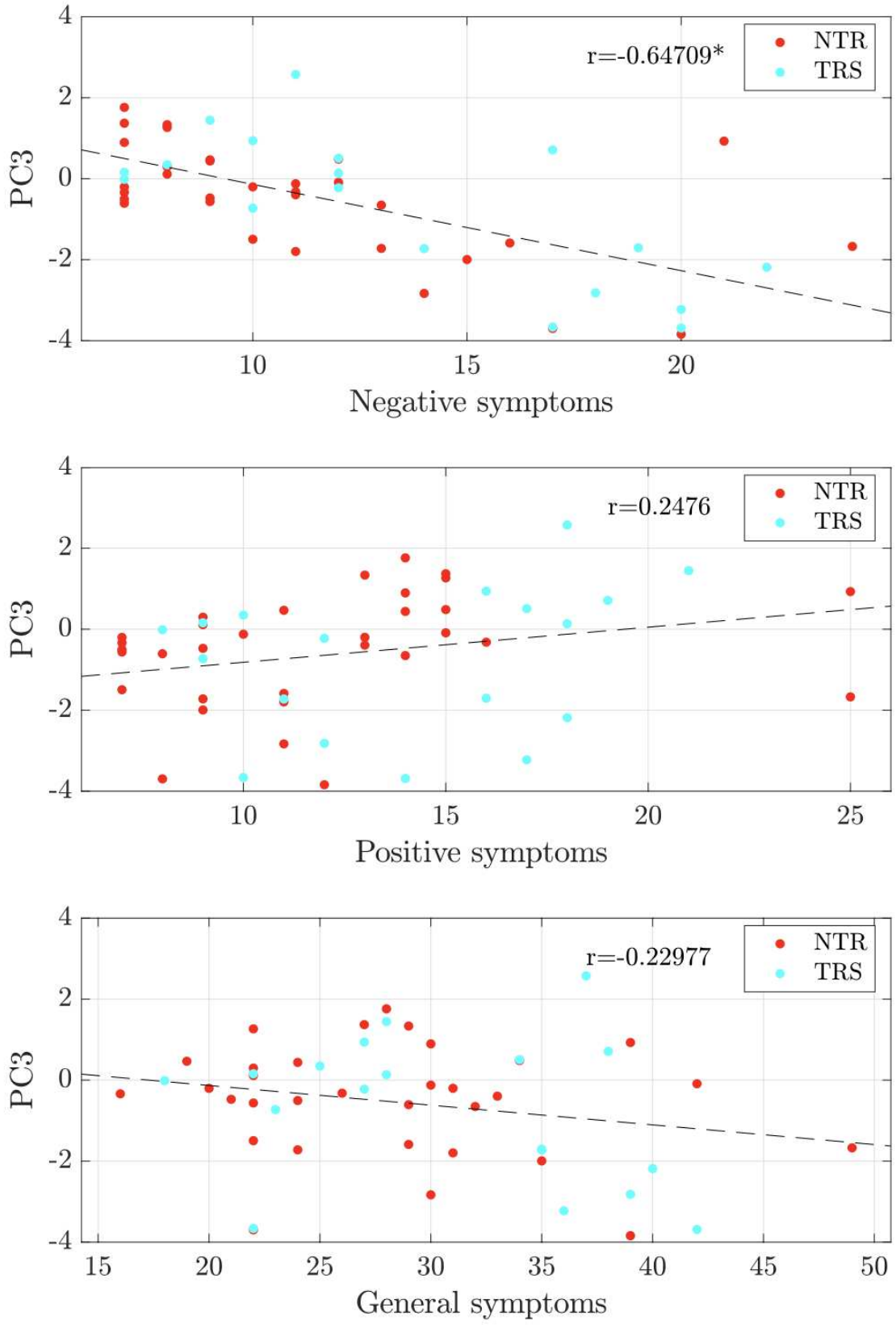


Figure 3. *Relationships between principal component 3* (PC3) *and PANSS subscales.* PC3 relates significantly to negative symptoms (top panel), but not to general and positive symptoms.

Discussion

In this study, we used Dynamic Causal Modelling (DCM) to assess differential effective connectivity in a cortico-striatal-limbic network between healthy controls (HCs) and patients with first-episode psychosis (FEP) (visit A) who were either treatment-resistant (TRS) or non-treatment resistant (NTR). Activity in this network was elicited by presenting participants with a reward-learning task in the scanner, which involved choosing between happy and angry faces associated with a contingent monetary reward. This study builds on findings from Horne *et al* (2021), who found in a chronic schizophrenia cohort that NTR patients have enhanced top-down effective connectivity from the anterior cingulate cortex (ACC) to the amygdala and the fusiform, while there was no difference in top-down connectivity between TRS patients and healthy controls. Horne and colleagues interpreted this to mean that enhanced reliance on top-down control is a compensatory mechanism available to NTR patients but not TRS patients – that this may be an explanation for the differences in their respective responses to dopaminergic antipsychotics. We used the same task and a similar network of interest in our current study, with the aim of exploring whether this differential connectivity replicates in patients at an earlier stage of the illness (FEP patients being approximately one year from their first psychotic episode, while chronic patients are at least five years beyond their first episode). Our analyses indeed similarly revealed a significant difference in top-down connectivity from ACC to amygdala in FEP NTR patients compared to HCs, but no significant difference in this connection between FEP TRS patients and HCs. No other connections within the network showed significantly different average estimated connection strengths between groups. ACC-amygdala connections have been shown to constitute an emotional processing system that underwrites innate fear responses, with the amygdala responding rapidly to threatening stimuli such as angry faces (McFadyen et al, 2019), and the ACC regulating emotional responses on extinction of threatening stimuli (Jinho et al, 2018; Etikin et al, 2011). This result, taken together with the findings of Horne *et al* (2021), suggests a significant role of this top-down regulation of emotion – underwritten by ACC to amygdala connectivity – in treatment-resistant schizophrenia.

In keeping with this interpretation, in a subsequent paper, Horne et al (2022) suggested that patients with early psychosis – especially those with more severe positive symptoms – overweight the prior bias for happy

faces and underweight the bottom-up evidence received in each trial as feedback. In other words, patients with severe positive symptoms are less able to inhibit their emotional bias in order to strategically obtain a reward. Schlagenhauf et al (2014) tell a similar story in unmedicated patients: they used computational modelling to assess the ability of patients with schizophrenia to use latent task structure (e.g., reward contingency) and found impaired reversal learning as well as decreased sensitivity to reward in patients compared to controls. It has been shown previously that cognitive control (as measured by the Stroop task) was significantly more impaired in TRS than NTR patients (Thomas et al, 2021).

However, in our study, NTR patients exhibit *reduced* top-down connectivity between the ACC and amygdala compared to healthy controls – as opposed to the *enhanced* effect in NTRs seen in the previous study. This may indicate a potential progressive trajectory of ACC-amygdala connectivity that is specific to patients whose symptoms do improve with dopaminergic antipsychotics. At this earlier stage of illness (FEP), NTR patients appear to engage *less* than HCs in top-down control of – potentially hypervigilant – processing of emotional inferences in the amygdala. However, as NTR patients progress towards a later stage, this top-down influence appears to develop such that they rely on it more than HCs at five or more years into the illness. If this is the case, it may be that dopaminergic antipsychotics indirectly influence top-down glutamatergic connections (for example, by creating conditions that enable the development of emotional regulation that mitigates hypervigilance to social threat). Indeed, there is evidence that the dopaminergic and glutamatergic systems influence each other bidirectionally in humans – for example, Jauhar et al (2018) show an inverse relationship between glutamate concentration in the ACC and striatal dopamine synthesis capacity in FEP patients. Howes et al (2015) propose that cortical glutamatergic dysfunction in patients with schizophrenia directly causes negative and cognitive symptoms and indirectly causes positive symptoms by disinhibiting dopamine neurons via projections to the midbrain. This would be an interesting explanation for why the differences in top-down control are seen in NTR patients but not TRS patients, who perhaps do not develop this top-down ACC to amygdala control facilitated by antipsychotic medications to the same degree. Inhibitory control of emotional responses to threatening facial expressions appears to be modulated by appetitive motivation (drive to obtain reward) (Carver, 2004; Gray, 1994; Passamonti et al, 2008). Loss of motivation and failure to update expected

value is a prevalent negative symptom of schizophrenia which worsens with chronic illness (Waltz & Gold, 2016), so this may also influence the trajectory of top-down control of emotional bias from early to later stages of illness.

In our downstream analyses, we did not find significant relationships between PANSS sub-components (positive, negative or general symptom composite scores, or hallucination or delusion scores) and connectivity in our group. This may further indicate that the group differences in connectivity we observed in the current study are a result of exposure and individualised responses to medication, as opposed to the mechanistic underpinnings of symptoms themselves. We also did not see significant relationships between the 'psychosis configuration' principal component (PC3) (proposed by Jie et al (2021) as a more neurobiologically robust alternative than the standard PANSS composite scores) and connectivity in patients in the current study. Interestingly, while the intuition Jie et al (2021) offer is that the more *positive* the PC3 score, the greater the presentation of delusional, grandiose or hallucinatory symptoms, and the more *negative* the PC3 score, the greater the presentation of negative symptoms such as social avoidance and flat affect, we find that PC3 for our patient cohort correlates only with *negative* symptoms. This may suggest that PC3 may not be as specific in different patient samples.

There are some limitations of the current study that are important to note. Grouping patients into TRS and NTR categories is challenging and as yet unstandardised, with considerable heterogeneity between treatment resistant patients (Horne et al, 2022; Potkin, 2020). We did not see a significant difference in ACC to amygdala connectivity when we *directly* compared the two patient groups – which suggests that the differences we observe are subtle – perhaps more a matter of degrees rather than clear-cut group difference – and likely influenced by the aforementioned heterogeneity in categorisation of patients into TRS and NTR groups. In future studies, longer term follow-ups would be necessary to confirm our interpretations of the difference between the results of the current study and those from Horne et al (2021). Exposure to antipsychotic medication can impact brain structure and function (Abbott et al., 2013, Smieskova et al., 2009), so it would be particularly important to assess ACC-amygdala connectivity in relation to medication status, or alterations in

this connection as a result of experimental exposure to dopaminergic antipsychotics. This would ideally include unmedicated patients.

Overall, our results support previous observations of altered top-down cognitive control of emotional responses to threatening stimuli in patients with schizophrenia, which appears to be exaggerated in patients who *do* respond to dopaminergic antipsychotics. Taken together with recent literature (Horne et al, 2021; Horne et al, 2022; Vanes et al, 2018; Schlagenauf et al 2014), these findings indicate a trajectory from reduced top-down control by the ACC compensating for a hypervigilant response of the amygdala in NTR FEP patients, to increased top-down connectivity in NTR chronic schizophrenia patients, potentially indicating that prolonged use of dopaminergic antipsychotics indirectly enhances emotional regulation. We also see that this trajectory is not present in TRS patients, suggesting that there may be a different neurobiological mechanism more relevant than top-down emotional regulation to symptoms in this population, which future studies could potentially elucidate. Our results also further support the potential relevance of glutamatergic top-down connectivity in the aetiology of treatment resistant schizophrenia, potentially guiding development of more effective medications for this population (Fleischhacker et al 2023).

Declaration of interest: The authors declare no conflict of interest.

Data availability: All code written for this study is available at https://github.com/abhat92/Schizophrenia-Treatment-Resistance-DCM.git. Imaging data is stored on NeuroVault.

Acknowledgements: We are grateful for funding from the UKRI grant ES/T01279X/1 supporting the work of AB and RM.

Supplementary Information

Supplementary Methods

Scanning parameters

Acquisition of functional scans was conducted on a 3 Tesla GE Excite-II Magnetic Resonance scanner (GE Healthcare, USA). A T2* echo planar sequence that was sensitive to a blood oxygenation level-dependent (BOLD) contrast was used (430 volumes, TR=2000 ms, TE=35 ms, field of view = 24 cm, slice thickness = 3 mm, matrix = 64 × 64, flip angle = 75°). Each participant's structural image was acquired with a T1-weighted magnetization prepared rapid acquisition gradient echo (MP RAGE) sequence (TR=7321 ms, TE=3 ms, TI = 400 ms, field of view = 240, slice thickness = 1.2 mm, 196 slices).

fMRI head motion quality control

Volumes corrupted by motion artifacts were identified using DVARS (Power et al 2012), implemented in FSL Motion Outliers, which detects 'spikes' in head movement for each participant. A resulting motion artifact confound matrix was added to the first-level fMRI analysis model (alongside the 6 standard motion parameters) as a 'regressor of no interest'.

Pre-processing of fMRI data

Structural and functional images were skull-stripped and manually reoriented to set the origin over the anterior commissure. Functional images were then realigned to correct for the effects of head motion, co-registered to the structural images and normalised to Montreal Neurological Institute (MNI) space. A high pass filter of 100 seconds was applied. Finally, the data were spatially smoothed using a 6mm full-width half maximum (FWHM) Gaussian kernel.

Reinforcement learning model

The model used to estimate reward learning was a *Q*-learning model, which essentially assesses the quality of an action (how effectively the action effects a transition to a desired state). In standard Rescorla-Wagner *Q*-learning, only the chosen option is updated by a prediction error. The current study uses double-update *Q*-learning (Schlagenhauf et al 2014), in which expected reward (Q) values for both the chosen and unchosen face stimuli were updated on every trial. If a reward is obtained for the chosen face, the *Q*-value for this choice is increased, while the *Q*-value of the unchosen face is reduced. This model represents the asymmetry of the two choices introduced by the task: if face 1 is the better choice (rewarded with 60% contingency), then face 2 is the worse choice (rewarded with 40% contingency). Participants' choices were considered ideal if their expected reward in a given trial for the face they chose, $Q_1(t)$, was greater than their expected reward for the face they did not choose, $Q_2(t)$. The aim here was to quantify how well the subjects translated their value representation into action. The probability, $P_1(t)$, of choosing the first option on a given trial t was estimated using a softmax function (Equation 1).

$$P_1(t) = \frac{exp\ exp\left(\beta \times Q_1(t)\right)}{exp\ exp\left(\beta \times Q_1(t)\right) + exp\ exp\left(\beta \times Q_2(t)\right)}$$

β = inverse temperature (choice randomness)

$Q_1(t)$ = expected reward for choice 1 on trial *t*

(1

Reward prediction error (Equation 2) was defined as the difference between the expected reward (Q) and obtained reward (R) on trial t.

$$RPE = R(t) - Q_1(t)$$

(2

Expected reward for the chosen face was updated in each trial by adding the RPE scaled by a learning rate parameter α (Equation 3).

$$Q_1(t+1) = Q_1(t) + \alpha(RPE) \quad \textbf{(3}$$

Similarly, the expected reward for the unchosen face was updated on each trial using the same update equation with the inverse reward value (Equation 4).

$$Q_2(t+1) = Q_2(t) + \alpha(\underline{R}(t) - Q_2(t)) \quad \textbf{(4}$$

The learning rate parameter α and the inverse temperature parameter β were estimated for each diagnostic group (TRS, NTR and HC) by minimizing the negative log likelihood of the observed data across all participants within the group.

References

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

Adams RA, Vincent P, Benrimoh D, Friston KJ, Parr T. 2021. Everything is connected: Inference and attractors in delusions. *Schizophr Res*

Adams, R, Konstantinos Tsirlis, Leonhardt Unruh, Aashna Mahajan, Ana Montero Horas, Laura Convertino et al. (2022). Computational modeling of electroencephalography and functional magnetic resonance imaging paradigms indicates a consistent loss of pyramidal cell synaptic gain in schizophrenia. *Biol Psych*, 91, 202-215.

Bartholomeusz CF, Cropley VL, Wannan C, Di Biase M, McGorry PD, Pantelis C. 2017. Structural neuroimaging across early-stage psychosis: Aberrations in neurobiological trajectories and implications for the staging model. *Aust N Z J Psychiatry* 51: 455-76

Benrimoh D, Parr T, Adams RA, Friston K. 2019. Hallucinations both in and out of context: An active inference account. *PLoS One* 14: e0212379

Bhat A, Irizar H, Thygesen JH, Kuchenbaecker K, Pain O, et al. 2021. Transcriptome-wide association study reveals two genes that influence mismatch negativity. *Cell Rep* 34: 108868

Coppens HJ, Slooff CJ, Paans AM, Wiegman T, Vaalburg W, Korf J. 1991. High central D2-dopamine receptor occupancy as assessed with positron emission tomography in medicated but therapy-resistant schizophrenic patients. *Biol Psychiatry* 29: 629-34

Demjaha A, Egerton A, Murray RM, Kapur S, Howes OD, et al. 2014. Antipsychotic treatment resistance in schizophrenia associated with elevated glutamate levels but normal dopamine function. *Biol Psychiatry* 75: e11-3

Demjaha A, Lappin JM, Stahl D, Patel MX, MacCabe JH, et al. 2017. Antipsychotic treatment resistance in first-episode psychosis: prevalence, subtypes and predictors. *Psychol Med* 47: 1981-89

Deserno L, Schlagenhauf F, Heinz A. 2016. Striatal dopamine, reward, and decision making in schizophrenia. *Dialogues Clin Neurosci* 18: 77-89

Egerton A, Broberg BV, Van Haren N et al. 2018. Response to initial antipsychotic treatment in first episode psychosis is related to anterior cingulate glutamate levels: a multicentre 1H-MRS study (OPTiMiSE). *Mol Psychiatry*. 23:2145-55.

FitzGerald TH, Dolan RJ, Friston K. 2015. Dopamine, reward learning, and active inference. *Front Comput Neurosci* 9: 136

Flagel SB, Clark JJ, Robinson TE, Mayo L, Czuj A, et al. 2011. A selective role for dopamine in stimulus-reward learning. *Nature* 469: 53-7

Friston K. 2009. The free-energy principle: a rough guide to the brain? *Trends Cogn Sci* 13: 293-301

Friston K, Brown HR, Siemerkus J, Stephan KE. 2016. The dysconnection hypothesis (2016). *Schizophr Res* 176: 83-94

Friston KJ, Harrison L, Penny W. 2003. Dynamic causal modelling. *Neuroimage* 19: 1273-302

Griffa A, Baumann PS, Klauser P, Mullier E, Cleusix M, et al. 2019. Brain connectivity alterations in early psychosis: from clinical to neuroimaging staging. *Transl Psychiatry* 9: 62

Horne CM, Sahni A, Pang SW, Vanes LD, Szentgyorgyi T, et al. 2022. The role of cognitive control in the positive symptoms of psychosis. *Neuroimage Clin* 34: 103004

Horne CM, Vanes LD, Verneuil T, Mouchlianitis E, Szentgyorgyi T, et al. 2021. Cognitive control network connectivity differentially disrupted in treatment resistant schizophrenia. *Neuroimage Clin* 30: 102631

Howes OD, Egerton A, Allan V, McGuire P, Stokes P, Kapur S. 2009. Mechanisms underlying psychosis and antipsychotic treatment response in schizophrenia: insights from PET and SPECT imaging. *Curr Pharm Des* 15: 2550-9

Howes OD, Kapur S. 2009. The dopamine hypothesis of schizophrenia: version III--the final common pathway. *Schizophr Bull* 35: 549-62

Howes O, McCutcheon R, Stone J. 2015. Glutamate and dopamine in schizophrenia: an update for the 21st century. *J Psychopharmacol (Oxf)*;29:97–115

Ji JL, Helmer M, Fonteneau C, Burt JB, Tamayo Z, Demšar J, Adkinson BD, Savić A, Preller KH, Moujaes F, Vollenweider FX, Martin WJ, Repovš G, Cho YT, Pittenger C, Murray JD, Anticevic 2021. A. Mapping brain-behavior space relationships along the psychosis spectrum. *Elife*. 10:e66968. doi: 10.7554/eLife.66968.

Jauhar S, McCutcheon R, Borgan F, Veronese M, Nour M, Pepper F, Rogdaki M, Stone J, Egerton A, Turkheimer F, McGuire P, Howes OD. 2018. The relationship between cortical glutamate and striatal dopamine in first-episode psychosis: a cross-sectional multimodal PET and magnetic resonance spectroscopy imaging study. *Lancet Psychiatry.* 5(10):816-823. doi: 10.1016/S2215-0366(18)30268-2. E

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

Kay SR, Fiszbein A, Opler LA. 1987. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull* 13: 261-76

Keiflin R, Janak PH. 2015. Dopamine Prediction Errors in Reward Learning and Addiction: From Theory to Neural Circuitry. *Neuron* 88: 247-63

Kirkpatrick B, Fenton WS, Carpenter WT, Jr., Marder SR. 2006. The NIMH-MATRICS consensus statement on negative symptoms. *Schizophr Bull* 32: 214-9

Knill DC, Pouget A. 2004. The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci* 27: 712-9

Lavin C, Melis C, Mikulan E, Gelormini C, Huepe D, Ibanez A. 2013. The anterior cingulate cortex: an integrative hub for human socially-driven interactions. *Front Neurosci* 7: 64

Leucht S, Leucht C, Huhn M, Chaimani A, Mavridis D, et al. 2017. Sixty Years of Placebo-Controlled Antipsychotic Drug Trials in Acute Schizophrenia: Systematic Review, Bayesian Meta-Analysis, and Meta-Regression of Efficacy Predictors. *Am J Psychiatry* 174: 927-42

Lindenmayer JP. 2000. Treatment refractory schizophrenia. *Psychiatr Q* 71: 373-84

McCutcheon RA, Krystal JH, Howes OD. 2020. Dopamine and glutamate in schizophrenia: biology, symptoms and treatment. *World Psychiatry*. 19(1):15-33. doi: 10.1002/wps.20693.

McFadyen J, Mattingley JB, Garrido MI. 2019. An afferent white matter pathway from the pulvinar to the amygdala facilitates fear recognition. *Elife* 8

Parr T, Friston KJ. 2019. Attention or salience? *Curr Opin Psychol* 29: 1-5

Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE. 2012. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage* 59: 2142-54

Rund BR. 2009. Is there a degenerative process going on in the brain of people with Schizophrenia? *Front Hum Neurosci* 3: 36

Sakurai H, Bies RR, Stroup ST, Keefe RS, Rajji TK, et al. 2013. Dopamine D2 receptor occupancy and cognition in schizophrenia: analysis of the CATIE data. *Schizophr Bull* 39: 564-74

Schultz W. 1998. Predictive reward signal of dopamine neurons. *J Neurophysiol* 80: 1-27

Siskind D, Orr S, Sinha S, Yu O, Brijball B, et al. 2022. Rates of treatment-resistant schizophrenia from first-episode cohorts: systematic review and meta-analysis. *Br J Psychiatry* 220: 115-20

Stevens FL, Hurley RA, Taber KH. 2011. Anterior cingulate cortex: unique role in cognition and emotion. *J Neuropsychiatry Clin Neurosci* 23: 121-5

Trubetskoy V, Pardiñas AF, Qi T, Panagiotaropoulou G, Awasthi S, Bigdeli TB, Bryois J, Chen CY, Dennison CA, Hall LS, Lam M, Watanabe K, Frei O, Ge T, Harwood JC, ... Schizophrenia Working Group of the Psychiatric Genomics Consortium. 2022. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature*. 604(7906):502-508. doi: 10.1038/s41586-022-04434-5.

Vanes LD, Mouchlianitis E, Collier T, Averbeck BB, Shergill SS. 2018. Differential neural reward mechanisms in treatment-responsive and treatment-resistant schizophrenia. *Psychol Med* 48: 2418-27

Wang S, Yu R, Tyszka JM, Zhen S, Kovach C, et al. 2017. The human amygdala parametrically encodes the intensity of specific facial emotions and their categorical ambiguity. *Nat Commun* 8: 14821

Weiner KS, Zilles K. 2016. The anatomical and functional specialization of the fusiform gyrus. *Neuropsychologia* 83: 48-62

Whitford TJ, Grieve SM, Farrow TF, Gomes L, Brennan J, et al. 2007. Volumetric white matter abnormalities in first-episode schizophrenia: a longitudinal, tensor-based morphometry study. *Am J Psychiatry* 164: 1082-9

Wolkin A, Barouche F, Wolf AP, Rotrosen J, Fowler JS, et al. 1989. Dopamine blockade and clinical response: evidence for two biological subgroups of schizophrenia. *Am J Psychiatry* 146: 905-8

Zeidman P, Jafarian A, Corbin N, Seghier ML, Razi A, et al. 2019a. A guide to group effective connectivity analysis, part 1: First level analysis with DCM for fMRI. *Neuroimage* 200: 174-90

Zeidman P, Jafarian A, Seghier ML, Litvak V, Cagnan H, et al. 2019b. A guide to group effective connectivity analysis, part 2: Second level analysis with PEB. *Neuroimage* 200: 12-25