OXFORD UNIVERSITY PRESS | Brain

# Acquisition of visual priors and induced hallucinations in chronic schizophrenia

| | |
|---|---|
| Journal: | *Brain* |
| Manuscript ID | BRAIN-2018-02223.R1 |
| Manuscript Type: | Original Article |
| Date Submitted by the Author: | n/a |
| Complete List of Authors: | Valton, Vincent; University College London Institute of Cognitive Neuroscience; University of Edinburgh, Informatics<br>Karvelis, Povilas; University of Edinburgh, Informatics<br>Richards, Katie; Univ of Edinburgh, Department of Psychiatry<br>Seitz, Aaron; Riverside University,<br>Lawrie, Stephen; Univ of Edinburgh, Department of Psychiatry<br>Series, Peggy; University of Edinburgh, Informatics |
| Subject category: | Neuropsychiatry |
| To search keyword list, use whole or part words followed by an *: | Schizophrenia < NEUROPSYCHIATRY, Computational psychiatry < NEUROPSYCHIATRY, Psychosis < NEUROPSYCHIATRY, Vision < SYSTEMS/DEVELOPMENT/PHYSIOLOGY, Positive symptoms < NEUROPSYCHIATRY, Visual hallucinations < NEUROPSYCHIATRY, Sensory systems < SYSTEMS/DEVELOPMENT/PHYSIOLOGY |
| | |

# Acquisition of visual priors and induced hallucinations in chronic schizophrenia

**Vincent Valton[1,2,5,¶], Povilas Karvelis[1,¶], Katie L. Richards[2], Aaron R. Seitz[3], Stephen M. Lawrie[2,4], Peggy Seriès[1,*]**

*[1] Institute for Adaptive and Neural Computation, University of Edinburgh, UK*

*[2] Department of Psychiatry, Royal Edinburgh Hospital, University of Edinburgh, UK*

*[3] Department of Psychology, University of California Riverside, CA, USA*

*[4]Patrick Wild Centre, University of Edinburgh, Edinburgh, UK*

*[5]Institute of Cognitive Neuroscience, University College London, London, UK*

**Correspondence should be addressed to: pseries@inf.ed.ac.uk*

¶ These authors contributed equally to this work

*Running Title: "Learning and Bayesian inference in schizophrenia"*

## Abstract

Prominent theories suggest that symptoms of schizophrenia stem from learning deficiencies resulting in distorted internal models of the world. To further test these theories, we here use a visual statistical learning task known to induce rapid implicit learning of the stimulus statistics (Chalk *et al.,* 2010). In this task, participants are presented with a field of coherently moving dots and need to report the presented direction of the dots (estimation task) and whether they saw any dots or not (detection task). Two of the directions were more frequently presented than the others. In controls, the implicit acquisition of the stimuli statistics influences their perception in two ways: 1- motion directions are perceived as being more similar to the most frequently presented directions than they really are (estimation biases); 2- in the absence of stimuli, participants sometimes report perceiving the most frequently presented directions (a form of hallucinations). Such behaviour is consistent with probabilistic inference, i.e. combining learnt perceptual priors with sensory evidence. We investigated whether patients with chronic, stable, treated schizophrenia (n=20) differ from controls (n=23) in the acquisition of the perceptual priors and/or their influence on perception. We found that, although patients were slower than controls, they showed comparable acquisition of perceptual priors, correctly approximating the stimulus statistics. This suggests that patients have no statistical learning deficits in our task. This may reflect our patients' relative wellbeing on antipsychotic medication. Intriguingly, however, patients experienced significantly fewer hallucinations of the most frequently presented directions than controls when the stimulus was absent or when it was very weak (prior-based lapse estimations). This suggests that prior expectations had less influence on patients' perception than on controls when stimuli were absent or below perceptual threshold.


Keywords: Schizophrenia, Inference, Statistical Learning, Hallucinations

## Introduction

An increasingly popular idea in neuroscience is that perception and decision-making can be well described in terms of probabilistic inference processes (Knill and Pouget, 2004; Fiser *et al.*, 2010; Friston, 2010; 2012). For example, statistical and perceptual learning studies show that the perceptual systems continuously extract and learn statistical regularities of the environment (for a review in visual perception, see e.g. Seriès and Seitz, 2013). This learning results in the construction of internal models of the environment, or expectations, which are used automatically and unconsciously to predict and disambiguate perceptual inputs in situations of uncertainty and to guide decisions.

In this context, it has been proposed that psychiatric disorders in general, and schizophrenia in particular, might be explained in terms of deficits in probabilistic inference (Friston, 2005; Frith and Friston, 2012; Adams *et al.,* 2013; Jardri and Denève, 2013, for reviews see: Friston *et al.* 2016; Valton *et al.,* 2017; Sterzer et al 2018). Impaired statistical learning and/or inference deficits would lead to distorted internal models of the world, which could then explain the existence of abnormal beliefs or delusions experienced by patients with schizophrenia, as well as their hallucinations.

Two different lines of research support this general idea. First, a number of studies using explicit probabilistic learning tasks, such as the "beads task" report that patients with schizophrenia show a deficit in integrating probabilistic information resulting in faster responses than control subjects, an effect called the 'jumping-to-conclusions' bias (Huq *et al*., 1988; Speechley *et al*., 2010; Averbeck *et al*., 2011; Evans *et al*., 2012). Interestingly, patients with stronger delusional symptoms fare worse at the task than those who do not (Huq *et al*., 1988; Speechley *et al*., 2010), and control subjects displaying delusional ideation also show similar impairments at the task (Freeman *et al*., 2008), suggesting a link between delusions and probabilistic inference (Garety *et al.*, 2013; Garety and Freeman 2013). Second, patients with schizophrenia do not experience visual illusions in the same way as controls do. For example, patients are less susceptible to certain visual illusions such as the hollow-mask illusion (Dima *et al*., 2009; 2010; Keane *et al*., 2013; for review see: Silverstein and Keane 2011a; 2011b; Notredame *et al.*, 2014). This suggests that they either have different implicit expectations about the environment (i.e. they would not have such a strong expectation that faces are convex) or

that these expectations do not affect patients' perception in the same way as observed in controls.

A few studies have recently tried to test the impaired Bayesian inference hypothesis more directly, but the findings are mixed. Teufel et al. (2015) found that early psychosis and schizotypal traits were associated with an increased influence of prior knowledge when disambiguating two-tone images. Powers et al. (2017) also reported an increased perceptual prior influence in experimentally-induced hallucinations in both patients and controls with higher propensity to hallucinatory experiences. Interestingly, however, a series of studies by Schmack et al. (2013; 2015; 2017) found an increased influence of cognitive priors on the perception of bi-stable stimuli in participants with schizotypal traits and clinical schizophrenia, but showed on the contrary a decreased influence of perceptual priors, suggesting that the level at which the prior operates in the inference hierarchy might lead to differential effects. Finally, Jardri et al. (2017) investigated probabilistic reasoning and found schizophrenia patients to be over-counting sensory evidence and under-weighting priors, which they described using a so-called 'circular inference' model. Together these findings paint a complicated picture of Bayesian inference in schizophrenia where priors can have either increased or decreased influence depending on the task, the stimulus, the type of priors involved (e.g. low-level perceptual prior *vs.* high-level cognitive prior) or depend on particular symptom characteristics (e.g. whether patients experience hallucinations vs. delusions, Stephan et al., 2016; Sterzer et al., 2018).

A general limitation of these studies however, is that it is typically unclear to what extent the effects are driven by deficits in statistical learning (i.e. forming and updating the priors) or impaired inference *per se*. Moreover, past studies usually only qualitatively compared the behavioural results they collected with the proposed Bayesian theories.

To address these issues, we simultaneously investigated the implicit acquisition of perceptual priors, how these priors are integrated with sensory information and the influence they have on perception when stimulus is absent (i.e. experimentally induced hallucinations) in patients with schizophrenia. We used a previously developed statistical learning task (Chalk *et al*., 2010; Gekas *et al*., 2013; Karvelis *et al*., 2018) that is known to induce the rapid acquisition of the statistics of motion stimuli. In this task, participants need to report the direction of motion of a cloud of dots (estimation task) and whether they have perceived the dots or not (detection task;

on some trials no stimulus is presented). Unbeknownst to the participants, two directions of motion are more frequently presented than others. Participants implicitly and unconsciously learn those stimulus statistics. This learning influences perception such that: 1) motion stimuli are perceived as being more similar to the most frequently presented stimuli than they really are (i.e. estimation biases); 2) participants sometimes report perceiving the most frequently presented stimuli in absence of visual stimuli (a form of hallucination). In previous work (Chalk *et al* (2010); Karvelis et al (2018)), we showed that Bayesian modelling could be applied to individual participants' performances to quantitatively monitor their acquisition and use of the statistics of the stimuli (perceptual prior). We apply the same techniques in the current study to compare the perceptual priors acquired by patients with schizophrenia to those of controls.

## Materials & Methods

### Participants

A sample of 25 (22 male) individuals with psychosis (diagnosed with either DSM-IV schizophrenia, $n = 21$; or schizoaffective disorder, $n = 4$), and 23 (13 male) controls with normal or corrected-to-normal vision were recruited. Patients were recruited from inpatient and outpatient adult mental health services across NHS Lothian. Diagnoses were determined using the Structured Clinical Interview for DSM-IV (SCID-I; First, Gibbons, Spitzer, & Williams, 2002). None of the control participants met DSM-IV criteria for a psychotic disorder, bipolar disorder, or schizotypal or schizoid personality disorder. Symptom severity was measured with the Positive and Negative Syndrome Scale (PANSS) (Kay, Fiszbein, & Opler, 1987), current IQ with the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999) and pre-morbid IQ with the National Adult Reading Test (NART; Nelson, & Willison, 1991). All patients were medicated (85% on second generation anti-psychotics, 50% of these were also on mood stabilisers). The study was conducted in accordance with the national and international ethical standards for human experimentation and research (Declaration of Helsinki, 2013; Good Clinical Practice, 2014). All participants provided fully informed written consent. The study received ethical approval from the South East Scotland Research Ethics Committee 01 and NHS Lothian Research & Development. Following previous studies using the same paradigm (Chalk *et al.* 2010), we determined that 20 participants per group would give us ≥80% power to detect the correct acquisition of the prior and significance between groups (see power calculations in

supplemental material). We thus aimed to recruit between 20-25 participant per group to account for possible exclusions due to poor performance at the task.

Twenty patients and twenty-three controls successfully performed the task (see below and Supplementary Figure 1 for exclusion criteria based on performance). The demographic details of the included participants are shown in Table 1. It is of note that the patients all had relatively low levels of symptoms on the PANSS, but ongoing functional impairment. Most were assigned a PANSS positive rating of 1 (absent), 2 (minimal/questionable) or 3 (mild) for delusions, hallucinatory behaviour, suspiciousness/persecution, and unusual thought content, suggesting that they were clinically stabilised and relatively homogenous in their current symptoms (see detailed symptoms demographics Table S1 and Suppl. Fig. S6).

## Apparatus, Stimuli & Procedure

The setup for this study was similar to that used by Chalk *et al.* (2010). Motion stimuli consisted of a field of dots with a density of 2 dots/deg$^2$, moving coherently (100%) at a speed of 9°/s.

Each trial was composed of two tasks arranged as follows (**Fig. 1A**): First, participants were presented with a fixation point (0.5° diameter) for 400 ms. With the fixation point still on-screen, the motion stimulus (field of dots) was displayed along with a red bar extending from this fixation point. During the presentation of the field of dots, participants were required to estimate the direction of motion by aligning the red bar into the perceived direction of motion (*Estimation task*). The angle of this bar was randomized on each trial and participants were instructed to focus their gaze on the fixation point throughout the estimation task. The display then cleared when either the participant clicked the mouse to validate their choice (estimation) or 3000 ms had elapsed. After the estimation, a 200 ms delay was enforced before the detection screen was presented. The new screen was divided in two equal areas reading 'Dots' and 'No Dots', giving the participants a two-alternative forced choice (2-AFC). Participants were required to move the cursor to the right or the left to indicate whether they detected dots or not, and click to validate their choice (*Detection task*). The cursor then flashed green or red for correct or incorrect responses respectively. No time-outs were enforced during the detection task. Finally, the screen was cleared for 400 ms before a new trial began. Every 20 trials, participants were presented with feedback on their estimation performance in terms of average estimation error (in degrees).

## Design

Participants completed 567 trials (i.e. lasting approximately 40 minutes *vs.* 850 trials lasting 60 minutes in Chalk *et al.* 2010) with opportunities for breaks every 170 trials to prevent fatigue. Stimuli were presented at four different randomly interleaved contrast levels. The highest contrast level was 1.7 cd/m$^2$ above the 5.2 cd/m$^2$ background. There were 167 trials at zero contrast (no stimulus condition) and 67 trials at high contrast. Contrasts of other stimuli were determined using a descending 4/1 and an ascending 2/1 staircase on detection performance. Throughout the experiment, there were 90 trials with the 2/1 staircase and 243 trials with the 4/1 staircase. The use of the 2/1 and 4/1 contrast staircases ensures that individual detection performances converge to contrast thresholds corresponding to 70.4% and 84.1% correct responses respectively (Levitt, 1971; Garcia-Perez, 1998), thus being comparable between participants. From the point of view of the observer, this meant that stimulus contrasts converged and remained around the limit of visibility, hence creating a lot of uncertainty in the stimulus (its presence and its direction). We had more trials in the descending staircase so as to facilitate acquisition of the prior. On a given trial, the direction of motion for the two staircased contrast levels could either be 0°, ±16°, ±32°, ±48°, and ±64° with respect to a central reference angle. This central reference angle was randomised for each participant.

Unbeknownst to participants, we manipulated their expectations about which motion directions were most likely to occur by presenting stimuli moving at ±32° more frequently (resulting in a bimodal distribution, **Fig. 1B**). At the highest contrast level, 50% of trials were at ±32° and 50% remaining trials at random directions (i.e. not just the predetermined directions).

## Behavioural data analysis

Performance on high contrast trials was used as an indicator of whether participants were performing the task adequately. Detection accuracy of at least 70% and estimation root mean square error (RMSE) of less than 30 degrees were the minimum criteria. All 23 controls met these criteria, while 5 out of 25 patients did not meet at least one of the criteria and thus were excluded from further data analysis (**Supplementary Fig. 1**).

The main data analysis was performed on 2/1 and 4/1 staircased contrast levels and only on confirmed trials (i.e. trials where participants validated their choice with a click within 3000 ms

in the estimation task and reported seeing dots in the subsequent detection task). The first 100 trials were excluded from the analysis to allow the staircases to converge to stable contrast levels (**Supplementary Fig. 2**). After removing these trials, the luminance levels achieved by the 2/1 and 4/1 staircases were found to be overlapping (**Supplementary Fig. 2**), thus they were combined for all further analysis. Finally, since the distribution of presented directions was symmetrical around a central reference angle, the behavioural measures at equal absolute distance from the reference angle were averaged together.

In the estimation task, the variance of participants' direction estimates was large. As in previous work (Chalk *et al*., 2010; Gekas *et al*., 2013), we hypothesized that this variability resulted from a proportion of estimations being made randomly. To account for this, we fitted the individual estimation responses for each stimulus direction to the following distribution:

$$(1-\alpha)\cdot V(\mu,\sigma) + {}^{\alpha}/_{2\pi} \quad (1)$$

where $V(\mu, \sigma)$ is the circular normal (i.e. von Mises; Mardia, 1972) distribution with mean $\mu$ and width $\sigma$ and $\alpha$ corresponds to the proportion of lapse estimations (i.e. seemingly random responses). The estimation bias is then defined as the difference between the estimated mean of this circular normal distribution $\mu$ and the true motion direction, while the estimation variability corresponds to the standard deviation $\sigma$ of that distribution. The lapse rate $\alpha$ enables us to capture the proportion of estimation responses that appear random, unrelated to the true motion direction. As a first approximation, these are here parameterised as being uniformly distributed between ±180◦.

In trials where no stimulus was presented, we reconstructed the probability distributions of participants' responses over motion directions using Kernel Density Estimation (Silverman, 1986; Wand, 1994). The KDE is a non-parametric method used to estimate the probability density function from discrete measures of a continuous variable. To do so, a kernel that defines the form of the probability density function (e.g. von Mises kernel) was placed at each of the observed measurement. Then, all the individual kernels were summed to create the probability density function of the random variable (motion direction).

To assess the effects of the acquired expectations in both groups and to simultaneously determine whether patients differed from controls, we performed 2-way mixed ANOVA with motion direction as a within-subjects factor and group as a between-subjects factor. We ran

Bonferroni-corrected pairwise comparisons to determine which motion directions were driving the effects. To control for possible deviations from normality, we also ran non-parametric tests to confirm the parametric 2-way ANOVA findings (see SI). All analysis was done using SPSS version 24. To determine the strength of the evidence for the null hypothesis (i.e. finding no group difference vs. finding evidence that the groups are similar) we also report Bayes factors using the Bayesian statistical software package JASP. We report Kendall correlation coefficients whenever the data is ordinal with rank ties and/or strong outliers.

## Modelling

To test for individual variability in the underlying perceptual inference and to obtain more direct measurements of the acquired expectations, we fitted a range of models to our data. The first class of models assumed that the biases were of a perceptual nature, as conceived in the Bayesian framework: sensory information is combined with a learned prior of the stimulus statistics in a probabilistic way. The simple 'BAYES' model assumed that the likelihood precision was constrained to be the same across all presented motion directions (corresponding to the hypothesis that there was no learning in the likelihood due to the distribution of the motion directions). An additional variant of the 'BAYES' model tested the hypothesis that the lapse estimations were not uniformly distributed but instead were sampled from the acquired prior expectations. We call these responses 'prior-based lapses' (**Fig. 4**). This model was termed 'BAYES_P' and was otherwise equivalent to 'BAYES'.

Another class of models assumed that task performance could be explained by response strategies that do not involve Bayesian integration (heuristic models). According to these models, on any given trial participants responses were based solely on either the prior expectations or sensory information. We considered four variations of response strategy models (see Methods and Supplementary information for details). Below we present the Bayesian models as they provided a better explanation to the data (see Figure 5, model comparison).

### Bayesian model

Following the Bayesian framework, we assumed that participants combined sensory information (likelihood) with their expectations about the motion direction (prior) on every trial. The sensory likelihood of the observed motion direction ($\theta_{sensory}$) was parameterised as a von Mises circular normal distribution with variance $\sigma_{sensory}$:

$$p_{likelihood}(\theta_{sensory}|\theta) = V(\theta, \sigma_{sensory}) \quad (2)$$

The mean of this distribution depended on the actual presented motion direction ($\theta_{actual}$), but to account for trial-to-trial variability it was sampled from another von Mises distribution centred on $\theta_{actual}$ with variance $\sigma_{sensory}$ (i.e. $V(\theta_{actual}, \sigma_{sensory})$). This is necessary because on any given trial the observer does not have direct access to the true motion direction. Therefore, the likelihood cannot be centred on the true motion direction, but instead has to be centred on a noisy estimate of it (Stocker & Simoncelli, 2006; Kording et al, 2007).

We then hypothesized that participants acquire an approximation of the 'true' prior from experience ($p_{prior}$), representing the participants' expectations of motion directions. The acquired priors were parameterized as the sum of two von Mises circular normal distributions, centred on motion directions $\theta_{expected}$ and $-\theta_{expected}$, each with variance $\sigma_{expected}$:

$$p_{prior}(\theta) = \frac{1}{2}[V(-\theta_{expected}, \sigma_{expected}) + V(\theta_{expected}, \sigma_{expected})] \quad (3)$$

Combining the prior and the likelihood gives us the posterior probability that the stimulus is moving in a direction $\theta$:

$$p_{posterior}(\theta|\theta_{sensory}) \propto p_{likelihood}(\theta_{sensory}|\theta) \cdot p_{prior}(\theta) \quad (4)$$

The perceived direction, $\theta_{perceived}$, was taken to be the mean of the posterior distribution (almost identical results would be obtained by using the maximum instead).

Finally, we accounted for motor noise (related to aligning and clicking the mouse) and lapse estimations, such that:

$$p(\theta_{estimate}|\theta_{perceived}) = (1 - \alpha_{prior\text{-}based}) \cdot V(\theta_{perceived}, \sigma_{motor}) + \alpha_{prior\text{-}based} \cdot p_{prior}(\theta) \quad (5)$$

where $\sigma_{motor}$ is the motor noise and $\alpha_{prior\text{-}based}$ describes the chance of making an estimation response that appears random in that it is unrelated to the actual stimulus. In the model that we call 'Bayes_P' for 'Bayes with Prior-based lapse estimations', we assume that these random estimations are drawn from the participants' acquired expectations – $p_{prior}(\theta)$. The motivation for

this model came from observing: i) that the lapse responses were found not to be uniformly distributed in the data, and ii) that if such random responses are effectively made when the stimulus is below detection threshold (a common occurrence since contrasts hover around threshold), and thus become invisible for the observer, those responses are formally equivalent to what we describe as hallucinations, which seem to be sampled from the prior (see also Laquitaine & Gardner 2018 for a related model).

Finally, we also tested variants of this model, where lapse estimations were uniformly distributed, rather than following the participants expectations (model 'BAYES'), or that due to increased exposure to stimuli at specific angles, sensory uncertainty $\sigma_{sensory}$ could vary across angles (0°, ±16°, ±32°, ±48°, ±64°; model 'BAYES_var'), or that sensory uncertainty varied only at the most presented directions (model 'BAYES_varmin' – see Methods and Supplementary information for details).



## Results

### Detection performances and contrast levels

Contrast staircases converged to stable luminance levels after about 100 trials for both groups (**Supplementary Fig. 2**); Controls converged to 0.41 cd/m$^2$ (±0.03) for the 2/1 staircase and 0.46 cd/m$^2$ (±0.03) for the 4/1 staircase, while patients converged to 0.57 cd/m$^2$ (±0.04) for the 2/1 staircase and 0.62 cd/m$^2$ (±0.05) for the 4/1 staircase. These results confirm previous findings (Skottun and Skoyles, 2007) suggesting that patients with schizophrenia display significantly poorer contrast-sensitivity in comparison to controls (2/1 staircase: $Z = 3.15$, $p = 0.002$; 4/1 staircase: $Z = 2.90$, $p = 0.004$, two-tailed rank-sum test).

### Statistical learning

First, we investigated whether participants acquired the statistics of the stimulus. To do so, we looked at patterns suggestive of statistical learning in each group, namely attractive biases towards the most frequent directions, decreased reaction times and improved detection performance for the most frequent directions (**Fig. 2**).

**Estimation performance**

In the analysis of the estimation task, we looked only at the staircase contrasts stimulus trials where participants both reported seeing a stimulus and clicked on the mouse during stimulus presentation to indicate their estimate of motion direction.

To investigate whether the participants' perceived motion-directions were biased, we measured the difference between the true motion direction and the motion direction reported by the participants. **Fig. 2A** displays the average estimation bias plotted against the true motion direction for each group. Overall, there was a significant effect of motion direction on the estimation bias ($F(2.45, 100.52) = 15.37$, $p < 0.001$, $\eta_p^2 = 0.273$, Greenhouse-Geisser correction $\varepsilon = 0.613$), but no differences between the groups (group main effect: $F(1, 41) = 0.83$, $p = 0.369$, $\eta_p^2 = 0.001$; with positive evidence for the null hypothesis, $BF_{01} = 3.99$); and no group*angle interaction ($F(2.45, 100.52) = 1.64$, $p = 0.193$, $\eta_p^2 = 0.038$). Pairwise comparisons (with Bonferroni correction) revealed that there was an attractive bias towards ±32° at ±48° and ±64° (MD = 4.858, p = 0.002; and MD = 14.395, p < 0.001, respectively), but not at ±16° (MD = 1.818, p = 0.955). Together, these results confirm that both patients and controls were biased towards perceiving motion directions as being more similar to the most frequently presented directions than they really were, consistent with having acquired the priors that approximate the statistics of the stimulus.

We also investigated whether the effects of acquired prior expectations were reflected in the variability of estimations, namely a decrease of variability for the expected directions (**Fig. 2B**). We found a significant main effect of motion direction ($F(3.07, 125.69) = 5.18$, $p = 0.002$, $\eta_p^2 = 0.112$, Greenhouse-Geisser correction $\varepsilon = 0.766$), but no differences between the groups (main effect of group: $F(1, 41) = 0.02$, $p = 0.880$, $\eta_p^2 = 0.001$; with positive evidence for the null hypothesis, $BF_{01} = 3.74$); and no group*angle interaction $F(3.07, 125.69) = 1.58$, $p = 0.196$, $\eta_p^2 = 0.037$). Pairwise comparisons showed that variability at 0° stood out the most, being significantly larger than at ±32° and ±48° (MD = 4.680, p = 0.012; and MD = 4.733, p = 0.025, respectively), although not different than at ±16° and ±64° (MD = 3.044, p = 0.239 and MD = 2.990, p = 0.541). The increased variability in the region between the two modes reflects their conflicting influence on the percepts in this region.

Finally, we analysed lapse estimations, which were captured by the 'α' term in Eq. (1) and which were assumed to arise from random responses on some of the trials (**Fig. 2C**). We found both motion direction and group main effects to be significant ($F(4,164) = 5.76$, $p < 0.001$, $\eta_p^2 = 0.123$ and $F(1, 41) = 6.41$, $p = 0.015$, $\eta_p^2 = 0.135$, respectively), with patients exhibiting fewer lapse estimations. Pairwise comparisons revealed that lapse rate at 0° was significantly smaller than at all other directions (±32°, MD = 4.814, p = 0.001; ±48°, MD = 6.010, p = 0.007; ±64°, MD = 4.951, p = 0.003), except for ±16° (MD = 3.043, p = 0.393). The finding that the estimated lapses would depend on the presented motion direction was surprising, since it suggests these lapses are not made completely randomly. Below, we show that these seemingly random responses can be best described as being sampled from the acquired prior distribution (see Modelling results).

**Reaction times and Detection Performance**

Next, we examined how participants' acquired expectations influenced reaction times and the detection of stimulus. The estimation reaction times **(Fig. 2D)** show a significant main effect of motion direction ($F(2.73,111.76) = 10.80$, $p < 0.001$ $\eta_p^2 = 0.209$, Greenhouse-Geisser correction $\varepsilon = 0.681$). This was driven by decreased reaction times at the most frequent directions as revealed by pairwise comparisons: reaction time at ±32° was significantly shorter than at all other directions (0°, MD = 0.104, p = 0.001; ±16°, MD = 0.068, p = 0.004; ±64°, MD = 0.139, p < 0.001), except for ±48° (MD = 0.027, p = 1.000). Furthermore, patients were found to also be significantly slower than controls ($F(1, 41) = 4.11$, $p = 0.049$, $\eta_p^2 = 0.091$), but we found no interaction between group and motion direction ($F(2.73, 111.76) = 0.66$, $p = 0.563$, $\eta_p^2 = 0.016$). Slow reaction time is a hallmark of schizophrenia that has been documented thoroughly in the literature in simple reaction-time tasks using visual and/or auditory stimuli (e.g., see Nuechterlein, 1977; Fioravanti *et al.*, 2012).

An even more direct way of assessing how the acquired expectations influenced the detection of stimulus is to analyse the fraction of trials where participants explicitly report seeing or not seeing the stimulus (**Fig. 2E**). We found that the detection of stimulus was greatly affected by the presented motion direction ($F(2.36,96.64) = 8.51$, $p < 0.001$, $\eta_p^2 = 0.172$, Greenhouse-Geisser correction $\varepsilon = 0.589$), with stimulus at ±32° being the most frequently detected direction as shown by pairwise comparisons: detection at ±32° was significantly better than at all other

directions (0°, MD = 9.59, p = 0.004; ±16°, MD = 6.48, p = 0.001; ±48°, MD = 6.54, p = 0.001; ±64°, MD = 12.35, p < 0.001). However, the groups were not found to be different (main group effect: F(1,41) = 3.62, p = 0.064, $\eta_p^2$ = 0.081; although there was no evidence for the null hypothesis either, $BF_{01}$ = 0.97); group*motion direction interaction was also non-significant: F(2.36,96.64) = 1.11, p = 0.340, $\eta_p^2$ = 0.026).

Overall, these results indicate that, in terms of detection responses (hit rates and reaction-time), similar benefits of statistical learning were present in both patient and control groups. Overall, behavioural measures suggest that prior effects (e.g. Bias, RT, and Hit rate) became significant as early as within first ~100 trials for both patients and controls (see Supplementary Figures 3-4).

**Perceived motion in absence of visual stimuli (hallucinations)**

Finally, we investigated whether the acquired statistics about the motion stimulus affected the participants' perception on trials where no stimulus was presented, but where participants reported both a motion direction and seeing a stimulus. We refer to this effect as hallucinations. These hallucinations in our perceptual task are of course different in terms of content and complexity from the visual hallucinations observed in psychosis. However, the underlying mechanisms might be informative to the understanding of illusions and hallucinations in schizophrenia (Silverstein and Keane 2011a; 2011b; Notredame *et al.*, 2014).

To quantify the probability ratio that participants made estimates that were closer to the most frequently presented motion directions relative to other directions, we multiplied the probability that participants estimated within 16° of these motion-directions by the total number of 32° bins:

$$p_{ratio} = p(\theta_{estimate} = \pm 32(\pm 16)^\circ) \cdot N_{bins} \qquad (6)$$

This probability would be equal to 1 if participants were equally likely to estimate within 16° of ±32° as they are to estimate within the other 16° bins.

We found that the median value of *'$p_{ratio}$'* was significantly greater than 1 for both patients and controls (median($p_{ratio}$) = 2.88, p = 0.003 and median($p_{ratio}$) = 2.75, p <0.001, respectively; two-tailed signed-rank test), indicating that both patients and controls were much more likely to

hallucinate the most frequent motion directions as opposed to all other directions (**Fig. 3A, B**). Bayesian statistical analysis provided positive evidence for the groups being the same in this measure ($BF_{01}$ = 3.32). Finally, hallucinations of the most frequent directions (i.e. hallucinations at ±32°±16°) were quick to develop during the course of the experiment: they became significant after only 150 trials for both controls (*p=0.036,* one-tailed signed-rank test; Supplementary Fig. 3D) and patients (*p=0.035*, one-tailed signed-rank test; Supplementary Fig. 4D).

While both patients and controls hallucinated predominantly towards the most frequently presented directions (i.e. prior-based hallucinations), patients exhibited fewer of such hallucinations (**Fig. 3D**; hallucinations within ±16º of ±32º; p = 0.016, two-sided rank-sum test), and also exhibited less hallucinations overall (**Fig. 3C**; p = 0.004, two-sided rank-sum test). We wanted to know whether the severity of the symptoms was predictive of the magnitude of this effect. However, we found no correlations between the number of hallucinations and the PANSS positive, negative, general or total scores nor duration of illness, or between the daily-dosage of anti-psychotics (Olanzapine equivalent; Leucht et al., 2015) and the total number of hallucinations.

We also investigated the distribution of responses when participants estimated the direction of motion but reported not seeing any dots. Interestingly, in this subset of trials, and unlike in our previous work (Chalk et al, 2010), estimations were also more likely than chance to be made around the most frequent motion directions by both patients and controls (median($p_{ratio}$) = 1.24, p = 0.002 and median($p_{ratio}$) = 1.41, p = 0.045, respectively; two-sided signed rank test). One explanation for this might be habitual effects - when the bar is moved towards the most frequent directions out of motor habit. Another possible explanation is that participants might have hallucinated stimuli in the expected prior directions, but their confidence about their percept being very low, they sometimes chose to report not seeing any dots in the hope to give the correct answer (each detection response was followed by immediate feedback).

## Modelling results

We fitted the models to the behavioral data and computed the Bayesian Information Criterion (BIC) for each participant, which quantifies goodness of fit while penalizing for extra model complexity (preventing overfitting). We found that the BAYES_P model had the smallest BIC

for both patients and controls (**Fig. 5**) – indicating best performance, with a difference in BIC between the winning model (BAYES_P) and the second best model (BAYES) being larger than 10. This is equivalent to a log Bayes factor larger than 10, and is considered to be very strong or 'decisive' evidence in favor of the winning model (Kaas & Raftery, 1995). Model fits also showed that BAYES_P was much better at fitting lapse estimations, confirming that such estimations were sampled from the acquired prior distribution instead of being uniformly distributed (**Fig. 6C and G**). We also ran a random effect Bayesian model selection analysis (Rigoux et al., 2014; Daunizeau et al., 2014) to ensure that the favored model was not being selected due to a subset of participants in each group. The analysis confirmed that BAYES_P was best at describing behavior for both groups (Fig. 5C and D). Moreover, we found a strong correlation between the participant's prior-based lapse rate parameter recovered via BAYES_P and their number of hallucinations ($\tau_b = 0.657$, $p < 0.001$; Kendall's correlation; **Fig. 7B**). This suggests that prior-based lapses (in the presence of a (very weak) stimulus) and hallucinations (in absence of a stimulus) originate from the same mechanism (**Fig. 7A**).

Finally, we compared patients and controls on the basis of BAYES_P parameter estimates (**Fig. 6I-L**). Consistent with the behavioral data analysis, we found no differences in the acquired prior expectations (**Fig. 6I, J;** the mean of acquired prior: $p = 0.874$, $BF_{01} = 3.32$; and the uncertainty in the acquired prior: $p = 0.401$; two-tailed rank-sum test; $BF_{01} = 2.95$). There were no differences in the precision of sensory likelihood (**Fig. 6K**, $p = 0.742$, two-tailed rank-sum test; $BF_{01} = 2.96$). Lastly, just as in the behavioral data, we found that patients made less prior-based lapse estimations (**Fig. 6L**, $\alpha_{\text{prior-based}}$ : $p = 0.024$, two-tailed rank-sum test), suggesting that patients were less likely to perceive a sample from their prior in the presence of a weak (possibly subthreshold) stimulus.

**Parameter recovery for model BAYES_P**

To gauge the reliability of our modelling results, we performed parameter recovery for the winning BAYES_P model. Parameter recovery consists of simulating synthetic data with different sets of known parameter values ('true parameters') for a given model and then fitting the same model to the synthetic data to estimate and recover these parameters ('recovered parameters'). The strength of correlation between the actual and recovered parameters measures the reliability of modelling results. Parameter recovery, just as the parameter estimation from behavioral data, is sensitive to any correlations that might be present among the model parameters, to the choice of parameter estimation methods and also to the amount of data used

for model fitting. Therefore, parameter recovery serves as a crucial step in validating the reliability of the modelling results (Palminteri, et al., 2017).

We found that the winning BAYES_P model recovered parameters very well, which was reflected in the coefficient of determination ($R^2$) for all recovered parameters being $R^2 \geq 0.84$ (**Fig. 8**).

## Discussion

We were interested in testing the emerging model of schizophrenia proposing that the disorder could stem from deficits in Bayesian inference (Corlett *et al*., 2009a, 2009b; Fletcher and Frith, 2009; Adams *et al*., 2013; Schmack et al., 2013, 2015, 2017; Teufel et al., 2015; Powers et al., 2017, Jardri et al., 2017). The experimental paradigm we chose is well suited to quantitatively assess the acquisition of sensory priors, how these priors are used in perception, as well as to quantify inter-individual variability in the learning and inference process (Chalk et al 2010, Karvelis et al, 2018).

### *Acquisition of visual prior expectations*

We found that both the control and patient groups implicitly learned the statistics of the motion stimuli and that those expectations modified their perception, consistent with them acquiring a Bayesian prior of the stimulus statistics and combining it with sensory evidence, replicating our previous results (Chalk et al, 2010). This was reflected by attractive estimation biases towards the frequently presented directions, faster reaction times and higher detection rates at these directions, as well as hallucinated motion directions in the absence of stimulus predominantly following the most frequent directions.

Patients with schizophrenia were not qualitatively, nor quantitatively different from controls in the measures used to assess learning of the task statistics. This suggest that while there are clearly domain-specific perceptual (e.g. hallucinations) and learning deficits in schizophrenia (e.g. jumping to conclusions), our study demonstrates that these deficits are not due to a domain-general impairment in the acquisition and/or utilization of statistical information in the environment. That is, we find that patients with chronic schizophrenia do not appear to be impaired in the acquisition of visual statistical priors in our task.

These results are consistent with studies finding no deficit in implicit learning in schizophrenia (Kéri *et al*., 2000; Danion *et al*., 2001; Marvel *et al*., 2005; for review see: Gold *et al*., 2009). In contrast with studies that assay explicit statistical learning and inference using more cognitive tasks (i.e. usually believed to involve frontal cortical regions), here we measured implicit statistical learning of visual stimuli that could be embodied in visual processing areas rather than frontal cortices (Kok *et al*., 2013). In fact, patients with schizophrenia appear relatively spared in implicit learning tasks that do not require integrating feedback after each trial (Gold *et al*., 2009). These results are also consistent with our previous study using the same paradigm showing intact statistical learning in participants with high schizotypal traits (Karvelis et al, 2018).

***Impact of acquired visual prior expectations***

We found no difference between patients and controls as to the influence of the acquired expectations on their performance regarding estimation of the motion directions. They were not more or less biased towards the most frequent directions, nor more or less variable in those estimations.

Patients were found to differ from controls in three ways, however. First, patients with schizophrenia displayed significantly poorer contrast discrimination thresholds and slower reactions times, as documented in previous studies. Second, and more interestingly, patients reported significantly fewer hallucinations at all directions and fewer hallucinations of the most frequently presented motion directions.. Third, patients exhibited fewer prior-based lapse estimations than controls on low contrast trials. We argue that prior-based lapses (defined for low contrast trials) and hallucinations (defined for no-stimulus trials) are underlined by the same mechanism. Because of the convergence of the contrast staircases, contrast levels in our task hover around the detection threshold, which means that on a significant number of trials the stimulus contrast falls below the participants' threshold of perception, effectively becoming invisible and equivalent to trials with no stimulus. If participants experience hallucinations on these trials (and thus report that they have perceived a stimulus), these will be expressed in our results as prior-based lapse estimations. In support for this interpretation, we found that there were significantly more lapse estimations made on trials when the stimulus contrast was below the participants' 75% detection threshold (based on their individual psychometric curve) than when it was above ($F(1,84) = 12.61$, $p < 0.001$; **Fig. 7A**) and we found a strong correlation between prior-based lapse rate on threshold contrast trials and number of hallucinations on no-

stimulus trials (**Fig. 7B**). Together, this suggests that although patients appear to acquire the same prior as controls, they tend to hallucinate this prior less often than controls when there is no stimulus, or when the stimulus is below detection threshold.

Interestingly, this diminished impact of prior expectations in the detection task was not observed in our previous work in participants with high schizotypal traits (Karvelis et al, 2018). suggesting that it might be characteristic of the chronic state of schizophrenia.

The fact that patients exhibit fewer prior-based lapses and hallucinations suggests that their perception is less influenced by their learned prior expectations than control participants. This is consistent with previously reported findings suggesting that patients with chronic schizophrenia are less sensitive to expectation-driven illusions (e.g. the hollow-mask illusion) than controls (Tschacher *et al*., 2006; Dima *et al*., 2009; Crawford *et al*., 2010; Horton and Silverstein, 2011; Keane *et al*., 2013, Notredame et al., 2014). This finding is also in line with results from Schmack et al. (2013, 2015), reporting a decreased influence of induced expectations (priors) on perception.

It is intriguing however that the influence of prior expectations is weaker in the detection task, but similar to that of controls for the estimation task. This may reflect the fact that conscious detection and estimation may depend on different processing stages, with possibly different representations for the prior distributions, or involve different mappings between representations and responses (Petzschner et al. (2015)). In future work we will aim to address the possible neural substrate for the effects we describe, by modelling how prior distributions could modulate visual responses or their read-out (Seriès, Stocker and Simoncelli, 2009) to explain differential biases in different dimensions.

Other factors might contribute to the absence of differences in the estimation bias.

One factor might be that very few of our patients reported clinically significant levels of hallucinations (PANSS items >3), and that this propensity might be more pertinent than the diagnosis of schizophrenia *per se* (Powers et al, 2017). We found some evidence supporting the fact that patients who express fewer or weaker positive symptoms differ more from controls in our task than patients who express more or stronger positive symptoms: PANSS Positive symptom score and prior-based lapses (estimated via BAYES_P) were close to being negatively correlated ($\tau_b$=-0.314, p= 0.063; Kendall's correlation; **Supplementary Fig. 5A**); this relationship was also backed up by a much stronger correlation for when lapse estimations were estimated directly from the behavioural data using Eq. (1) ($\tau_b$ = -0.465, p = 0.006; Kendall's correlation; **Supplementary Fig. 5B**). However, it should be noted that no correlation was

found between the PANSS Positive symptom score and the number of hallucinations exhibited on no-stimulus trials.

Similarly, the absence of stronger effects might be related to illness duration: weaker estimation biases may be characteristic of earlier stages of the illness but may not be detectable anymore in our patients as they have been generally ill for a long time, due to medication or compensatory mechanisms. We find some support for this idea with a significant correlation between duration of illness and magnitude of estimation bias ($\tau_b$ = 0.523, p = 0.003; Kendall's correlation; **Supplementary Fig. 5C**).

Finally, since the differences appear only for the detection part of the task, it might be that chronic patients have simply developed an increased perceptual threshold. Following this idea, patients would require stronger evidence (i.e. sharper posterior) in order to perceive a stimulus or to make a decision about the presence of a stimulus. This is consistent with the fact that patients required higher stimulus contrasts and integrated information over longer periods of time before responding (slower reaction times during the estimation task). We therefore hypothesize that it is a possible adaptation strategy used by patients over time to minimize responses to stimuli that were not truly present (i.e. their psychotic hallucinations).

Taken together, our results suggest that statistical learning is intact in relatively well patients with chronic schizophrenia on stable doses of second generation antipsychotic medication. The impact of their acquired priors is also the same as that of controls in the estimation task, but weaker in the detection task. These results are surprising in view of the current prominent theories proposing that schizophrenia is a disorder of predictive processing or Bayesian inference, and suggest ways in which these accounts need to be nuanced. The similarity of controls and patients' performance in our task may however be related to the success of treatment or some other adaptive process related to illness duration. Future work will aim at testing participants at earlier stages of the illness, and ideally before pharmacological treatment has begun.

## Funding

VV was supported by grants EP/F500386/1 and BB/F529254/1 for the University of Edinburgh School of Informatics Doctoral Training Centre in Neuroinformatics and Computational Neuroscience (www.anc.ed.ac.uk/dtc/) from the UK Engineering and Physical Science Research Council (EPSRC), and the UK Medical Research Council (MRC). PK was funded by Engineering and Physical Sciences Research Council. PS was supported by the NARSAD grant (19271). ARS was funded by NSF Grant BCS-1057625 and NIH grant 1R01EY023582. The collaboration between PS and ARS was supported by a Marie Curie Exchange Grant FP7-PEOPLE-2009-IRSES-247543. Some of this work was completed while PS was visiting the Simons Institute for the theory of Computing. Psychophysical testing, neuropsychological testing and structured clinical interviews were conducted at the University of Edinburgh, Royal Edinburgh Hospital. The funding bodies had no control over the design, analysis or acquisition of the data for this study.

## Conflict of interest

In the past three years, SML has received personal financial support from Janssen, Otsuka and Sunovion and research grant income from Janssen and Lundbeck in connection with therapeutic initiatives for psychosis. The authors VV, PK, KR, PS and ARS have no conflict of interest.

| Characteristics | Controls (n=23) | Patients with schizophrenia (n=20) | Significance level p-value |
|---|---|---|---|
| **Gender (males)** | 13 | 17 | 0.04 |
| **Age (years)** | 33.86 (12.08) | 39.40 (9.43) | 0.07 |
| **Premorbid IQ** | 115.55 (4.41) | 113.18 (8.76) | 0.58 |
| **Current IQ** | 117.45 (7.28) | 111.15 (10.70) | 0.06 |
| **PANSS Positive Scale** | 8.86 (2.29) | 12.55 (5.01) | <0.01 |
| **PANSS Negative Scale** | 8.41 (2.24) | 12.20 (5.05) | <0.01 |
| **PANSS General Scale** | 22.27 (6.48) | 26.10 (8.35) | 0.12 |
| **PANSS Total** | 39.55 (9.22) | 50.85 (15.86) | <0.01 |
| **GAF** | 74.81 (11.42) | 54.75 (14.36) | <0.001 |
| **OLZ eq. (mg/day)** | --- | 12.61 (6.24) | N/A |
| **Illness duration (years)** | --- | 13.33 (9.01) | N/A |

**Table 1**. Participants' demographics. PANSS = Positive and Negative Symptom Scale (lower score is better), GAF = Global Assessment of Functioning (higher score is better), OLZ eq. = Olanzapine equivalent dosage in mg/day. Values indicate mean and (standard deviation). For gender, group comparisons were done using Chi-square test, for all other measures, two-sided Wilcoxon rank-sum test was used. See Table S1 and Supplementary Figure 6 for a more detailed description of the clinical characteristics.

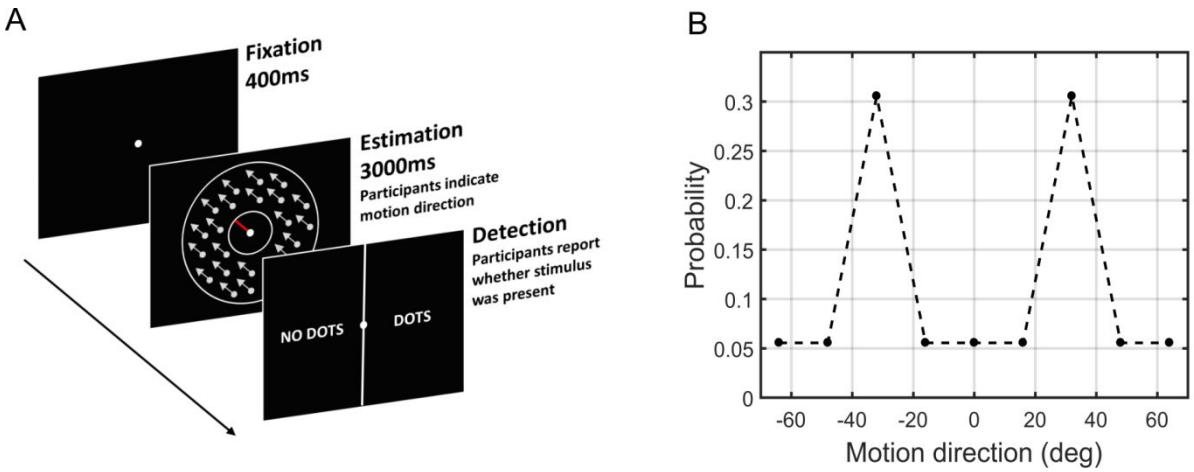



**Figure 1: (A)** Experimental procedure. Participants were presented with a fixation point followed by the motion stimulus and a response bar (red bar) that they were instructed to align to the perceived motion-direction. The screen was cleared either when participants clicked to validate their estimation or 3000 ms had elapsed. A new screen appeared with a two-alternative forced choice task (2-AFC), requiring participants to indicate whether they perceived the dots during the estimation task. **(B)** Probability distribution of the motion directions. Unbeknownst to participants, the distribution of motion direction was bimodal (i.e. stimuli appeared most often at ±32° from a central direction). The central direction was randomised for each participant.

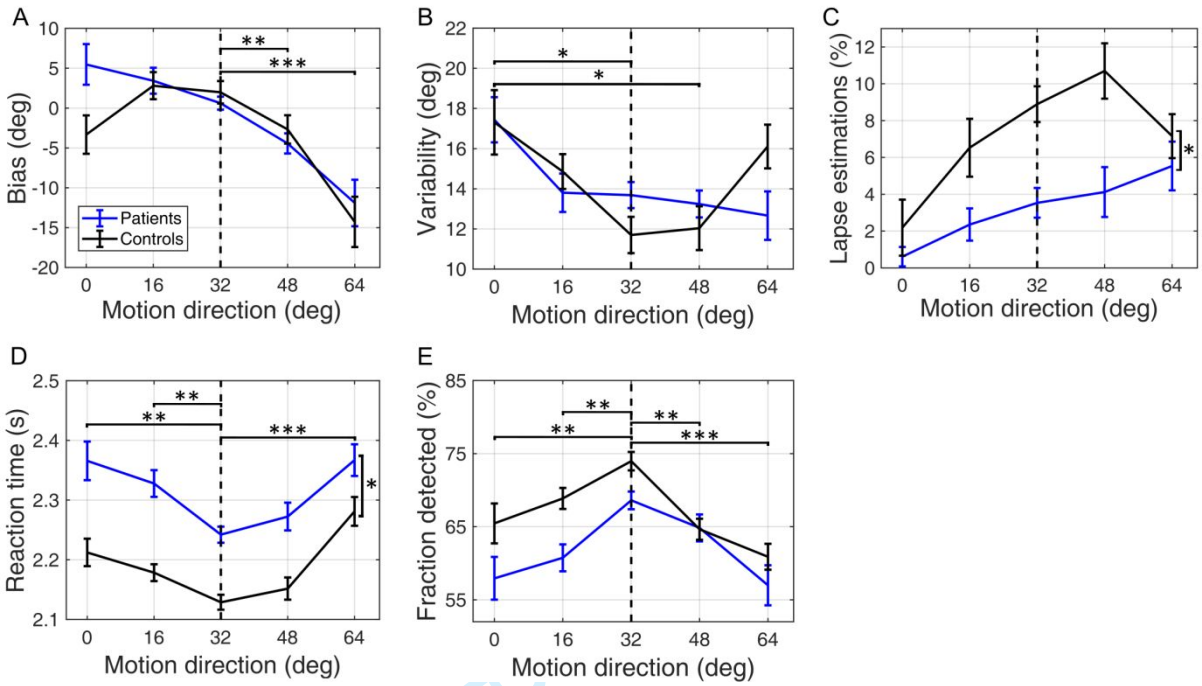



**Figure 2:** Performance on low contrast trials by patients (blue lines) and controls (black lines). (**A**) Mean estimation bias as a function of the true motion direction. (**B**) Estimation standard deviation (i.e. variability) as a function of the presented of motion direction. (**C**) Lapse estimations as a function of motion direction, estimated using (eq. 1). (**D**) Reaction times during the estimation task as a function of motion direction. (**E**) The fraction of trials in which the stimulus was detected as a function of the presented motion direction. The error bars represent within-subject standard error. The vertical dashed lines correspond to the most frequently presented motion directions (i.e. ±32°). *, ** and *** indicate significance levels at $p<0.05$, $p<0.01$, $p<0.001$ respectively. These results are not affected by any violation of the normality assumption (see SI).

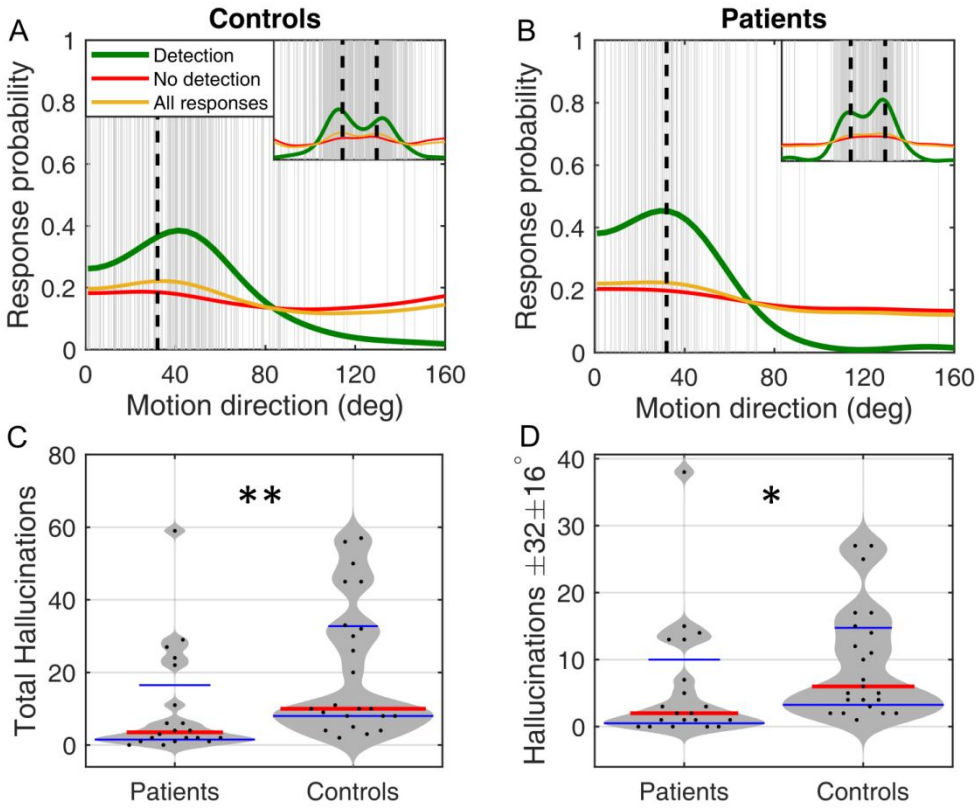



**Figure 3.** Estimation responses in the absence of stimulus. **(A, B)** Distribution of the estimation responses by patients and controls, respectively. The vertical grey lines represent reported motion directions when no stimulus was present (i.e. hallucinations) pooled across each group. The green line denotes the probability distribution of these hallucinations, which was produced using Kernel Density Estimation. The red line denotes probability distribution of responses that were followed by participants' reporting seeing no stimulus. The orange line denotes all estimations regardless of the detection response. In the main plots the data is averaged across the central motion direction, while the insets show the corresponding distributions across the full range. **(C, D)** Comparison of patients and controls by **(C)** the total number of hallucinations ($p = 0.004$, two-sided rank-sum test) and **(D)** the number of hallucinations around the most frequently presented motion directions (within ±16º of ±32º; $p = 0.016$, two-sided rank-sum test). Red horizontal lines denote median values; blue horizontal lines denote 25th and 75th percentiles. Black dots denote individual participants, grey areas represent density of the data points. * and ** indicate significance levels at $p< 0.05$ and $p<0.01$ respectively.

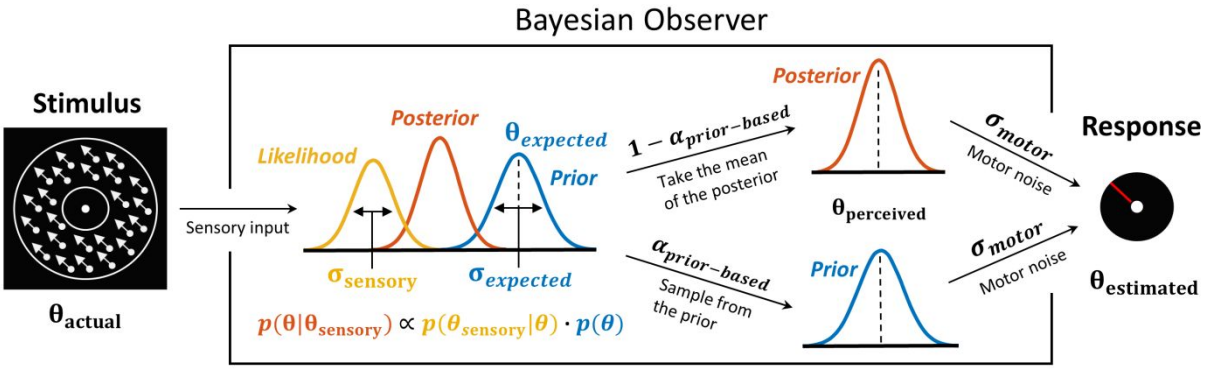



**Figure 4**. Bayesian model of estimation response for a single trial. The actual motion direction ($\theta_{actual}$) is corrupted by sensory uncertainty ($\sigma_{sensory}$), and then combined with prior expectations (mean $\theta_{expected}$ and uncertainty $\sigma_{expected}$) to form a posterior distribution. The perceived motion direction ($\theta_{perceived}$) then corresponds to the mean of the posterior distribution. However, on a fraction of trials, determined by the lapse rate ($\alpha_{prior\text{-}based}$), the perceived motion direction is sampled from the prior. Finally, in both cases, the response ($\theta_{estimated}$) is made by perturbing $\theta_{perceived}$ with motor noise ($\sigma_{motor}$). This results in 4 free model parameters: $\sigma_{sensory}$, $\sigma_{expected}$, $\theta_{expected}$ and $\alpha_{prior\text{-}based}$. The motor noise ($\sigma_{motor}$) is estimated from high contrast trials and is used as a fixed parameter.

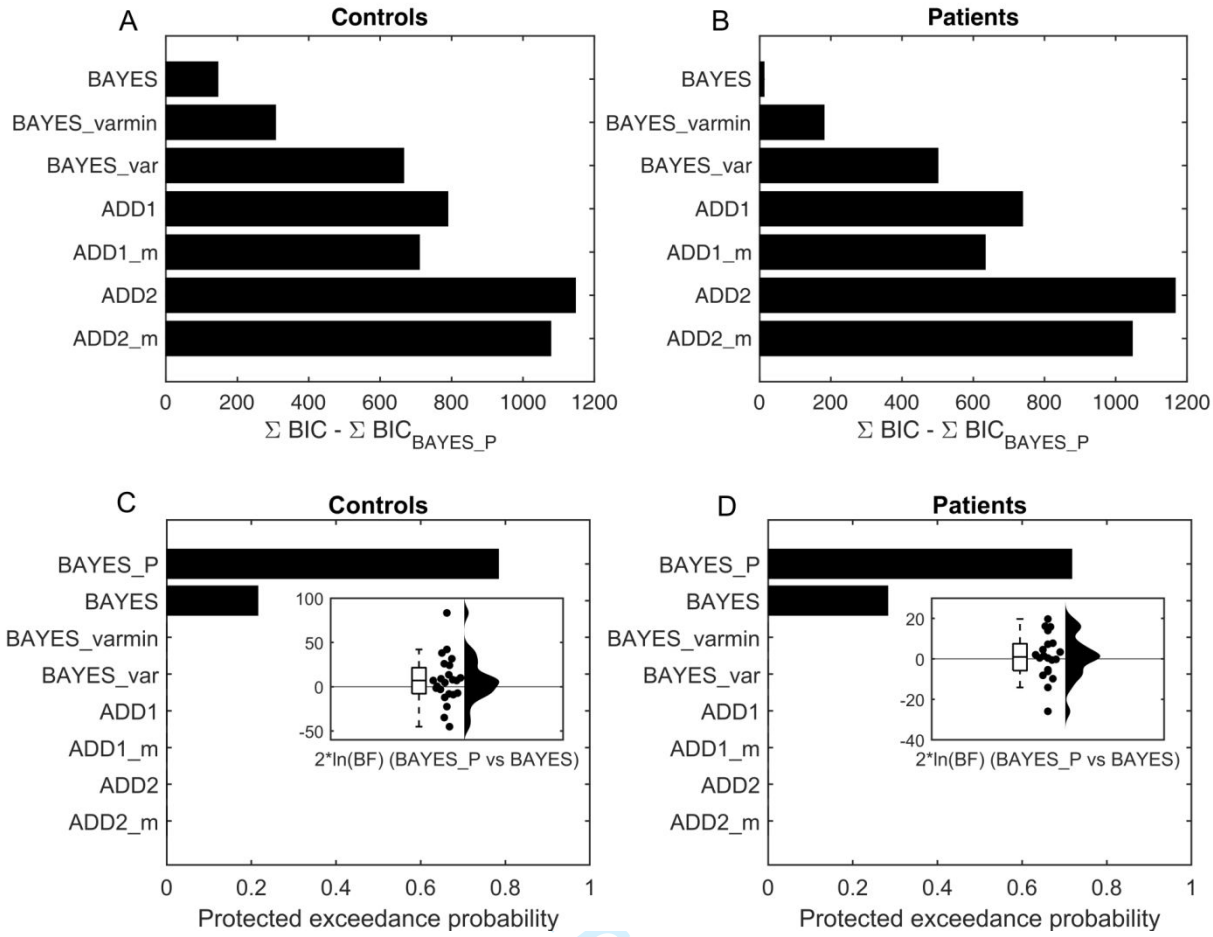



**Figure 5.** Model comparison and selection. (A,B) Fixed effects model selection using Bayesian Information Criterion (BIC) for (A) controls and (B) patients. X-axis measures the relative difference between BIC of each model (as indicated on Y-axis) and BIC of BAYES_P (winning model) summed across participants. Smaller BIC indicate a better model fit while penalizing for added model complexity. For both patients and controls BAYES_P provided the best model evidence, 12 BIC units better than BAYES for patients and 146 BIC than BAYES for controls. (C, D) Random effect Bayesian model selection for (C) controls and (D) patients. Higher protected exceedance probability indicates a model having a higher likelihood of being more frequent among the subjects. For both patients and controls BAYES_P was the most likely model. Insets show the distribution of Bayesian Factor for BAYES_P vs. BAYES summarized by a boxplot, jittered data scatter and a probability density that was estimated using a normal kernel.

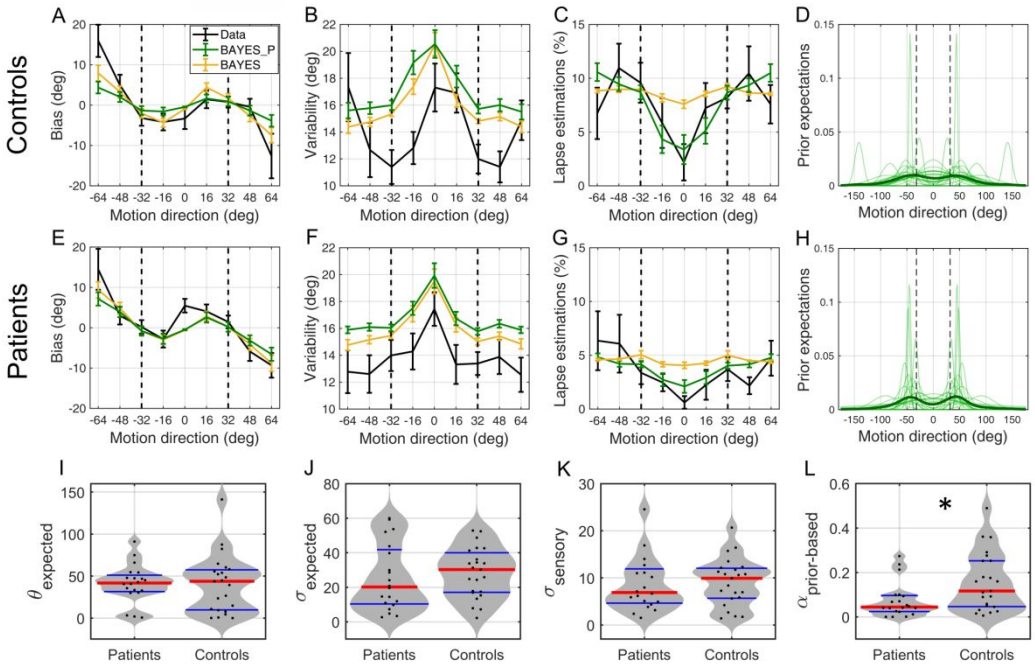



**Figure 6.** Model fits and parameter estimates. **(A-H)** Model fits for the best fitting model BAYES_P (green) and the second best model BAYES (yellow), to the behavioral data (black). **(A-D)** controls and **(E-H)** patients. **(A, E)** estimation bias, **(B, F)** estimation variability, **(C, G)** estimation lapse rate, **(D, H)** prior expectations of each individual (transparent green) and group average (thick green) as estimated via BAYES_P model. The vertical dashed lines correspond to the most frequently presented motion directions (i.e. ±32°). The error bars represent within-subject standard error. **(I-L)** Comparison of BAYES_P model parameter estimates of patients and controls. **(I)** $\theta_{expected}$ – the mean of acquired prior ($p = 0.874$, two-tailed rank-sum test; BF01 = 3.32), **(J)** $\sigma_{expected}$ – the uncertainty in the acquired prior ($p = 0.401$, two-tailed rank-sum test; BF01 = 2.95), **(K)** $\sigma_{sensory}$ – the uncertainty of sensory likelihood ($p = 0.742$, two-tailed rank-sum test; BF01 = 2.96), **(L)** $\alpha_{prior\text{-}based}$ – prior-based lapse rate ($p = 0.024$, two-tailed rank-sum test). Red horizontal lines denote median values; blue horizontal lines denote 25th and 75th percentiles. Black dots denote individual participants, grey areas represent density of the data points. * indicates significance level at $p < 0.05$. See Suppl. Fig. 7 and 8 for the model fits of individual biases.

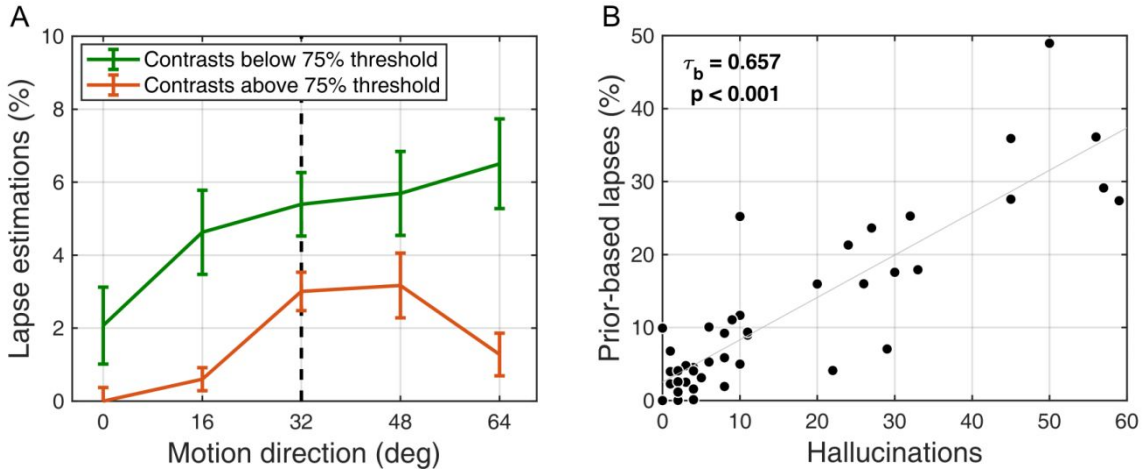



**Figure 7. Relationship between lapse estimations and hallucinations.** (**A**) The amount of lapse estimations at different stimulus contrast levels. Contrast staircase trials were split into two subsets along the 75% detection threshold, which was determined for each individual from their psychometric curves. The data then was pooled across both patient and control groups. We found that the number of lapse estimations depended on the presented contrast level with more lapse estimations being made at lower contrasts ($F(1,84) = 12.61$, $p < 0.001$). This supports our interpretation that these estimations can be interpreted, at least partly, as hallucinations on trials when the stimulus was too weak to be visible. The error bars represent within-subject errors. (**B**) Prior-based lapses and hallucinations. A strong positive correlation between prior-based lapse rate parameter (recovered via BAYES_P model) on low contrast trials and number of hallucinations on no-stimulus trials ($\tau_b = 0.657$, $p < 0.001$; Kendall's correlation) provided further support that the two phenomena are driven by the same mechanism (i.e. sampling from the prior).

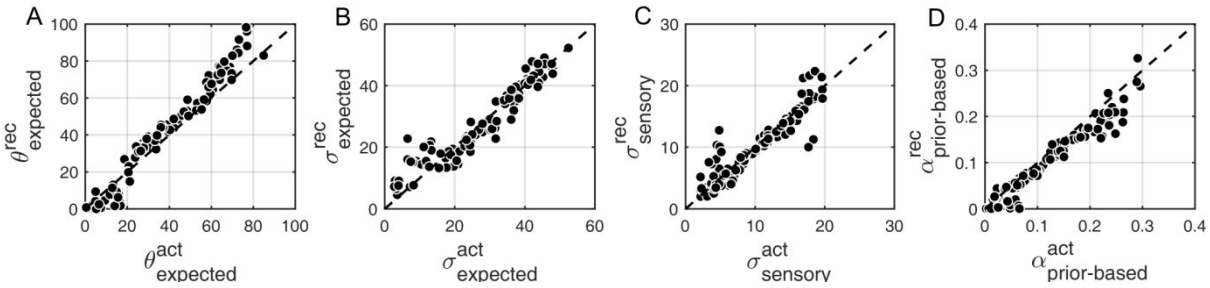



**Figure 8.** Parameter recovery with BAYES_P model. **(A)** $\theta_{expected}$ - mean of the prior expectations ($R^2$ = 0.96), **(B)** $\sigma_{expected}$ - uncertainty of the prior distribution ($R^2$ = 0.89), **(C)** $\sigma_{sensory}$ - uncertainty in the sensory likelihood ($R^2$ = 0.84), **(D)** $\alpha_{prior\text{-}based}$ – prior-based lapse rate ($R^2$ = 0.94). X-axes – actual parameters used for simulating the data (denoted with the superscript 'act'), Y-axes – recovered parameters (denoted with the superscript 'rec') from fitting the model to the simulated data. The dashed diagonal line is a reference line indicating perfect parameter recovery.

## Bibliography

Adams, R.A., Stephan, K.E., Brown, H.R., Frith, C.D., Friston, K.J., 2013. The Computational Anatomy of Psychosis. Frontiers Psychiatry 4.

American Psychiatric Association, 2000. Diagnostic and statistical manual of mental disorders (4th ed., text rev. - DSM-IV-R). Washington, DC,

Andreasen, N.C., Pressler, M., Nopoulos, P., Miller, D., Ho, B.-C., 2010. Antipsychotic Dose Equivalents and Dose-Years: A Standardized Method for Comparing Exposure to Different Drugs. Biological Psychiatry 67, 255–262.

Averbeck, B.B., Evans, S., Chouhan, V., Bristow, E., Shergill, S.S., 2011. Probabilistic learning and inference in schizophrenia. Schizophrenia Research 127, 115–122.

Chalk, M., Seitz, A.R., Seriès, P., 2010. Rapidly learned stimulus expectations alter perception of motion. Journal of Vision 10(8), 2.

Corlett, P.R., Frith, C.D., Fletcher, P.C., 2009a. From drugs to deprivation: a Bayesian framework for understanding models of psychosis. Psychopharmacology 206, 515–530.

Corlett, P.R., Simons, J.S., Pigott, J.S., Gardner, J.M., Murray, G.K., Krystal, J.H., Fletcher, P.C., 2009b. Illusions and delusions: relating experimentally-induced false memories to anomalous experiences and ideas. Frontiers Behavioural Neuroscience 3, 53.

Crawford, T.J., Hamm, J.P., Kean, M., Schmechtig, A., Kumari, V., Anilkumar, A.P., Ettinger, **U.**, 2010. The perception of real and illusory motion in schizophrenia. Neuropsychologia 48, 3121–3127.

Danion, J.-M., Meulemans, T., Kauffmann-Muller, F., Vermaat, H., 2001. Intact Implicit Learning in Schizophrenia. American Journal of Psychiatry 158, 944–948.

Daunizeau, J., Adam, V, Rigoux, L., 2014. VBA: a probabilistic treatment of nonlinear models for neurobiological and behavioural data. PLoS Comp Biol 10(1): e1003441.

Dima, D., Dietrich, D.E., Dillo, W., Emrich, H.M., 2010. Impaired top-down processes in schizophrenia: A DCM study of ERPs. NeuroImage 52, 824–832.

Dima, D., Roiser, J.P., Dietrich, D.E., Bonnemann, C., Lanfermann, H., Emrich, H.M., Dillo, W., 2009. Understanding why patients with schizophrenia do not perceive the hollow-mask illusion using dynamic causal modelling. NeuroImage 46, 1180–1186.

Evans, S., Almahdi, B., Sultan, P., Sohanpal, I., Brandner, B., Collier, T., Shergill, S.S., Cregg, R., Averbeck, B.B., 2012. Performance on a probabilistic inference task in healthy subjects receiving ketamine compared with patients with schizophrenia. Journal of

Psychopharmacology 26, 1211–1217.

Fiser, J., Berkes, P., Orbán, G., Lengyel, M., 2010. Statistically optimal perception and learning: from behavior to neural representations. Trends Cognitive Science (Regul Ed) 14, 119–130.

Fletcher, P.C., Frith, C.D., 2009. Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. Nature Reviews Neuroscience 10, 48–58.

Freeman, D., Pugh, K., Garety, P., 2008. Jumping to conclusions and paranoid ideation in the general population. Schizophrenia Research 102, 254–260.

Friston, K.J., 2005. Hallucinations and perceptual inference. Behavioral and Brain Sciences 28, 764–766.

Friston, K.J., 2010. The free-energy principle: a unified brain theory? Nature Review Neuroscience 11, 127–138.

Friston, K.J., 2012. The history of the future of the Bayesian brain. NeuroImage 62, 1230–1233.

Friston, K., Brown, H.R., Siemerkus, J. and Stephan, K.E., 2016. The dysconnection hypothesis (2016). Schizophrenia research, 176(2), pp.83-94.

Frith, C.D., Friston, K.J., 2012. False perceptions and false beliefs: understanding schizophrenia. Working Group on Neurosciences and the Human Person: New Perspectives on Human Activities, The Pontifical academy of Sciences 8–10.

García-Pérez, M.A., 1998. Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties. Vision Research. 38, 1861–1881.

Garety, P., Joyce, E.M., Jolley, S., Emsley, R., Waller, H., Kuipers, E., Bebbington, P., Fowler, D., Dunn, G., Freeman, D., 2013. Neuropsychological functioning and jumping to conclusions in delusions. Schizophrenia Research.

Garety PA, Freeman D (2013): The past and future of delusions research: from the inexplicable to the treatable. British Journal of Psychiatry 203: 327–333.

Gekas, N., Chalk, M., Seitz, A.R., Series, P., 2013. Complexity and specificity of experimentally-induced expectations in motion perception. Journal of Vision 13.

Gold, J.M., Hahn, B., Strauss, G.P., Waltz, J.A., 2009. Turning it upside down: areas of preserved cognitive function in schizophrenia. Neuropsychological Reviews 19, 294–311.

Horton, H.K., Silverstein, S.M., 2011. Visual Context Processing Deficits in Schizophrenia: Effects of Deafness and Disorganization. Schizophrenia Bulletin 37, 716–726.

Huq, S.F., Garety, P.A., Hemsley, D.R., 1988. Probabilistic judgments in deluded and non-deluded subjects. Quarterly Journal of Experimental Psychology A 40, 801–812.

Jardri, R., Denève, S., 2013. Circular inferences in schizophrenia. Brain.

Jardri, R., Duverne, S., Litvinova, A.S. and Denève, S., 2017. Experimental evidence for

circular inference in schizophrenia. Nature communications, 8, p.14218.

Kaas, R.E., Raftery A.E., 1995. Bayes Factors. Journal of the American Statistical Association 90(430), 773-795.

Karvelis, P., Seitz, A.R., Lawrie, S.M. and Seriès, P., 2018. Autistic traits, but not schizotypy, predict increased weighting of sensory information in Bayesian visual integration. eLife, 7, p.e34115.

Keane, B.P., Silverstein, S.M., Wang, Y., Papathomas, T.V., 2013. Reduced depth inversion illusions in schizophrenia are state-specific and occur for multiple object types and viewing conditions. Journal of Abnormal Psychology 122, 506–512.

Kéri, S., Kelemen, O., Szekeres, G., Bagóczky, N., Erdélyi, R., Antal, A., Benedek, G., Janka, Z., 2000. Schizophrenics know more than they can tell: probabilistic classification learning in schizophrenia. Psychological Medicine 30, 149–155.

Knill, D.C., Pouget, A., 2004. The Bayesian brain: the role of uncertainty in neural coding and computation. Trends Neuroscience 27, 712–719.

Kok, P., Brouwer, G.J., van Gerven, M.A.J., de Lange, F.P., 2013. Prior Expectations Bias Sensory Representations in Visual Cortex. Journal of Neuroscience 33, 16275–16284.

Laquitaine S., Gardner J., 2018. A Switching Observer for Human Perceptual Estimation. Neuron 97(2), pp.462-474

Leucht, S., Samara, M., Heres, S., Patel, M.X., Furukawa, T., Cipriani, A., Geddes, J. and Davis, J.M., 2015. Dose equivalents for second-generation antipsychotic drugs: the classical mean dose method. Schizophrenia bulletin, 41(6), pp.1397-1402.

Levitt, H., 1971. Transformed up-down methods in psychoacoustics. Journal of Acoustical Society of America. 49, Suppl 2:467–.

Mardia, K.V., 1972. Statistics of directional data, Academic Press, Inc. (London) LTD.

Marvel, C.L., Schwartz, B.L., Howard, D.V., Howard, J.H., 2005. Implicit learning of non-spatial sequences in schizophrenia. Journal of the International Neuropsychological Society: JINS 11, 659–667.

Nelson, H. E., & Willison, J. R., 1991. The National Adult Reading Test (NART): test manual (2nd ed.). Windsor: NFER-Nelson

Notredame C.E., Pins D., Denève S., 2014: What Visual illusions teach us about schizophrenia. Frontiers in Integrative physiology

Nuechterlein, K.H., 1977. Reaction time and attention in schizophrenia. Schizophrenia Bulletin 3.

Petzschner, F.H., Glasauer S., Stephan, K.E., 2015. A Bayesian perspective on magnitude

estimation. Trends Cogn Sci. May; 19(5):285-93.

Palminteri, S., Wyart, V., and Koechlin, E. (2017). The importance of falsification in computational cognitive modeling. Trends in Cognitive Sciences, 21(6): 425-433.

Powers, A.R., Mathys, C. and Corlett, P.R., 2017. Pavlovian conditioning–induced hallucinations result from overweighting of perceptual priors. Science, 357(6351), pp.596-600.

Rigoux, L., Stephan, K.E., Friston, K.J., Daunizeau, J., 2014. Bayesian model selection for group studies – revisited. Neuroimage, 84:971-985.

Sato, Y. and Kording, K.P., 2014. How much to trust the senses: Likelihood learning. Journal of Vision, 14(13), pp.13-13.

Seriès, P., Seitz, A.R., 2013. Learning what to expect (in visual perception). Frontiers in human neuroscience 7.

Seriès P., Stocker A.A., Simoncelli E.P., 2009. Is the homunculus "aware" of sensory adaptation? Neural Comput.; 21(12):3271-304.

Schmack, K., de Castro, A.G.C., Rothkirch, M., Sekutowicz, M., Rössler, H., Haynes, J.D., Heinz, A., Petrovic, P. and Sterzer, P., 2013. Delusions and the role of beliefs in perceptual inference. Journal of Neuroscience, 33(34), pp.13701-13712.

Schmack, K., Schnack, A., Priller, J. and Sterzer, P., 2015. Perceptual instability in schizophrenia: Probing predictive coding accounts of delusions with ambiguous stimuli. Schizophrenia Research: Cognition, 2(2), pp.72-77.

Schmack, K., Rothkirch, M., Priller, J. and Sterzer, P., 2017. Enhanced predictive signalling in schizophrenia. Human brain mapping, 38(4), pp.1767-1779.

Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. CRC Press.

Silverstein S.M., Keane B.P., 2011a: Vision science and schizophrenia research: toward a re-view of the disorder. Editors' introduction to special section. Schizophrenia Bulletin 37: 681–689.

Silverstein S.M., Keane B,P., 2011b: Perceptual organization impairment in schizophrenia and associated brain mechanisms: review of research from 2005 to 2010. Schizophrenia Bulletin 37: 690–699.

Skottun, B.C., Skoyles, J.R., 2007. Contrast sensitivity and magnocellular functioning in schizophrenia. Vision Research. 47, 2923–2933.

Speechley, W.J., Whitman, J.C., Woodward, T.S., 2010. The contribution of hypersalience to the “jumping to conclusions” bias associated with delusions in schizophrenia. Journal of Psychiatry Neuroscience 35, 7–17.

Stephan K.E., Bach D.R., Fletcher P.C., Flint J., Frank M.J., Friston K.J., Heinz A., Huys Q.J.M., Owen M.J., Binder E.B., Dayan P., Johnstone E.C., Meyer-Lindenberg A., Montague P.R., Schnyder U., Wang X.J., Breakspear M., 2016. Charting the landscape of priority problems in psychiatry, part 1: classification and diagnosis. Lancet Psychiatry 3(1): 77-83

Sterzer, P., Adams R. A., Fletcher, P., Frith, C., Lawrie, S. M., Muckli L., Petrovic, P., Uhlhaas, P., Voss, M., Corlett, P.R., 2018. The predictive coding account of psychosis. Biological Psychiatry 84:634-643.

Tschacher, W., Schuler, D., Junghan, U., 2006. Reduced perception of the motion-induced blindness illusion in schizophrenia. Schizophrenia Research 81, 261–267.

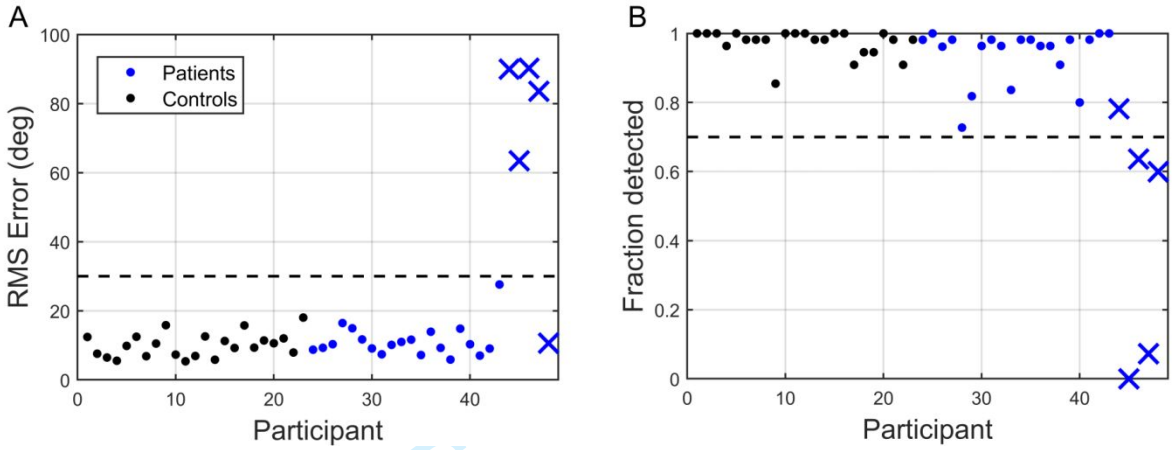
Valton, V., Romaniuk, L., Steele, J.D., Lawrie, S. and Seriès, P., 2017. Comprehensive review: Computational modelling of schizophrenia. Neuroscience & Biobehavioral Reviews. 83:631-646

Wand, M.P., Jones, M.C., 1994. Kernel Smoothing. CRC Press.

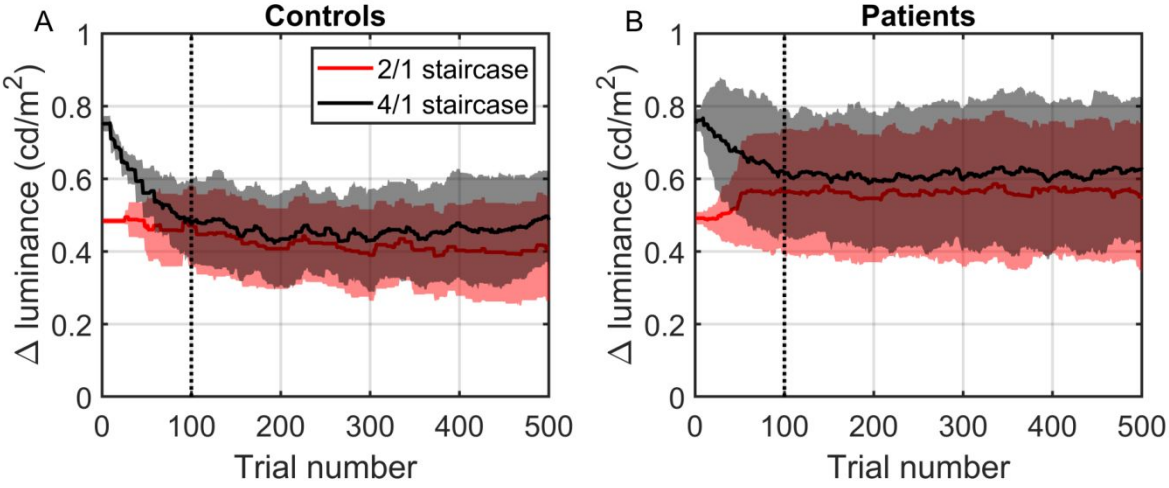
Wechsler, D. (1999). Wechsler Abbreviated Scale of Intelligence. The Psychological Corporation: Harcourt Brace & Company. New York, NY.

# Acquisition of visual priors and induced hallucinations in chronic schizophrenia: Supplementary information

**V. Valton, P. Karvelis, K. L. Richards, A. R. Seitz, S. M. Lawrie, P. Seriès**




**Supplementary Figure 1:** Performance on high contrast trials. **(A)** Root mean square error of estimations. **(B)** Fraction of trials on which the stimulus was detected. The dots represent included participants; the cross marks represent excluded participants. Dashed lines denote inclusion criteria (30 degree estimation and 70% detection).

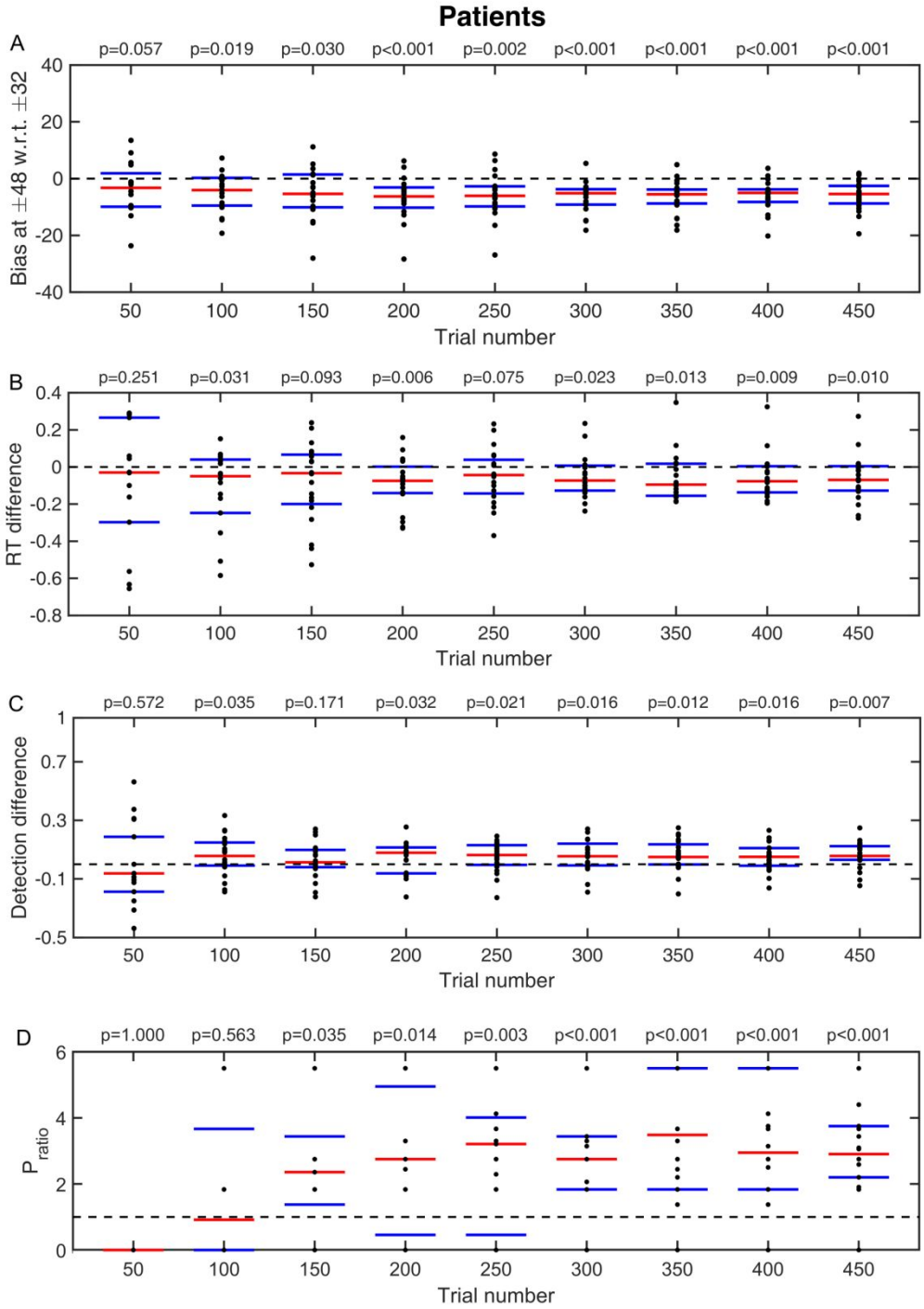



**Supplementary Figure 2:** Convergence of 2/1 and 4/1 staircase luminance levels. **(A)** Controls, **(B)** Patients. Both groups reached convergence after ~100 trials. These trials were excluded from further data analysis.

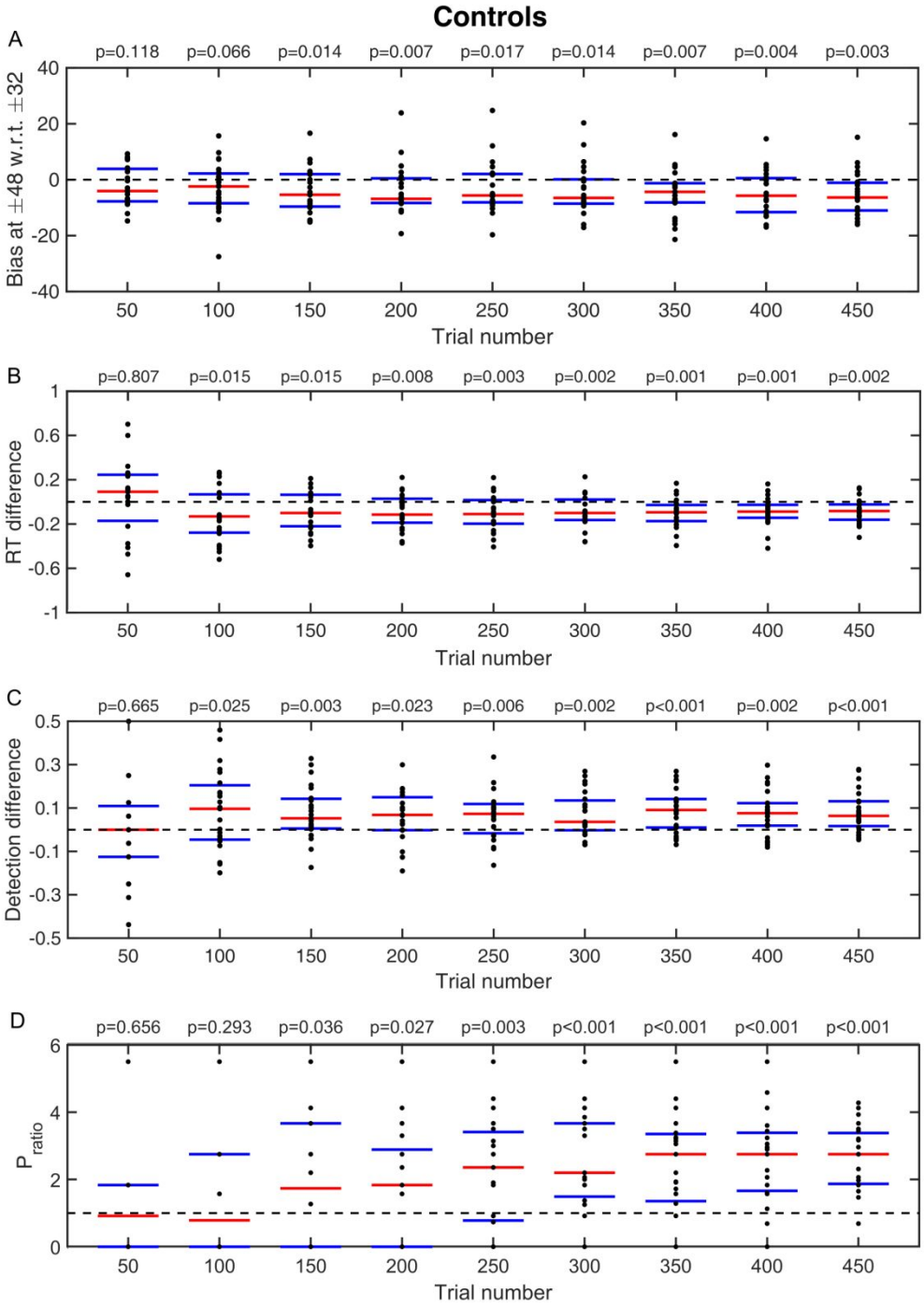
## Emergence of prior effects

We wanted to determine when the acquired expectations started to have a significant effect on performance and whether this was the same for both groups. First and most importantly, we wanted to know when the estimations on low contrast trials became biased towards ±32°. To do so, we computed cumulative moving averages at every 50 trials for the bias at ±48° with respect to bias at ±32°. Next, we looked at estimation reaction times (RT) on low contrast trials and compared mean RT of each individual at ±32° with mean RT at all other directions. Similarly, we looked at the average detection performance on low contrast trials and compared the fraction of trials in which stimulus was detected at ±32° with the mean fraction detected over all other presented directions.

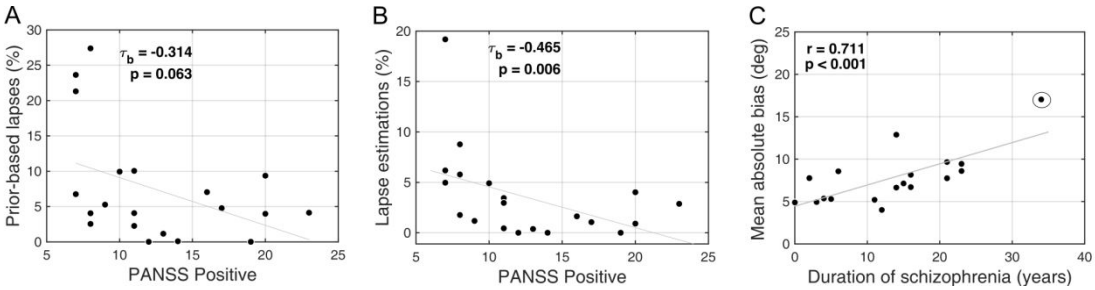
We found that both patients and controls showed signs of very rapid acquisition of the priors. For the bias to become statistically significant, it took less than 100 trials for patients (**Supplementary Fig. 3A**) and less than 150 trials for controls (**Supplementary Fig. 4A**). Similarly, both patients and controls became significantly faster and significantly better at detecting stimulus moving at ±32° within the first 100 trials (**Supplementary Fig. 3B, C and 4B, C respectively**).




**Supplementary Figure 3:** Emergence of prior effects in patient group. **(A)** Cumulative moving averages of bias at ±48° with respect to bias at ±32°. **(B)** Cumulative moving averages of median differences between estimation RTs at ±32° and RTs at all other directions. **(C)** Cumulative moving averages of median differences between fraction of detected stimuli at ±32° and fraction detected at all other directions. **(D)** Cumulative moving averages of the probability ratio of hallucinating predominantly around ±32 on no-stimulus trials. Red bars indicate median values and blue bars indicate 25th and 75th percentiles. p-values are denoted above each plot (one-tailed Wilcoxon signed rank test).




**Supplementary Figure 4:** Emergence of prior effects in control group. **(A)** Cumulative moving averages of bias at ±48° with respect to bias at ±32°. **(B)** Cumulative moving averages of median differences between estimation RTs at ±32° and RTs at all other directions. **(C)** Cumulative moving averages of median differences between fraction of detected stimuli at ±32° and fraction detected at all other directions. **(D)** Cumulative moving averages of the probability ratio of hallucinating predominantly around ±32 on no-stimulus trials. Red bars indicate median values and blue bars indicate 25th and 75th percentiles. p-values are denoted above each plot (one-tailed Wilcoxon signed rank test).



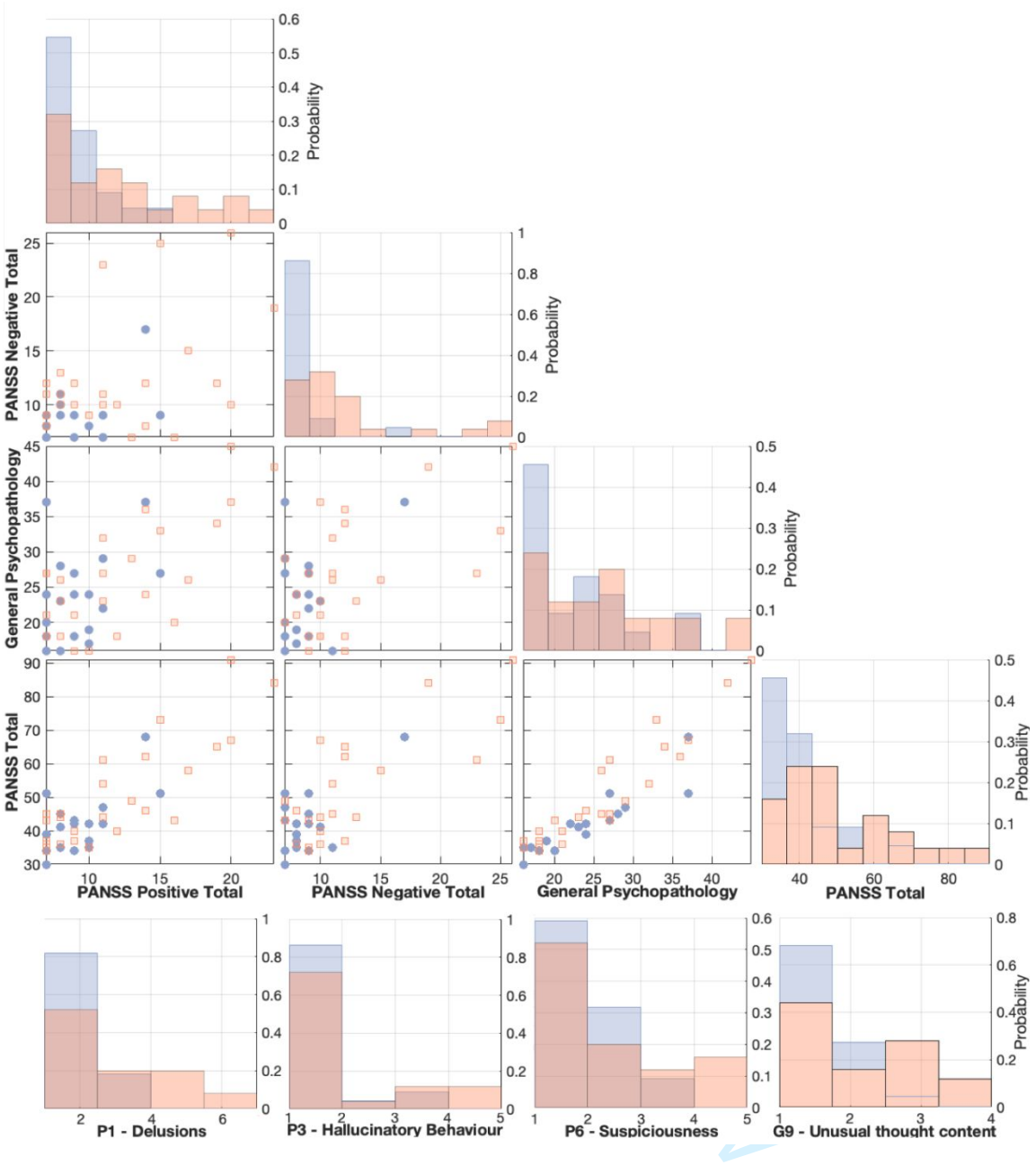

**Supplementary Figure 5:** Exploratory analysis. (**A**) Prior-based lapses (estimated via BAYES_P) as a function of PANSS positive score ($\tau_b$ = -0.314, p = 0.063; Kendall's correlation). (**B**) Lapse estimations (determined via Eq.(1)) and PANSS positive score ($\tau_b$ = -0.465, p = 0.006; Kendall's correlation). (**C**) Mean absolute bias as a function of duration of illness. The plot shows a positive relationship between the duration of illness and the amount of bias exhibited during the task (r = 0.711, p < 0.001; Pearson's correlation. The correlation remained highly significant after excluding the outlier (circled) (r = 0.523, p = 0.003; Pearson's correlation).

## Detailed clinical characteristics

The demographic and clinical characteristics of the included participants can be seen in Table S1 and Supplementary Figure 6. Patients had relatively low levels of symptoms on the PANSS, but ongoing functional impairment. Most were assigned a PANSS rating of 1 (absent), 2 (minimal/questionable) or 3 (mild) for delusions, hallucinatory behaviour, suspiciousness/persecution, and unusual thought content, reflecting the clinically stable state of the psychosis group. Except for one patient, all had been taking anti-psychotic medication at the time of testing (atypical, *n* = 18; typical, *n* = 1). Nine were also taking anti-depressants and two were taking sodium valproate.

| | Patients (n=20) | Controls (n=23) | *Statistic* | p-value |
|---|---|---|---|---|
| Delusions (P1) | 2.60 (1.90) | 1.48 (0.79) | *W* = 158 | < .05 |
| Hallucinatory behaviour (P3) | 1.85 (1.31) | 1.22 (0.60) | *W* = 175 | .07 |
| Suspiciousness/Persecution (P6) | 2.00 (1.26) | 1.61 (0.84) | *W* = 198 | .40 |
| Unusual thought content (G9) | 2.05 (1.10) | 1.35 (0.57) | *W* = 150 | < .05 |

**Table S1.** Demographic and detailed Clinical Characteristics (mean and standard deviation in parentheses). P# refer to specific items/symptoms of the PANSS (Positive and Negative Symptom Scale). For all measures Wilcoxon rank-sum tests (normal approximation with continuity correction) were used.




**Supplementary Figure 6:** Histogram and pair-plots for general PANSS subscales, and histograms for items P1, P3, P6 and G9 of the PANSS. Blue data represents controls, while red data represents patients. The patient group appears relatively homogeneous and stabilised with respect to symptom severity when compared to the control group.

## Normality and Non-parametric testing

Despite ANOVAs frequent robustness in this respect (Glass et al., 1972; Harwell et al., 1992; Lix et al., 1996), we were concerned that our results may have depended on normality assumptions. We ran normality tests among repeated measures for each group separately, as well as for each group averaged across angles (Table S2).

| Angle | Group | Bias | | Variability | | Lapses | | RT | | Detection | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | SW Stat. | p-value | SW Stat. | p-value | SW Stat. | p-value | SW Stat. | p-value | SW Stat. | p-value |
| 0° | Patients | .964 | .616 | .944 | .290 | .236 | **<.001** | .963 | .602 | .964 | .622 |
| | Controls | .953 | .337 | .929 | .103 | .486 | **<.001** | .968 | .632 | .960 | .472 |
| ±16° | Patients | .914 | .075 | .931 | .163 | .530 | **<.001** | .963 | .602 | .970 | .757 |
| | Controls | .967 | .614 | .953 | .341 | .705 | **<.001** | .984 | .966 | .861 | **.004** |
| ±32° | Patients | .899 | **.039** | .883 | **.020** | .906 | .053 | .831 | **.003** | .927 | .133 |
| | Controls | .904 | **.030** | .915 | .053 | .860 | **.004** | .951 | .300 | .939 | .170 |
| ±48° | Patients | .839 | **.003** | .951 | .385 | .505 | **<.001** | .959 | .527 | .982 | .956 |
| | Controls | .922 | .074 | .866 | **.005** | .847 | **.002** | .953 | .340 | .963 | .531 |
| ±64° | Patients | .702 | **<.001** | .897 | **.036** | .740 | **<.001** | .958 | .503 | .934 | .183 |
| | Controls | .790 | **<.001** | .901 | **.027** | .861 | **.004** | .963 | .519 | .972 | .731 |
| Mean | Patients | .583 | **<.001** | .948 | .340 | .682 | **<.001** | .894 | **.033** | .919 | .096 |
| | Controls | .743 | **<.001** | .937 | .153 | .886 | **.015** | .967 | .626 | .933 | .116 |

**Table S2.** Shapiro-Wilk normality tests for estimation bias, variability, lapses, reaction time and detection rate at each stimulus motion direction (0°, ±16°, ±32° ±48° ±64°) and averaged across motion directions (mean) for each group (patients and controls). Degrees of freedom are df = 20 for patients and df = 23 for controls for all angles. Normality violations below $\alpha = 0.05$ significance level are highlighted in bold.



Some deviations from normality were identified for estimation biases and lapses. For these measures, we re-ran group comparisons using a two-sided Wilcoxon rank-sum test (that does not require normality assumptions). To our knowledge there is no well-established non-parametric equivalent to the 2-way mixed ANOVA, therefore, we focused on verifying group comparisons on the mean values of bias and lapses. Most importantly, despite being more conservative than parametric tests, the non-parametric rank-sum test results were found to be in complete agreement with the ANOVA test results: patients exhibited significantly fewer lapse estimations than controls ($W = 324$, $p = 0.005$, two-sided rank-sum test) and there were no differences in the mean absolute bias between the groups ($W = 399$, $p = 0.324$, two-sided rank-sum test). Furthermore, we also report Bayesian statistical analyses on all of the measures reported, which also confirm the frequentist parametric and non-parametric analyses findings. Therefore, we feel confident that the findings reported in the manuscript do not suffer from the normality assumptions being violated.

## Modelling

### Variant Bayes models

The model 'Bayes' is similar to the model 'Bayes_P' described in the main text, but has uniform lapse estimations instead of 'prior-based' lapse estimations, such that:

$$p(\theta_{estimate}|\theta_{perceived}) = (1 - \alpha) \cdot V(\theta_{perceived}, \sigma_{motor}) + \alpha/2\pi$$

The model 'Bayes_var' is similar to the model 'Bayes', except that to account for the possibility of exposure effect to sensory uncertainty, we allow sensory uncertainty $\sigma_{sensory}$ to vary for each of the following angles 0°, ±16°, ±32°, ±48°, ±64°, resulting in five $\sigma_{sensory}$ free parameters.

The model 'Bayes_varmin' is similar to the model 'Bayes', except that we allow $\sigma_{sensory}$ to vary only for the most presented angles ±32°, resulting in two $\sigma_{sensory}$ free parameters ($\sigma_{sensory}$ at ±32°, and $\sigma_{sensory}$ at all other angles).

We did not include models without the lapse parameter ($\alpha_{prior\text{-}based}$ in the case of model 'Bayes_P', or $\alpha$ in the case of model 'Bayes') in the current report because they performed much worse (~3000 BIC units larger) than the worst performing model in the current model comparison (ADD2).

### Response strategy models

We wanted to control for the possibility that the task behaviour might be explained by simple behavioural strategies that do not involve Bayesian integration. This class of models assumed that participants did not combine their expectations with sensory information, but relied on either of them alone on any given trial.

The first model, 'ADD1', assumed that estimations derived from prior expectations were simply sampled from a learnt prior distribution, $p_{expected}(\theta)$, which was parameterized as in Eq (4). However, on trials when participants did perceive motion direction, it was based solely on the sensory input, $p_{sensory}(\theta_{sensory}|\theta_{actual}) = V(\theta_{actual}, \sigma_{sensory})$.

Putting together the estimations derived from sensory input and the ones derived from learnt expectations, and the possibility of random estimations, the average distribution of estimation responses for a single participant is:

$$p(\theta_{estimate}|\theta_{actual}) = (1-\alpha)\cdot[(1-a(\theta))\cdot p_{sensory}(\theta_{sensory}|\theta_{actual}) + a(\theta)\cdot p_{expected}(\theta_{estimate})] * V(0,\sigma_{motor}) + \alpha, \quad (9)$$

where $a(\theta)$ determines the proportion of trials in which participants sampled from the acquired prior, $p_{expected}(\theta)$; asterisk ($*$) denotes convolution. The resulting 'ADD1' model has 9 free parameters ($\theta_{expected}$, $\sigma_{expected}$, $a(\theta)$ (which can take a different value for each of the 5 angles: $0^\circ$, $\pm16^\circ$, $\pm32^\circ$, $\pm48^\circ$, $\pm64^\circ$), $\sigma_{sensory}$ and $\alpha$).

The second model, 'ADD2', was the same as 'ADD1' except that it had more complex strategy for trials when participants relied on the prior: instead of sampling from the complete acquired prior distribution ranging from $-180^\circ$ to $+180^\circ$ (Eq. (4)), they sampled from only one half of it, negative ($-180°$ to $0°$) or positive ($0°$ to $+180°$), depending on which side of the distribution the actual stimulus occurred on:

$$p_{expectedN}(\theta) = V(-\theta_{expected}, \sigma_{expected}) \quad (10)$$

$$p_{expectedP}(\theta) = V(\theta_{expected}, \sigma_{expected}) \quad (11)$$

Incorporating this into the distribution of estimation responses results in:

$$p(\theta_{estimate}|\theta_{actual}) = (1-\alpha)\cdot[((1-a(\theta)-b(\theta))\cdot p_{sensory}(\theta_{senssory}|\theta_{actual}) + a(\theta)\cdot p_{expectedN}(\theta_{estimate}) + b(\theta)\cdot p_{expectedP}(\theta_{estimate})] * V(0,\sigma_{motor}) + \alpha, \quad (12)$$

where asterisk ($*$) denotes convolution; $a(\theta)$ and $b(\theta)$ determine the proportion of trials in which participants sample from either negative or positive parts of the prior distribution, respectively.

Finally, we also considered two variations of the 'ADD1' and 'ADD2' models. These were identical to 'ADD1' and 'ADD2' except from setting $\sigma_{expected}$ to zero (i.e. no uncertainty); that is, on trials when perceptual estimates were derived only from expectations, they were

equal to the mode of the learnt distribution. These models are referred to as 'ADD1_m' and 'ADD2_m'.

**Parameter estimation**

We used the performance in trials with the highest contrast level to estimate motor noise, $\sigma_{motor}$, for each individual. We assumed that at this level sensory uncertainty was close to zero ($\sigma_{sensory} \approx 0$). The motor noise was determined by fitting estimation responses at the highest contrast level to the distribution in Eq. (2) using the actual motion direction, $\theta_{actual}$, as the mean. The estimated motor noise was used in all the subsequent model fitting as a fixed parameter.

The rest of the free parameters of each model were estimated by fitting the response data from the two staircased contrast levels (~200 trials per participant). For each model with a set of free parameters M, we computed the probability distribution $p(\theta_{estimate}|\theta_{actual}; M)$ of making an estimate $\theta_{estimate}$ given the actual stimulus direction $\theta_{actual}$. For the response strategy models, by definition, the $p(\theta_{estimate}|\theta_{actual}; M)$ corresponds to average behaviour in the task (Equations 9 and 12). Bayesian models, on the other hand, explicitly model trial-to-trial variability in the posterior estimate, which in our case is the mean of the posterior (Eq. (6)). To relate this to the behavioural data we built a distribution of 1,000 samples for each presented angle (where each sample is the mean of the posterior obtained via Eq. (6) and perturbed by motor noise via Eq. (7) or (8)).

The parameters were estimated by maximizing the fit of the log likelihood function for the experimental data for each participant individually:

$$M = \underset{M}{\operatorname{argmax}} \left[\sum_i^n \log\left(p(\theta_{estimate} = \theta_{i,\,data}|\theta_i)\right)\right], \qquad (13)$$

where $\theta_{i,data}$ is participant's estimation response, $\theta_i$ is the actual presented motion direction on the ith trial and n is the number of trials. The maximum likelihood was found using *fminsearchbnd* function in Matlab, by minimizing negative log-likelihood. Parameters α, a(θ) and b(θ) were bounded between 0 and 1, while $\theta_{expected}$, $\sigma_{expected}$ and $\sigma_{sensory}$ were bounded from 0 to ∞. To reduce the possibility of convergence at a local maxima we constructed a grid of initial $\sigma_{expected}$ and $\sigma_{sensory}$ parameter values covering the range as found in previous

studies ($\sigma_{expected}$ from 10° to 25° in 3° increments and $\sigma_{sensory}$ from 5° to 15° in 3° increments, which resulted in 22 different initializations). A set of parameters with the largest log-likelihood was selected as the best fit.

**Model Comparison**

To compare the model fits we used Bayesian Information Criterion (BIC), which approximates the log of model evidence (e.g., see Burnham and Anderson, 2004):

$$-2 \cdot \log(P(D|M)) \approx BIC = -2 \cdot \log(P(D|M, \Theta)) + k \cdot \log(n) \quad , \qquad (14)$$

where M is model, D is observed data and P (D|M, Θ) is the likelihood of generating the experimental data given the most likely set of parameters, Θ; k is the number of model parameters and n is the number of data points (or equivalently, the number of trials). BIC evaluates the model by how it fits the data by also penalizing for the number of parameters (i.e. model complexity) to avoid over-fitting. Lower BIC score indicates a better model. We also ran a random effect Bayesian model selection analysis for group studies (Rigoux et al., 2014). We used the VBA Matlab toolbox (Daunizeau et al., 2014) to perform this analysis, and used participant-level BIC as an approximation of the log-model evidence.

**Parameter recovery**

To test the reliability of the parameter estimates of our winning model we performed parameter recovery. This allowed us to simultaneously test whether parameters are identifiable (e.g., whether likelihood and prior uncertainty is not correlated and can be distinguished) and whether having ~200 trials (the amount of low contrast trials in our data) for data fitting and using maximum likelihood estimation are sufficient to give reliable results.

First, we generated 100 sets of parameters (i.e. 100 synthetic individuals) by randomly sampling each parameter from a Gaussian distribution that had a mean and variance as the parameter estimates from the collected participant data. Second, for each set of parameters we simulated data for 200 trials with the winning model by randomly sampling from the estimation probability distribution, which, as for the behavioural data, was built from a 1000 posterior means (Eq. (6)), each perturbed by motor noise (Eq. (8)) Finally, we fitted the winning model to the simulated data. To evaluate the goodness of recovered parameters we

computed the coefficient of determination ($R^2$) for a linear regression, which quantified how well the actual parameters predicted the recovered ones.

## Power calculations

**Power calculations (a-priori): Detecting effect of the prior on behaviour:**

We performed power calculations using an independent dataset from a previously published study that used the exact same task design but had 2 sessions of 850 trials (Chalk *et al*., 2010). We used the first 567 trials of the first session for the power analysis, to conform to the same trial structure present in the current study.

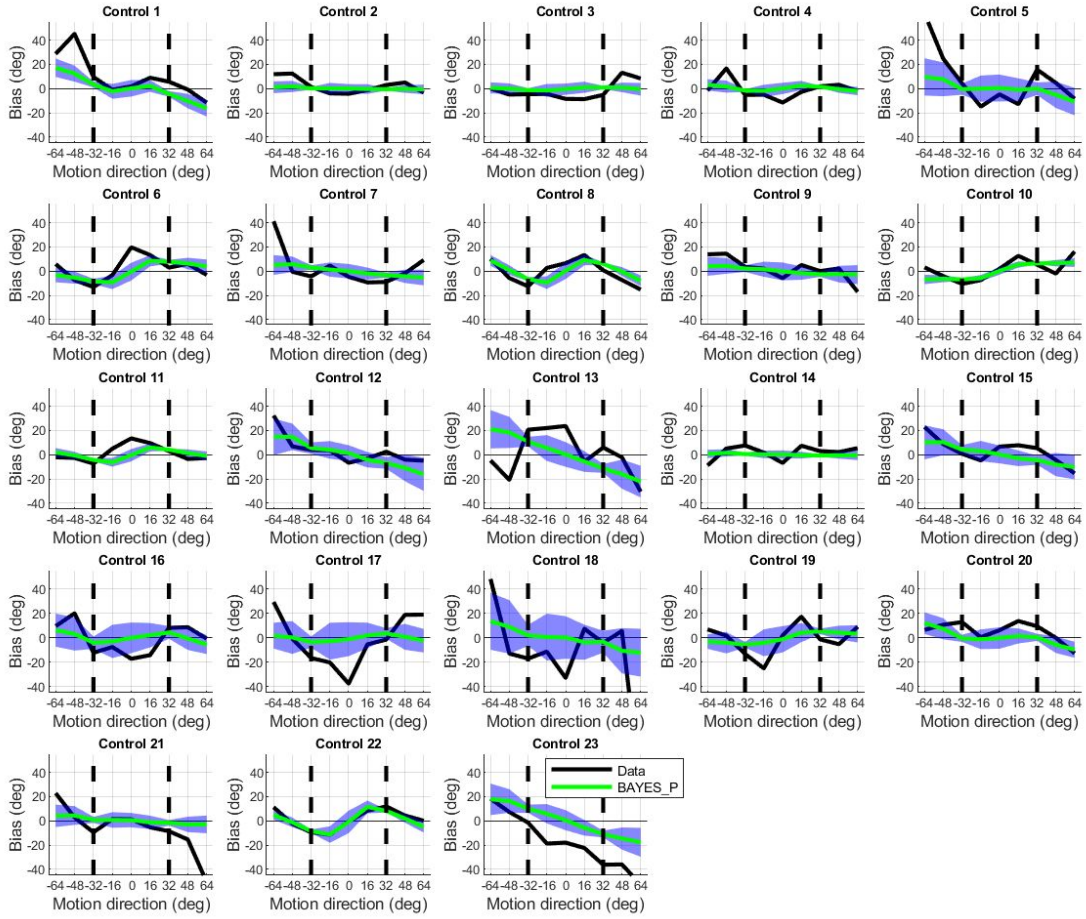
'Bias' and 'Detection rate' were selected as the behavioural measures relevant for determining the effect of the prior on behaviour. Given these parameters, power calculations for 'Bias' and 'Detection rate' revealed that 18-20 subjects are required respectively to detect an effect with 80% power.

**Power calculations (a-priori): Detecting group differences:**

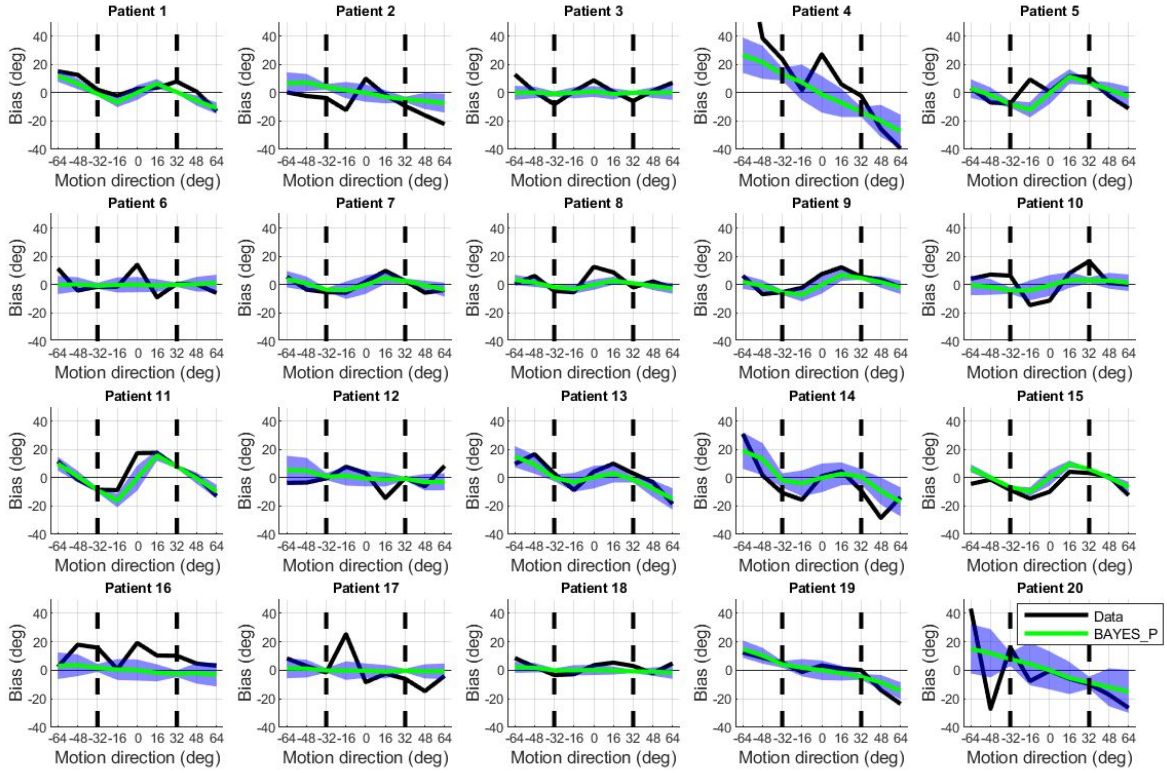
To our knowledge, the current task design has never been tested in a chronic Schizophrenia sample. We therefore turned to the literature to estimate the effect size between patients and controls in motion perception (Chen, 2011) and illusions (Dima et al., 2009; Dima et al., 2011). When Cohen's d was not available from the literature, we used reported statistics and converted them to Cohen's d prior to running power analyses. All a-priori analyses used a type I error rate 'α' of 0.05.

We determined that 20 participants in each group should allow us to detect an effect size of at least 0.9 (Cohen's d) with 80% power, which is a smaller effect than previous effect sizes reported in motion perception (Chen et al., 1999a; Chen et al., 1999b), perceptual illusions (Dima et al., 2009; Dima et al., 2011), or cognition generally (Fioravanti *et al.,* 2012).

Based on this analysis we aimed to recruit approximately 25 subjects per group, which would permit to reliably detect the effect of: the prior on perception, as well as differences in motion perception or illusions between groups, even accounting for a potentially large (e.g. 20%) exclusion rate based on task performance.




**Supplementary Figure 7: BAYES_P bias fits for each subject in the control group**. Black solid line is the behavioural data, green line is the mean bias and blue shaded area is the standard deviation of the bias as predicted by BAYES_P. The model predictions were generated using the parameter values obtained from maximum likelihood estimation. The model was simulated to produce the same number of responses as in the empirical data of each subject. For each subject, the simulation was repeated 30 times to obtain a distribution of biases, which then was used to calculate the mean and the standard deviation of the bias.




**Supplementary Figure 8: BAYES_P bias fits for each subject in the patient group**. Black solid line is the behavioural data, green line is the mean bias and blue shaded area is the standard deviation of the bias as predicted by BAYES_P. The model predictions were generated using the parameter values obtained from maximum likelihood estimation. The model was simulated to produce the same number of responses as in the empirical data of each subject. For each subject, the simulation was repeated 30 times to obtain a distribution of biases, which then was used to calculate the mean and the standard deviation of the bias.

**References**

Burnham, K.P. and Anderson, D.R., 2004. Multimodel inference: understanding AIC and BIC in model selection. Sociological methods & research, 33(2), 261-304.

Chalk, M., Seitz, A.R., Seriès, P., 2010. Rapidly learned stimulus expectations alter perception of motion. Journal of Vision 10, 2.

Chen, Y., Nakayama, K., Levy, D.L., Matthysse S., Holzman P.S. 1999. Psychophysical isolation of a motion-processing deficit in schizophrenics and their relatives and its association with impaired smooth pursuit. Proc Natl Acad Sci U S A. 96(8): 4724–4729.

Chen, Y., Palafox, G. P., Nakayama, K., Levy, D. L., Matthysse, S., Holzman P. S. 1999. Motion perception in schizophrenia. Arch Gen Psychiatry. 1999 Feb; 56(2):149-54.

Daunizeau, J., Adam, V. and Rigoux, L., 2014. VBA: a probabilistic treatment of nonlinear models for neurobiological and behavioural data. *PLoS Computational Biology*, *10*(1), p.e1003441.

Dima, D., Dietrich, D.E., Dillo, W., Emrich, H.M.. 2010. Impaired top-down processes in schizophrenia: A DCM study of ERPs. NeuroImage 52, 824–832.

Dima, D., Roiser, J.P., Dietrich, D.E., Bonnemann, C., Lanfermann, H., Emrich, H.M., Dillo, W., 2009. Understanding why patients with schizophrenia do not perceive the hollow-mask illusion using dynamic causal modelling. NeuroImage 46, 1180–1186.

Fioravanti M, Bianchi V, Cinti ME. Cognitive deficits in schizophrenia : an updated meta-analysis of the scientific evidence. BMC Psychiatry. 2012; 12:64

Glass, G.V., P.D. Peckham, and J.R. Sanders. 1972. Consequences of failure to meet assumptions underlying fixed effects analyses of variance and covariance. Rev. Educ. Res. 42: 237-288.

Harwell, M.R., E.N. Rubinstein, W.S. Hayes, and C.C. Olds. 1992. Summarizing Monte Carlo results in methodological research: the one- and two-factor fixed effects ANOVA cases. J. Educ. Stat. 17: 315-339.

Lix, L.M., J.C. Keselman, and H.J. Keselman. 1996. Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance F test. Rev. Educ. Res. 66: 579-619.

Rigoux, L., Stephan, K.E., Friston, K.J. and Daunizeau, J., 2014. Bayesian model selection for group studies—revisited. *Neuroimage*, *84*, pp.971-985.