

# **N-Methyl-D-aspartate receptor hypofunction causes recurrent and transient failures of perceptual inference**

## **Authors:**

Veith Weilhhammer<sup>1,2,3,*ec*</sup>, Marcus Rothkirch<sup>1,4,*ec*</sup>, Deniz Yilmaz<sup>1,5,6,7,*ec*</sup>, Merve Fritsch<sup>1</sup>, Lena Esther Ptasczynski<sup>1,5</sup>, Katrin Reichenbach<sup>1</sup>, Lukas Rödiger<sup>1</sup>, Philip Corlett<sup>8</sup>, Philipp Sterzer<sup>9</sup>

## **Affiliations:**

<sup>1</sup> Department of Psychiatry, Charité-Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Germany

<sup>2</sup> Berlin Institute of Health, Charité-Universitätsmedizin Berlin and Max Delbrück Center, Germany

<sup>3</sup> Helen Wills Neuroscience Institute, University of California Berkeley, USA

<sup>4</sup> Medical School Berlin, Hochschule für Gesundheit und Medizin, Germany

<sup>5</sup> Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, Germany

<sup>6</sup> Max Planck School of Cognition, Leipzig, Germany

<sup>7</sup> Department of Psychiatry and Psychotherapy, LMU University Hospital, Munich, Germany

<sup>8</sup> Department of Psychiatry, Yale University School of Medicine, New Haven, USA

<sup>9</sup> Department of Psychiatry (UPK), University of Basel, Switzerland

## **Contributions:**

*ec* Equal contribution

## **Corresponding Author:**

Veith Weilhhammer, Helen Wills Neuroscience Institute, University of California Berkeley, USA, email: [veith.weilhhammer@gmail.com](mailto:veith.weilhhammer@gmail.com)

# 1 Abstract

2 Perception is known to alternate between an external mode, driven by sensory inputs, and an  
3 internal mode, shaped by prior knowledge about the world. Using a double-blind, placebo-  
4 controlled, cross-over experiment in healthy human participants, we investigated the effects  
5 of the N-Methyl-D-aspartate receptor (NMDAR) antagonist S-ketamine on the alternation  
6 between external and internal modes. We found that pharmacologically induced NMDAR  
7 hypofunction causes a shift of perception toward the external mode. Similarly, a case-control  
8 study revealed that individuals with paranoid schizophrenia, a disorder repeatedly associated  
9 with NMDAR hypofunction, spend more time in the external mode. This NMDAR-dependent  
10 increase in the external mode suggests that the symptoms of schizophrenia are caused by  
11 recurring dissociations of perception from prior knowledge about the world.

## 12 2 Main

13 Imagine a dimly lit room at a crowded party, where unclear visual signals, indistinct sounds,  
14 and complex social interactions allow for multiple - and sometimes false - interpretations. In  
15 such ambiguity, failures of perceptual inference, the ability to contextualize sensory inputs  
16 with prior knowledge about the world, can lead to profound departures from reality: Faces  
17 obscured in shadow may appear distorted, random noise could be perceived as a whisper,  
18 and friendly smiles might seem derogatory.

19 This example illustrates why a disruption of perceptual inference is likely to play a crucial  
20 role in schizophrenia (Scz), a chronic and severe mental disorder characterized by psychotic  
21 symptoms such as delusions and hallucinations<sup>1-3</sup>. Yet despite considerable progress in the  
22 computational understanding of psychosis, two key questions have remained unanswered.

23 The first question concerns the neural mechanisms that cause perceptual inference to fail in  
24 Scz. Several lines of evidence point to N-Methyl-D-aspartate receptor (NMDAR) hypofunction  
25 as a key factor in the pathophysiology of psychosis<sup>4</sup>. NMDAR antibodies<sup>5</sup> and antagonists  
26 such as ketamine<sup>6</sup> mimic the symptoms of Scz, which is itself associated with a reduction  
27 of NMDAR density in prefrontal cortex<sup>7</sup>. NMDARs control the ratio of neural excitation  
28 and inhibition<sup>8</sup>, block the release of midbrain dopamine<sup>9</sup>, enable cortical feedback<sup>10</sup> and  
29 support synaptic short term plasticity<sup>11</sup>. While these NMDAR-dependent mechanisms are  
30 likely critical for perceptual inference, it is yet to be determined how NMDAR hypofunction  
31 may cause the psychotic symptoms that characterize Scz.

32 The second unresolved question concerns the temporal dynamics of psychotic experiences,

which often unfold as short-lived events spanning from seconds to minutes, especially at early stages of Scz<sup>12,13</sup>. The transient nature of psychotic experiences challenges models that assume a constant disruption of perceptual inference<sup>1-3</sup>. A solution to this problem is suggested by the recent observation that perceptual inference is subject to spontaneous fluctuations over time that occur at a timescale compatible with the duration of individual psychotic experiences<sup>14-16</sup>. Such fluctuations have been related to two opposing modes of inference, during which perception is driven predominantly either by external inputs or by internal predictions that stem from recent perceptual experiences<sup>17</sup> (Figure 1A). Although preliminary evidence indicates a tendency toward the external mode in people with Scz<sup>18</sup>, the neural mechanisms of mode fluctuations and their potential implications in psychosis have remained elusive.

The objective of the current study was therefore twofold: First, to test whether NMDAR hypofunction causes changes in perceptual inference that characterize Scz; and second, to explore the effect of NMDAR hypofunction on ongoing fluctuations in perceptual inference that may explain the transient nature of psychotic experiences. We addressed these aims in a double-blind placebo-controlled cross-over experiment in 28 healthy human participants. The participants attended two experimental sessions during which they received a continuous intravenous infusion of either the NMDAR antagonist S-ketamine at a dose of 0.1 mg/kg/h or a saline placebo. In each session, the participants viewed ten 120 sec blocks of an ambiguous structure-from-motion (SFM) stimulus that induced the experience of a sphere rotating around a vertical axis, and reported changes in the perceived direction of rotation (leftward vs. rightward movement of the front surface) as well as their confidence in the choice (Figure 1B and Supplemental Video S1).

The ambiguity of the display induced the phenomenon of bistable perception, where spontaneous changes in the perceived direction of rotation occurred in average intervals of  $13.75 \pm 3.09$  sec. In line with previous results<sup>19,20</sup>, these changes in perception occurred with a probability of  $0.11 \pm 8.67 \times 10^{-3}$  at brief depth-symmetric configurations of the stimulus (see Supplemental Video S1 and Supplemental Figure S1A). We therefore divided the continuous behavioral reports into a sequence of discrete states  $t$ . Each state was associated with a perceptual experience  $y_t$ , confidence  $c_t$  and the external input  $s_t$ .

Bistable perception can be conceptualized as an inferential process about the cause of  $s_t$ , in which previous experiences ( $y_{t-1}$ ) reflect internal predictions that provide prior knowledge about the interpretation  $y_t$  of the ambiguous stimulus<sup>20</sup> (Figure 1C). To test how NMDAR antagonism altered the balance between external inputs and internal predictions, we attached a 3D signal to a fraction of the stimulus dots. The signal-to-ambiguity ratio (SAR) ranged

from complete ambiguity to full disambiguation across five levels and remained constant in each block of the experiment. By changing the direction of rotation enforced by the 3D signal at random in average intervals of 10 sec, we created dynamic conflicts between the SAR-weighted input  $s_t$  and the stabilizing internal prediction  $y_{t-1}$ . As expected, we found that  $y_t$  was driven by both  $s_t$  ( $\beta_S = 3.01 \pm 0.06$ ) and  $y_{t-1}$  ( $\beta_P = 2.06 \pm 0.03$ ). Importantly, S-ketamine caused perception to shift toward  $s_t$  ( $0.45 \pm 0.08$ ,  $z = 5.6$ ,  $p = 1.71 \times 10^{-7}$ ; Figure 2A and Supplemental Figure S2), indicating a stronger weighting of external inputs over internal predictions during pharmacologically induced NMDAR hypofunction.

Next, we performed the same analysis on data from a previous case-control study using an analogous task in patients with Scz<sup>19</sup>. In Scz patients and controls,  $y_t$  was influenced by the SAR-weighted input  $s_t$  ( $\beta_S = 2.77 \pm 0.11$ ) and the stabilizing prediction  $y_{t-1}$  ( $\beta_P = 1.5 \pm 0.03$ ). Similar to S-ketamine,  $s_t$  had a larger impact on perception in Scz patients than controls ( $0.75 \pm 0.15$ ,  $z = 4.96$ ,  $p = 5.6 \times 10^{-6}$ ; Figure 2E and Supplemental Figure S3). Together, these results align with the canonical predictive processing theory of Scz<sup>1-3</sup>: Pharmacologically-induced NMDAR hypofunction and Scz are associated with a shift of perceptual inference toward external inputs, and away from stabilizing internal predictions. NMDAR hypofunction may thus trigger psychotic experiences by causing erratic inferences about ambiguous sensory information.

As a mechanism for symptoms that are transient and recurring, NMDAR-dependent changes in perceptual inference should not be constant, but fluctuate dynamically at a timescale that is compatible with the duration of individual psychotic experiences. We tested this prediction in Hidden Markov Models (HMM) that inferred transitions between two latent states, each linked to an independent general linear model (GLM) that predicted  $y_t$  from  $s_t$  and  $y_{t-1}$ . The  $\beta$  weights quantified the sensitivity to ambiguous sensory information ( $\beta_S \times s_t$ ) relative to the stabilizing effect of internal predictions provided by preceding experiences ( $\beta_P \times y_{t-1}$ ), and allowed us to evaluate dynamic changes in the balance  $\Delta_{S-P} = \beta_S - \beta_P$  between the two.

Consistent with recent findings in humans and mice<sup>16,17</sup>, Bayesian model comparison indicated a clear superiority of the two-state GLM-HMM over the standard one-state GLM in the S-ketamine experiment ( $\delta_{BIC} = -3.65 \times 10^3$ ). According to the two-state GLM-HMM, perception fluctuated between an internal mode, shaped by the stabilizing internal prediction  $y_{t-1}$ , and an external mode, dominated by the SAR-weighted input  $s_t$ . External mode increased  $\Delta_{S-P}$  by  $2.8 \pm 0.29$  ( $T(81) = 9.5$ ,  $p = 5.22 \times 10^{-13}$ ; Figure 2B-C). Switches between modes occurred in intervals of  $179.97 \pm 19.39$  sec.

The presence of slow fluctuations between external and internal modes suggests that, instead of causing a constant increase in the sensitivity to external inputs, NMDAR hypofunction

may affect perception by shifting the dynamic balance between the two modes. Indeed, S-ketamine did not alter the weights of the two-state GLM-HMM (Figure 2C), but increased the probability of external at the expense of internal mode ( $1.01 \pm 0.03$ ,  $z = 30.7$ ,  $p = 4.26 \times 10^{-206}$ ; Figure 2D). This effect was stable over time and present across the full range of SAR (Figure 2D). Inter-individual differences in the effects of S-ketamine confirmed that NMDAR hypofunction raised the sensitivity to sensory information (Figure 2A) by modulating the time participants spent in external and internal modes, respectively ( $\rho = 0.41$ ,  $T(26) = 2.3$ ,  $p = 0.03$ ).

Strikingly, the data from the Scz-control study mirrored the effect of S-ketamine on the balance between external and internal mode: The two-state GLM-HMM outperformed the standard one-state GLM (patients:  $\delta_{BIC} = -981.65$ ; controls:  $\delta_{BIC} = -862.91$ ) and revealed two opposing modes ( $\Delta_{S-P} = 1.44 \pm 0.33$ ,  $T(44) = 4.33$ ,  $p = 3.39 \times 10^{-4}$ ; Figure 2F) that alternated in intervals of  $265.38 \pm 57.76$  sec for patients and  $230.99 \pm 65.04$  sec for controls. Patients and controls did not differ with respect to the weights of the two-state GLM-HMM (Figure 2G). Instead, Scz patients spent more time in external mode ( $0.52 \pm 0.03$ ,  $z = 16.88$ ,  $p = 1.23 \times 10^{-63}$ ; Figure 2H).

We could not attribute between-mode transitions to fatigue, task difficulty (Figure 2D and H), executive function as reflected by response times (RT, Supplemental Figure S1C and S4), psychotomimetic effects of S-ketamine, global psychosis proneness, the clinical severity of Scz, stereodisparity thresholds, or subjective arousal, intoxication and nervousness (Supplemental Figure S5).

To our knowledge, these results are the first to uncover a neural mechanisms for the slow, task-related fluctuations in perceptual inference that have been observed across variety of tasks in humans and mice<sup>14-17</sup>. We found that healthy individuals who receive the NMDAR antagonist S-ketamine and patients diagnosed with Scz are prone to an external mode of perception. The external mode partly decouples perceptual inference from internal predictions that reflect prior knowledge about the world. In health, this may prevent circular inferences in recurrent neural networks, where predictive feedback modulates activity even at early stages of sensory processing<sup>21,22</sup>. A predominance of external mode, on the other hand, exposes perception to the destabilizing effects of ambiguity. Such transient failures of perceptual inference may cause individuals to be deluded by spurious connections between unrelated events<sup>3</sup>, to attribute the sensory consequences of their actions to an outside force, and to hallucinate signals in noise<sup>3</sup>.

During external mode, healthy participants were more confident in their choices ( $0.72 \pm 0.07$ ,  $z = 9.92$ ,  $p = 7.85 \times 10^{-22}$ , Supplemental Figure S1E) and scored higher on the

*Clinician-Administered-Dissociative-States-Scale*<sup>23</sup> (CADSS,  $1.05 \pm 0.54$ ,  $T(208.05) = 1.95$ ,  $p = 0.05$ , Supplemental Figure S5B). Given the known association of psychosis with elevated confidence<sup>24</sup> and dissociative symptoms<sup>25</sup>, intervals of external mode may thus reflect the computational correlate of individual psychotic experiences. The dynamic nature of between-mode transitions illustrates how constant and potentially heritable dysfunctions of the NMDAR, such as GRIN2A mutations in Scz<sup>26</sup>, may produce symptoms of psychosis that are recurrent and transient in nature.

## 3 Methods

### 3.1 Ressource availability

#### 3.1.1 Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Veith Weinhhammer (veith.weinhhammer@gmail.de).

#### 3.1.2 Materials availability

This study did not generate new unique reagents.

#### 3.1.3 Data and code availability

All data and code associated with this study will be made available on the associated Github repository [https://github.com/veithweinhhammer/modes\\_ketamine\\_scz](https://github.com/veithweinhhammer/modes_ketamine_scz) upon publication. Key resources are listed in Supplemental Table S1.

### 3.2 S-ketamine vs. placebo

The S-ketamine experiment consisted in a total of three experimental sessions. During the first session, we screened participants for S-ketamine contraindications (arterial hypertension, prior psychiatric or neurological diagnoses including substance use disorder, use of psychoactive medication), and assessed psychosis proneness using the 40-item *Peters Delusion Inventory* (PDI<sup>27</sup>) and the 32-item *Cardiff Anomalous Perception Scale* (CAPS<sup>28</sup>). Moreover, we conducted three experimental pre-test runs that tested the ability to process stereodisparity (run 1, SAR = 1, cut-off: perceptual accuracy > 0.75), ensured the experience of spontaneous switches during bistable perception (run 2, SAR = 0, cut-off: perceptual stability < 0.96, corresponding to phase durations < 40 sec), and familiarized participants with the main experiment (run 3, see below for details).

In the subsequent two sessions, participants received a continuous intravenous infusion of either S-ketamine at 0.1 mg/kg/h or a saline placebo. Health screenings were repeated before each session to ensure the participants remained eligible. At each day of testing, we checked for alcohol intoxication using a breathalyzer and recent illicit substance use via a urine drug screen.

Our experimental protocol was double-blinded: The order of S-ketamine and placebo administration was counter-balanced across participants, with at least a two week interval between

174 sessions. The participants, as well as the experimenters tasked with collecting the behavioral  
175 and psychometric data, were unaware of whether S-ketamine or placebo was administered  
176 by an independent group of clinicians who excluded undiagnosed psychotic illness using the  
177 *Brief Psychiatric Rating Scale* (BPRS<sup>29</sup>), established the intravenous line, started the infusion  
178 15 min prior to the experiment, monitored the participants for side effects (blood pressure,  
179 drowsiness and vasovagal reactions, psychotomimetic effects), and removed the intravenous  
180 line at the end of the experiment, after which participants were monitored for at least 30  
181 min. Deblinding occurred after data collection was complete.

### 182 3.2.1 Sample characteristics

183 We screened a total of 87 right-handed individuals with (corrected-to-) normal vision, who  
184 were naive to the purpose of the study and gave written informed consent before participating.  
185 All experimental procedures were approved by the ethics committee at Charité Berlin.

186 From the group of screened participants, 31 did not meet our pretest criteria (6 due to  
187 perceptual accuracy  $< 0.75$ , 15 due to perceptual stability  $> 0.96$ , 8 due to substance use,  
188 1 due to do a diagnosis of ADHD, and 1 due to medication with sertraline). Out of the  
189 remaining 56 participants who were eligible for the S-ketamine experiment, we aborted  
190 the main experiment in 1 participant due to high blood pressure at baseline (RR  $> 140/80$   
191 mmHG), in 2 participants due to strong psychotomimetic effects (micropsia) or dizziness under  
192 S-ketamine, and in 1 participant due to a vasovagal syncope during intravenous insertion.  
193 24 participants were not available for the main experiment after successful pre-testing. We  
194 therefore report the data from a total of 28 participants (mean age:  $28.93 \pm 1.35$  years, 18  
195 female) who met all inclusion criteria and completed all experimental sessions.

### 196 3.2.2 Experimental paradigm

197 We presented the experiment using Psychtoolbox 3<sup>30</sup> running in Matlab R2021b (session  
198 1: CRT-monitor at 85 Hz, 1280 x 1024 pixels, 60 cm viewing distance and 39.12 pixels per  
199 degree visual angle; session 2 and 32: CRT-monitor at 85Hz, 1280 x 1024 pixels, 40 cm  
200 viewing distance and 26.95 pixels per degree visual angle).

201 **Procedure:** Throughout the experiment, participants reported their perception of a discon-  
202 tinuous SMF stimulus (Supplemental Video S1). In this stimulus, random dots distributed on  
203 two intersecting rings induce the perception of a spherical object (diameter:  $15.86^\circ$ , rotational  
204 speed: 12 sec per rotation, rotations per block: 10, individual dot size:  $0.12^\circ$ ) that rotates  
205 around a vertical axis with the front surface to the left or right<sup>20</sup>. Stimuli were presented in  
206 120 sec blocks, separated by 10 sec fixation intervals.



Participants viewed the stimuli through a custom mirror stereoscope. In the pretest experiment, we presented stimuli at complete disambiguation (run 1,  $SAR = 1$ ), full ambiguity (run 2,  $SAR = 0$ ) and across five levels ranging from full ambiguity to complete disambiguation across five levels (run 3-5,  $SAR \in \{0, 0.1, 0.25, 0.5, 1\}$ ). The SAR, which was constant within blocks, defines the fraction of stimulus dots that received a disambiguating 3D signal. Within each block, the direction of rotation enforced by the disambiguating 3D signal changes in average intervals of 10 sec (i.e., at a probability of 0.15 per stimulus overlap, see below). We pseudo-randomized the order of SAR across blocks and the direction of disambiguation within blocks.

Participants were naive to the potential ambiguity in the visual display, passively experienced the stimulus and reported changes in their perception alongside their confidence via button-presses on a standard USB keyboard (right middle-finger on d: rotation of the front-surface to the right at high confidence; right index-finger on f: rotation of the front-surface to the right at low confidence; left middle-finger on k: rotation of the front-surface to the left at high confidence; left index-finger on j: rotation of the front-surface to the left at low confidence; thumb on space bar: unclear direction of rotation). Unclear perceptual states occurred at a rate of  $0.03 \pm 0.01$  and were excluded from further analyses.

The direction of rotation enforced by  $s_t$  (i.e., whether the parametric 3D signal enforced leftward or rightward rotation of the front surface) changed at a rate of 0.15 per overlap (i.e., on average every 10 sec). Changes in  $s_t$  and the order of blocks, each corresponding to one level of SAR, were pseudo-random.

In session 1 (pre-test), each run (runs 1 to 3) consisted of six blocks. In session 2 and 3 (main experiment), each run (run 4 and 5) consisted of 10 blocks. After every third block, the main experiment was paused to allow for the monitoring of the participants' vital signs (blood pressure and pulse rate) and dynamic changes in psychotomimetic experiences. The latter was assessed using the 6 item *Clinician-Administered-Dissociative-States-Scale* (CADSS<sup>23</sup>) and three additional questions (Q1: *How awake do you feel?*, Q2: *How intoxicated do you feel?*, Q3: *How nervous do you feel?*) to which participants responded by clicking on a continuous line that encoded responses from *not at all* to *very much*. To measure global psychotomimetic effects of S-ketamine vs. placebo, participants completed the Questionnaire for the *Assessment of Altered States of Consciousness* (5D-ASC<sup>31</sup>) at the end of session 2 and 3. In addition, we collected responses on a debriefing questionnaire, in which we asked participants to describe whether they were able to accurately perceive the two directions of rotation induced by the SFM stimulus, whether they noticed any differences between blocks, whether they would guess that they received S-ketamine or placebo, and whether they had

experienced any effects that they would attribute to a psychoactive substance.

**Stereodisparity thresholds:** At the beginning of the session 2 and 3, we conducted an independent stereo-acuity test to detect a potential effect of S-ketamine on stereodisparity thresholds<sup>19</sup>. We presented 5000 dots (each at  $0.15^\circ$  visual angle) within a square of  $11^\circ \times 11^\circ$  around a central fixation cross ( $0.10^\circ$ ). We added a stereodisparity signal to all dots on a Landolt C, i.e., a circle ( $1.37^\circ$  radius,  $2.06^\circ$  width) with a  $90^\circ$  gap located either at the left, top, right or bottom. Stimuli were presented for 1 sec, after which participants reported the location of the gap by pressing the up-, down-, left- or right-arrow key within a 2 sec response interval, followed by 5 sec of fixation before the next trial.

We adjusted the stereodisparity of the Landolt C in a two-up-one-down staircase across 40 trials (initial stereodisparity:  $0.0045^\circ$ , correct response: decrease in the available stereodisparity by one step; incorrect response: increase by two steps, initial step-size:  $0.001^\circ$ , reduction to  $0.0005^\circ$  after first reversal). Stereodisparity thresholds were defined by the average stereodisparity present at the last 10 trials of the staircase.

**Scores and Questionnaires:** Supplementary Table S2 provides an overview of our psychometric data.

### 3.3 Scz patients vs. healthy controls

To test whether Scz patients show similar changes in bimodal inference as healthy participants who receive the NMDAR-antagonist S-ketamine, we re-analyzed data from a previously published case-control study<sup>19</sup> that compared Scz patients to healthy participants in paradigm analogous to the S-ketamine experiment described above.

#### 3.3.1 Sample characteristics

We report data from 23 patients diagnosed with paranoid Scz (ICD-10: F20.0, 18 male, age =  $37.13 \pm 2.42$ ) and 23 controls (17 male, age =  $33.57 \pm 1.74$ ) that were matched for gender, age and handedness<sup>19</sup>.

#### 3.3.2 Experimental paradigm

Stimuli were presented using Psychtoolbox 3<sup>30</sup> running in Matlab R2007b (CRT-Monitor at 60 Hz, 1042x768 pixels, 59.50cm viewing distance, 30.28 pixels per degree visual angle).

**Main Experiment:** Throughout the experiment, participants reported their perception of a discontinuous SFM stimulus (see Supplemental Video S2) via button-presses on a standard

USB keyboard. In contrast to the S-ketamine experiment, the 300 dots ( $0.05^\circ$ ) that composed the stimulus ( $2.05^\circ \times 2.05^\circ$ ) were not placed on rings, but on a Lissajous band defined by the perpendicular intersection of two sinusoids ( $x(p) = \sin(A * p)$  and  $y(p) = \cos(B * p + \delta)$  with  $A = 3$ ,  $B = 6$ , with  $\delta$  increasing from 0 to  $2\pi$  at 0.15 Hz. Overlapping configurations of the stimulus occurred in intervals of 3.33 sec. Participants viewed the stimuli through a mirror stereoscope. Fusion was supported by rectangular fusion-frames and a background of random dot noise (700 dots of  $0.05^\circ$  which moved at a speed of  $1.98^\circ$  per sec and changed their direction at a rate of 1 Hz).

We presented participants with 3 sessions of the main experiment, each consisting of 14 40.08 sec blocks that were separated by 5 sec of fixation and differed with respect to the SAR, ranging from full ambiguity to complete disambiguation in 8 levels ( $SAR \in \{0, 0.01, 0.04, 0.9, 0.16, 0.26, 0.50, 1\}$ ). The frequency of changes in the direction of the disambiguating signal corresponded to the frequency of spontaneous changes that participants perceived during full ambiguity<sup>19</sup> ( $SAR = 0$ ). In contrast to the S-ketamine experiment, participants only reported the perceived direction of rotation  $y_t$  (left vs. rightward movement of the front surface).

**Stereodisparity thresholds:** We assessed stereodisparity thresholds in Scz patients and controls using the procedure described above.

**Scores and Questionnaires:** We used the PDI<sup>27</sup> and the CAPS<sup>28</sup> to measure delusional ideation and perceptual anomalies in Scz patients and controls. Clinical symptom severity was assessed using the *Positive and Negative Syndrome Scale* (PANSS)<sup>32</sup>.

### 3.4 Quantification and statistical procedures

This manuscript was written in RMarkdown. All data and summary statistics can be reviewed by cloning the Github repository [https://github.com/veithweilnhammer/Ketamin\\_RDK](https://github.com/veithweilnhammer/Ketamin_RDK) and running the file *ketamine\_scz\_fmri\_modes.Rmd*, which will be made public at the time of publication.

The SFM stimuli used in the above studies share an important feature: Even though physically ambiguous at all angles of rotation, spontaneous changes in the perceived direction of rotation are limited to overlapping configurations of the stimuli<sup>19,20</sup> (see also Supplemental Figure S1 and S4). This is because depth-symmetry, which is a prerequisite for changes in subjective experiences during bistable SFM<sup>19,20</sup>, is limited to timepoints when the bands that compose the stimuli overlap (Supplemental Video S1 and S2).

We therefore discretized the perceptual timecourse of all experiments into a sequence of

overlaps that occur at times  $t$  (1.5 sec inter-overlap interval for the S-ketamine and fMRI experiment, 3.33 sec inter-overlap interval for the case-control experiment). Each inter-overlap interval is characterized by the primary independent variable  $s_t = [-1, 1] \times SAR$  (the SAR-weighted input ranging from maximum information for leftward rotation to maximum information for rightward rotation). As secondary independent variables, we considered block and session index (reflecting the time participants were exposed to the experiment), participant identifiers and, if applicable, treatment or group identifiers. Primary dependent variables were  $y_t = [0, 1]$  (the experience of either leftward or rightward rotation),  $r_t$  (the time between the button-press indicating a perceptual event relative to the preceding overlap) and, if applicable,  $c_t = [0, 1]$  (low vs. high confidence). As secondary dependent variables, we computed perceptual accuracy (the probability of  $y_t \cong s_t$ ) and perceptual stability (the probability of  $y_t = y(t - 1)$ ). We report averages as mean  $\pm$  s.e.m.

### 3.4.1 Conventional statistics

The goal of our conventional statistics was to quantify the effect of NMDAR hypofunction, whether due to an antagonism with S-ketamine or due to a diagnosis of Scz, on the interpretation of ambiguous sensory information. To this end, we performed standard logistic and linear regression by fitting (general) linear effects models using the R-packages lmer, glmer and afex (see Supplemental Table S2). We predicted  $y_t$ ,  $c_t$ , perceptual accuracy and perceptual stability in logistic regression, and  $r_t$  in linear regression. We estimated random intercepts defined within participants in the S-ketamine experiment and nested random intercepts for participants within groups in the case-control experiment. We applied a Bonferroni-correction for the number of main effects and interactions within models. For non-normally distributed secondary dependent variables, we performed rank-based tests to assess correlations (Spearman) and distribution differences (Wilcoxon).

### 3.4.2 Computational modeling

Having established the effect of NMDAR hypofunction on the interpretation of ambiguous sensory information, we used computational modeling to arbitrate between two mechanistic explanations on how S-ketamine and schizophrenia may alter perceptual inference.

**Hypothesis H1: Unimodal inference** In one scenario, NMDAR hypofunction may induce a global increase in the sensitivity to external inputs relative to the stabilizing internal prediction. This unimodal scenario would be reflected by S-ketamine- or Scz-related changes in the weights  $w \equiv \{\beta_S, \beta_P, \beta_B\}$  of a GLM that predicts percepts  $y_t$  from the input vector  $x_t$ , which consists in the SAR-weighted external input  $s_t$ , the stabilizing internal prediction  $y_{t-1}$

338 and a constant bias  $b$ :

$$P(y_t = 1|x_t) = \frac{1}{1 + e^{-x_t \times w}} \quad (1)$$

$$x_t \times w = s_t \times \beta_S + y_{t-1} \times \beta_P + b \times \beta_B \quad (2)$$

339 According to the unimodal hypothesis 1, NMDAR hypofunction increases  $\beta_S$  at the expense  
340 of  $\beta_P$ , leading to an increase of  $\Delta_{S-P} = \beta_S - \beta_P$ .

341 **Hypothesis H2: Bimodal inference** In an alternative scenario, NMDAR hypofunction  
342 does not change the weights of the GLM directly, but modulates the transition between  
343 latent modes<sup>17</sup> or decision-making strategies<sup>16</sup> that differ with respect to the balance between  
344 external inputs  $s_t$  and the stabilizing internal prediction provided by  $y_{t-1}$ . In the bimodal  
345 scenario, perceptual inference is characterized by two latent modes  $z_t$  (i.e., states in a HMM)  
346 that alternate at a probability per overlap that is defined by a 2 x 2 transition matrix  $A$ :

$$P(z_t = k|z_{t-1} = j) = A_{kj} \quad (3)$$

347 Each state  $z_t$  is associated by an independent GLM defined by the weights  $w_k$ :

$$P(y_t = 1|x_t, z_t) = \frac{1}{1 + e^{-x_t \times w_k}} \quad (4)$$

$$x_t \times w_k = s_t \times \beta_{S,k} + y_{t-1} \times \beta_{P,k} + b \times \beta_{B,k} \quad (5)$$

348 The bimodal hypothesis H2 differs from the unimodal hypothesis H1 in two ways: First,  
349 bimodal inference is characterized by two (as opposed to one) GLMs that differ with respect  
350 to  $\Delta_{S-P}$ : During external mode,  $\beta_S$  is increased relative to  $\beta_P$ , whereas during internal mode,  
351  $\beta_P$  is increased relative to  $\beta_S$ . Second, during bimodal inference, NMDAR hypofunction does  
352 not alter the weights within the external and internal GLMs, but modulates the transition  
353 probability between the two.

354 **Procedure:** To contrast hypotheses H1 and H2, we fitted unimodal and bimodal GLM-HMMs  
355 using SSM<sup>33</sup> (Supplemental Table S2), compared models via Bayesian Information Criterion  
356 (BIC), and assessed the effects of S-ketamine or Scz on the posterior model parameters, i.e.,  
357 HMM transition probabilities and the mode-dependent GLM weights  $w_k$ . Model fitting using

SSM is governed by the hyperparameters  $\sigma^2$  and  $\alpha$ .  $\sigma^2$  denotes the variance of a prior over the GLM weights  $w_k$ . Smaller values of  $\sigma^2$  shrink  $w_k$  toward 0, whereas  $\sigma = \infty$  leads to flat priors. We set  $\sigma^2$  to 100 for GLMs that predicted group-level data, and to 1 for GLMs that predicted participant- or session-level data, which were initialized with group-level estimates of  $w_k$ .  $\alpha$  defines the Dirichlet prior over the transition matrix  $A$  and is flat for  $\alpha = 1$ . We set  $\alpha$  to 1 for all group-level and participant-level fits.

For each experiment, computational modeling was carried out in a sequence of 3 steps: In a first step, we fitted a unimodal GLM initialized with noisy weights to the group-level data (i.e., data pooled across participants within an individual experiment) for a total of  $n = 100$  iterations and computed the average posterior weights  $w_n$ . In a second step, we fitted the group-level data with the unimodal and the bimodal GLM-HMM initialized by  $w_n$ , extracted the posterior parameters  $w_k$ , and compared the models using BIC.

In a third step, we fitted the unimodal and the bimodal GLM-HMM to session-level data (S-ketamine experiment) and participant-level data (case-control experiment). Models were initialized by the average weights  $w_n$  of the corresponding group-level model. For all bimodal group-, participant- and session-level GLM-HMMs, we defined the latent mode associated with the higher posterior  $\beta_S$  estimate as external. Our definition of mode is thus agnostic with respect to  $\beta_P$  and  $\Delta_{S-P}$ . This allowed us to contrast external-to-internal bimodal inference (hypothesis H2) with a third alternative where perception fluctuates between latent states that differ with respect to decision noise (**hypothesis H0**). Low decision noise is characterized by higher posterior estimates of  $\beta_S$  and  $\beta_P$ , whereas high decision noise is characterized by lower posterior estimates of  $\beta_S$  and  $\beta_P$ .  $\Delta_{S-P}$  therefore discerns fluctuations in decision noise (H0, no changes in  $\Delta_{S-P}$ ) from external-to-internal bimodal inference (H2, mode-associated changes in  $\Delta_{S-P}$  with higher estimates during external mode).

For summary statistics, we extracted the posterior weights  $w_k$  (separately for external and internal mode) and the dynamic posterior probability of external mode  $z_t = e$ .

## 4 Declaration of interest

The authors declare no competing interests.

## 5 Acknowledgements

THIS work was funded by the Leopoldina Academy of Sciences (grant number: LDPS2022-16, <https://www.leopoldina.org/en/leopoldina-home/>), the German Research Foundation DFG (grant number: STE 1430/8-1, <https://www.dfg.de>), the Berlin Institute of Health Clinician Scientist Program (<https://www.bihealth.org/en/translation/innovation-enabler/academy/bih-charite-clinician-scientist-program>), and the German Ministry for Research and Education (ERA-NET NEURON program, grant number: 01EW2007A, <https://www.neuron-eranet.eu/>). The funders had no role in study design, data collection, data analysis, decision to publish, or preparation of the manuscript.

## 6 Figures

### 6.1 Figure 1

#### Figure 1.

**A.** When inferring whether the world is one state or another ( $P(y_t = 1)$  or  $P(y_t = 0)$ , respectively), the brain integrates ambiguous sensory signals  $s_t$  with internal predictions that reflect prior knowledge about the environment. One source of prior knowledge is the temporal autocorrelation of natural stimuli, in which the recent past often predicts the near future. The integration of external inputs and internal predictions depends on the weights assigned to incoming sensory data ( $\beta_S \times s_t$ ) and to internal prediction derived from previous experiences ( $\beta_P \times y_{t-1}$ , dotted versus solid lines, simulated data), respectively.  $\beta_S$  determines the slope, and  $\beta_P$  the shift of the psychometric function that links  $s_t$  and  $y_t$ . Importantly, the balance  $\Delta_{S-P} = \beta_S - \beta_P$  is known to alternate between two opposing modes: During the external mode (left), perception is largely determined by  $\beta_S \times s_t$ , which is reflected by a steep slope and a small shift of the psychometric curve. Conversely, during the internal mode (right), perception is shaped by  $\beta_P \times y_{t-1}$ , resulting in a shallow slope and a large shift of the psychometric curve.

**B.** We conducted a double-blind placebo-controlled experiments in 28 healthy human participants, who received a continuous infusion with either the NMDAR antagonist S-ketamine or saline. During the infusion, the participants viewed an ambiguous SFM stimulus that was compatible with two mutually exclusive subjective experiences (left vs. rightward rotation of the front surface, red arrows). This induced the phenomenon of bistable perception: While the ambiguous stimulus remained constant, participants perceived only one direction of rotation (blue arrow), before switching to the competing alternative.

**C.** Changes in the perceived direction of rotation of the SFM stimulus occur at brief depth-symmetric configurations of the stimulus (Supplemental Video S1). We therefore transformed the behavioral responses into a sequence of states  $t$  (80 1.5 sec intervals per block), each associated with a combination of the SAR-weighted input  $s_t$  and the perceived direction of rotation  $y_t$ . We used GLMs to quantify the weights  $\beta_S$ ,  $\beta_P$  and  $\beta_B$ , which reflect how inferences  $y_t$  were determined by the external inputs  $\beta_S \times s_t$  and internal predictions  $\beta_P \times y_{t-1}$ .



## 6.2 Figure 2

### Figure 2.

**A.** The percepts  $y_t$  were more likely to match the stimuli  $s_t$  at higher levels of SAR ( $3.01 \pm 0.06$ ,  $z = 50.39$ ,  $p = 0$ ). The positive effect of SAR on  $P(y_t \cong s_t)$  was more pronounced under S-ketamine (red) relative to placebo (blue;  $0.45 \pm 0.08$ ,  $z = 5.6$ ,  $p = 1.71 \times 10^{-7}$ ).

**B.** In the S-ketamine experiment, the HMM identified two modes that differed with respect to the relative weighting of external sensory data and internal predictions: Perception fluctuated between an external mode, determined by the input  $s_t$  (upper panel, steep slope and small shift of the psychometric curve), and an internal mode, dominated by a stabilizing prediction that biased perception toward previous experiences  $y_{t-1}$  (lower panel, shallow slope and large shift of the psychometric curve). Within modes, there was no significant effect of S-ketamine (red) versus placebo (blue) on the relation of  $y(t)$  with  $s(t)$  and  $y(t-1)$ .

**C.**  $\Delta_{S-P}$ , the balance between the external input and the stabilizing internal predictions, was larger during external than during internal mode ( $2.8 \pm 0.29$ ,  $T(-81) = -9.5$ ,  $p = 5.22 \times 10^{-13}$ ). Importantly, we found no significant effect of S-ketamine (red) vs. placebo (blue) on  $\Delta_{S-P}$  within modes ( $-0.03 \pm 0.29$ ,  $T(81) = -0.1$ ,  $p = 1$ ).

**D.** S-ketamine (red) increased the probability of external mode ( $1.01 \pm 0.03$ ,  $z = 30.7$ ,  $p = 4.26 \times 10^{-206}$ ) relative to placebo (blue; left panel). The effect of S-ketamine on mode was present from the start of the session ( $1.77 \pm 0.07$ ,  $z = 26.9$ ,  $p = 8.88 \times 10^{-159}$ , upper panel), with no significant effect of time ( $-0.18 \pm 0.08$ ,  $z = -2.17$ ,  $p = 0.12$ ). Relative to placebo, S-ketamine increased the probability of external mode across all SARs ( $0.85 \pm 0.06$ ,  $z = 14.14$ ,  $p = 8.33 \times 10^{-45}$ , lower panel). Higher SARs were associated with an increased probability of external mode ( $1.34 \pm 0.09$ ,  $z = 15.01$ ,  $p = 2.49 \times 10^{-50}$ ), in particular under S-ketamine ( $0.62 \pm 0.11$ ,  $z = 5.52$ ,  $p = 1.32 \times 10^{-7}$ ). Alternations between external and internal mode were found at all SARs: From full ambiguity to complete disambiguation, the probability of external mode increased by only 0.11 under S-ketamine and 0.07 under placebo.

**E.** In patients (red) and controls (blue), percepts  $y_t$  were more likely to match the stimuli  $s_t$  at higher levels of SAR ( $\beta_S = 2.77 \pm 0.11$ ,  $z = 24.85$ ,  $p = 2.18 \times 10^{-135}$ ). Patients followed the external inputs more closely than controls ( $0.75 \pm 0.15$ ,  $z = 4.96$ ,  $p = 5.6 \times 10^{-6}$ ).

**F.** In analogy to the S-ketamine experiment, the HMM identified two opposing modes in Scz patients (red) and controls (blue). The external mode increased the sensitivity toward  $s_t$  (slope of the psychometric function) and weakened the effect of the stabilizing internal

457 prediction  $y_{t-1}$  (shift between the dotted and solid line) relative to the internal mode. Within  
458 modes, there was no effect of group on the relation of  $y(t)$  with  $s(t)$  and  $y(t-1)$ .

459 **G.** The external mode increased  $\Delta_{S-P}$ , the balance between external inputs and internal  
460 predictions, in patients (red) and controls (blue;  $1.44 \pm 0.33$ ,  $T(44) = 4.33$ ,  $p = 3.39 \times 10^{-4}$ ),  
461 with no significant effect of group ( $-0.28 \pm 0.54$ ,  $T(87.97) = -0.52$ ,  $p = 1$ ).

462 **H.** Relative to controls (blue), patients (red) spent more time in external mode ( $0.52 \pm$   
463  $0.03$ ,  $z = 16.88$ ,  $p = 1.23 \times 10^{-63}$ , left panel). In both group, biases toward external mode  
464 increased over time after session onset ( $2.41 \pm 0.11$ ,  $z = -21.37$ ,  $p = 1.02 \times 10^{-100}$ ; upper  
465 panel), with a stronger effect in patients ( $1.84 \pm 0.14$ ,  $z = 12.97$ ,  $p = 7.09 \times 10^{-38}$ ). Patients  
466 were more likely than controls to be in external mode across all levels of SAR ( $0.51 \pm 0.03$ ,  $z$   
467  $= 14.56$ ,  $p = 1.89 \times 10^{-47}$ , lower panel). External mode increased with SAR ( $0.63 \pm 0.1$ ,  $z$   
468  $= 6.47$ ,  $p = 3.85 \times 10^{-10}$ ), with no significant difference between groups ( $0.15 \pm 0.13$ ,  $z =$   
469  $1.16$ ,  $p = 0.98$ ). As in the S-ketamine experiment, alternations between external and internal  
470 mode were found at all SARs: From from full ambiguity to complete disambiguation, the  
471 probability of external mode increased by only 0.12 in patients and 0.18 in controls.

## 7 Supplemental Information

### 7.1 Supplemental Figure S1

**Supplemental Figure S1. The effects of ketamine and bimodal inference on RT.**

**A.** RT were non-uniformly distributed across the inter-overlap interval ( $D = 0.09$ ,  $p = 5.38 \times 10^{-9}$ , one-sample Kolmogorov-Smirnov test). This corroborates that changes in perception aligned with the overlapping configurations of the stimulus after S-ketamine (red) and placebo (blue).

**B.** RT showed a quadratic relationship with  $s_t$  ( $-6.87 \pm 1.68$ ,  $T(6.2 \times 10^3) = -4.1$ ,  $p = 5.1 \times 10^{-4}$ ), indicating faster responses when sensory information was reliable ( $|s_t| \gg 0$ ; note that SAR as shown in Figure 2A and 2E is equal to  $|s_t|$ ). We observed no main effect of S-ketamine (red) vs. placebo (blue) on RT ( $-3.35 \times 10^{-3} \pm 0.01$ ,  $T(6.2 \times 10^3) = -0.32$ ,  $p = 1$ ).

**C.** We found no additional effect of mode on RT ( $0.02 \pm 0.03$ ,  $z = 5.96 \times 10^3$ ,  $p = 0.78$ ).

**D.** Confidence showed a quadratic relationship with  $s_t$  ( $74.83 \pm 2.39$ ,  $z = 31.32$ ,  $p = 3.22 \times 10^{-214}$ ), confirming that participants were more confident when sensory information was reliable ( $|s_t| = SAR \gg 0$ ). Relative to placebo (blue), S-ketamine (red) reduced choice confidence ( $-0.21 \pm 0.04$ ,  $z = -5.9$ ,  $p = 4.36 \times 10^{-8}$ ), and decreased the quadratic effect of  $s_t$  on confidence ( $-19.95 \pm 2.36$ ,  $z = -8.45$ ,  $p = 3.48 \times 10^{-16}$ ).

**E.** External mode increased confidence globally ( $0.72 \pm 0.07$ ,  $z = 9.92$ ,  $p = 7.85 \times 10^{-22}$ ) and by elevating the quadratic effect of  $s_t$  on confidence ( $242.61 \pm 18.43$ ,  $z = 13.16$ ,  $p = 3.37 \times 10^{-38}$ ). When controlling for mode, the negative effect of S-ketamine (red) vs. placebo (blue) on confidence and on the quadratic relationship of confidence with  $s_t$  remained significant.

## 7.2 Supplemental Figure S2

### Supplemental Figure S2. Extended data on the effects of S-ketamine and mode on perceptual inference (related to Figure 2A-C).

**A.** Here, we show psychometric curves (percept  $y_t$  versus input  $s_t$ ) under S-ketamine (red) and placebo (blue). The plot separates times  $t$  for which the previous experience was leftward rotation ( $y_{t-1} = -1$ , upper panel) and rightward rotation ( $y_{t-1} = +1$ , lower panel). As expected,  $y_t$  was driven by both the external input  $s_t$  ( $\beta_S = 3.01 \pm 0.06$ ,  $z = 50.39$ ,  $p < 2.2 \times 10^{-308}$ ) and the previous percept  $y_{t-1}$  ( $\beta_P = 2.06 \pm 0.03$ ,  $z = 80.58$ ,  $p < 2.2 \times 10^{-308}$ ). We found no significant interaction between the  $s_t$  and  $y_{t-1}$  ( $-0.06 \pm 0.06$ ,  $z = -1.06$ ,  $p = 1$ ). Relative to placebo, S-ketamine caused a shift of  $y_t$  toward  $s_t$  ( $0.45 \pm 0.08$ ,  $z = 5.6$ ,  $p = 1.71 \times 10^{-7}$ ), with no significant effect on  $y_{t-1}$  ( $0.08 \pm 0.04$ ,  $z = 2.39$ ,  $p = 0.13$ ). We found no significant three-way-interaction (drug  $\times s_t \times y_{t-1}$ ,  $-0.07 \pm 0.08$ ,  $z = -0.9$ ,  $p = 1$ ).

**B.** This panel shows the data from panel (A) separately for times  $t$  where the HMM identified the mode of perceptual inference as external (left panels) or internal (right panels). When the mode of perceptual processing was added to the prediction of  $y_t$  from  $s_t$  and  $y_{t-1}$ , the effect S-ketamine (red) vs. placebo (blue) on  $s_t$  disappeared ( $0.24 \pm 0.11$ ,  $z = 2.13$ ,  $p = 0.53$ ). Instead, changes in the balance between  $s_t$  and  $y_{t-1}$  were loaded onto fluctuations between external and internal mode, which caused perception to shift away from external inputs  $s_t$  ( $-4.23 \pm 0.21$ ,  $z = -20.01$ ,  $p = 7.54 \times 10^{-88}$ ) and toward previous experiences  $y_{t-1}$  ( $0.78 \pm 0.09$ ,  $z = 8.64$ ,  $p = 8.81 \times 10^{-17}$ ).

**C.** Here, we plot the weights from the GLM  $y_t = \beta_S \times s_t + \beta_P \times y_{t-1} + \beta_B \times 1$ , alongside the balance between external inputs and previous experiences  $\Delta_{S-P} = \beta_S - \beta_P$  during external and internal mode. Colors indicate S-ketamine (red) and placebo (blue).  $\beta_S$ , the weight associated with the external input  $s_t$ , was positive in external mode, but reduced to zero in internal mode ( $-3.55 \pm 0.23$ ,  $T(81) = -15.44$ ,  $p = 4.78 \times 10^{-24}$ ). We found no additional effect of S-ketamine (red) versus placebo (blue;  $-0.25 \pm 0.23$ ,  $T(81) = -1.1$ ,  $p = 1$ ) and no significant interaction ( $0.21 \pm 0.33$ ,  $T(81) = 0.65$ ,  $p = 1$ ).  $\beta_B$ , the weight associated with the constant response bias  $b$  toward rightward rotation, was not different from zero ( $\beta_B = 0.04 \pm 0.11$ ,  $T(98.36) = 0.31$ ,  $p = 1$ ). We found no effect of drug ( $-0.11 \pm 0.14$ ,  $T(81) = -0.74$ ,  $p = 1$ ) or mode ( $-0.02 \pm 0.14$ ,  $T(81) = -0.12$ ,  $p = 1$ ) on the bias weight  $\beta_B$ .  $\beta_P$ , the weight associated with the previous percept  $y_{t-1}$  was not modulated by S-ketamine ( $-0.22 \pm 0.26$ ,  $T(81) = -0.87$ ,  $p = 1$ ) or mode ( $-0.75 \pm 0.26$ ,  $T(81) = -2.92$ ,  $p = 0.29$ ). There was no significant interaction between drug and mode with respect to  $\beta_P$  ( $0.35 \pm 0.36$ ,  $T(81) = 0.97$ ,  $p = 1$ ). The balance  $\Delta_{S-P}$  between external inputs and internal predictions was determined

528 by mode ( $2.8 \pm 0.29$ ,  $T(81) = 9.5$ ,  $p = 5.22 \times 10^{-13}$ ), with no significant effect of S-ketamine  
529 ( $0.03 \pm 0.29$ ,  $T(81) = 0.1$ ,  $p = 1$ ) and no interaction ( $0.14 \pm 0.42$ ,  $T(81) = 0.34$ ,  $p = 1$ ).

### 7.3 Supplemental Figure S3

#### Supplemental Figure S3. Extended data on external and internal mode in Scz patients and healthy controls (related to Figure 2E-H).

**A.** Here, we show psychometric curves (percept  $y_t$  versus input  $s_t$ ) in patients (red) and controls (blue). The plot separates times  $t$  for which the previous experience was leftward rotation ( $y_{t-1} = -1$ , upper panel) and rightward rotation ( $y_{t-1} = +1$ , lower panel). Perception was driven by  $s_t$  ( $\beta_S = 2.77 \pm 0.11$ ,  $z = 24.85$ ,  $p = 2.18 \times 10^{-135}$ ) and  $y_{t-1}$  ( $\beta_P = 1.5 \pm 0.03$ ,  $z = 58.2$ ,  $p < 2.2 \times 10^{-308}$ ), with no significant interaction between  $s_t$  and  $y_{t-1}$  ( $-5.41 \times 10^{-3} \pm 0.11$ ,  $z = -0.05$ ,  $p = 1$ ). Patients were more sensitive to  $s_t$  ( $0.75 \pm 0.15$ ,  $z = 4.96$ ,  $p = 5.6 \times 10^{-6}$ ). We found no significant three-way-interaction (group  $\times s_t \times y_{t-1}$ ,  $-0.37 \pm 0.15$ ,  $z = -2.45$ ,  $p = 0.11$ ).

**B.** This panel shows the data from panel (A) separately for times  $t$  where the HMM identified the mode of perceptual inference as external (left panels) or internal (right panels). When the mode of perceptual processing was added to the prediction of  $y_t$  from  $s_t$  and  $y_{t-1}$ , the difference between patients (red) and controls (blue) in the effect of  $s_t$  on  $y_t$  disappeared ( $-0.02 \pm 0.22$ ,  $z = -0.08$ ,  $p = 1$ ). Instead, changes in the balance between  $s_t$  and  $y_{t-1}$  were loaded onto fluctuations between external and internal mode, which caused perception to shift away from external inputs  $s_t$  ( $-3.47 \pm 0.29$ ,  $z = -11.95$ ,  $p = 1.01 \times 10^{-31}$ ) and toward previous experiences  $y_{t-1}$  ( $0.5 \pm 0.07$ ,  $z = 6.85$ ,  $p = 1.15 \times 10^{-10}$ ).

**C.** Here, we plot the weights from the GLM  $y_t = \beta_S \times s_t + \beta_P \times y_{t-1} + \beta_B \times 1$ , alongside the balance between external inputs and previous experiences  $\Delta_{S-P} = \beta_S - \beta_P$  during external and internal mode. Colors indicate the group (patients in red, controls in blue).  $\beta_S$ , the weight associated with the external input  $s_t$ , was positive in external mode, but reduced to zero in internal mode ( $-2.19 \pm 0.24$ ,  $T(44) = -9.13$ ,  $p = 4.07 \times 10^{-11}$ ). We found no additional effect of group ( $-0.11 \pm 0.37$ ,  $T(87.69) = -0.3$ ,  $p = 1$ ) and no significant interaction ( $-0.25 \pm 0.34$ ,  $T(44) = -0.74$ ,  $p = 1$ ).  $\beta_B$ , the weight associated with the constant response bias  $b$  toward rightward rotation, was not different from zero ( $0.05 \pm 0.18$ ,  $T(1.62 \times 10^{-8}) = 0.29$ ,  $p = 1$ ). We found no effect of group ( $-0.09 \pm 0.25$ ,  $T(1.62 \times 10^{-8}) = -0.37$ ,  $p = 1$ ). There was a trend for a positive effect of internal mode ( $0.6 \pm 0.24$ ,  $T(88) = 2.47$ ,  $p = 0.06$ ) on the bias weight  $\beta_B$ .  $\beta_P$ , the weight associated with the previous percept  $y_{t-1}$ , was reduced in internal mode ( $-0.75 \pm 0.26$ ,  $T(88) = -2.92$ ,  $p = 0.02$ ), but not modulated by group ( $0.17 \pm 0.32$ ,  $T(9.88 \times 10^{-10}) = 0.54$ ,  $p = 1$ ). There was no significant interaction between group and mode with respect to  $\beta_P$  ( $0.11 \pm 0.36$ ,  $T(88) = 0.3$ ,  $p = 1$ ). The balance  $\Delta_{S-P}$  between external inputs and internal predictions was determined by mode ( $1.44 \pm 0.33$ ,  $T(81) = 9.5$ ,

<sup>564</sup>  $p = 3.39 \times 10^{-4}$ ), with no significant effect of group ( $0.28 \pm 0.54$ ,  $T(87.97) = 0.52$ ,  $p = 1$ )  
<sup>565</sup> and no interaction ( $0.36 \pm 0.47$ ,  $T(44) = 0.76$ ,  $p = 1$ ).

## 7.4 Supplemental Figure S4

### Supplemental Figure S4. RT and bimodal inference in Scz patients and controls.

**A.** RT were non-uniformly distributed across the inter-overlap interval ( $D = 0.22$ ,  $p < 2.2 \times 10^{-308}$ , one-sample Kolmogorov-Smirnov test against uniformity) in patients (red) and controls (blue). This confirmed that changes in perception were aligned with the overlapping configurations of the stimulus.

**B.** RT did not differ between patients (red) and controls (blue;  $-0.07 \pm 0.08$ ,  $T(66.96) = -0.87$ ,  $p = 1$ ). We found no quadratic relationship between RT and  $s_t$  ( $-3.54 \pm 2.34$ ,  $T(5.33 \times 10^3) = -1.51$ ,  $p = 1$ ).

**C.** We found no effect of mode on RT ( $0.03 \pm 0.04$ ,  $z = 4.89 \times 10^3$ ,  $p = 0.76$ ).



## 7.5 Supplemental Figure S5

### Supplemental Figure S5. Scores and Questionnaires.

**A.** Responses to Q1 (*How awake do you feel?*) indicated that participants felt more tired under S-ketamine (red) than placebo (blue;  $-1.53 \pm 0.6$ ,  $z = -2.57$ ,  $p = 0.04$ ), with no significant effect of time or a between-factor interaction. Responses to Q2 (*How intoxicated do you feel?*) indicated that participants felt more intoxicated under S-ketamine ( $3.32 \pm 1.44$ ,  $z = 2.3$ ,  $p = 0.09$ ), with no significant effect of time or a between-factor interaction. Responses to Q3 (*How nervous do you feel?*) revealed no effect of S-ketamine ( $-3.01 \pm 2.62$ ,  $z = -1.15$ ,  $p = 1$ ), time, nor a significant between-factor interaction. CADSS scores were elevated under S-ketamine ( $1.01 \pm 0.34$ ,  $T(185.32) = 2.99$ ,  $p = 0.01$ ) with a borderline trend for an increase over time ( $0.09 \pm 0.04$ ,  $T(185.61) = 2.24$ ,  $p = 0.1$ ) and no significant between-factor interaction.

**B.** Q1-3 and CADSS scores were collected after blocks 1, 3, 6 and 9. To assess how the mode of perceptual inference was linked to dissociative symptoms, we separated the participants ratings according to the mode that dominated perception at the very end of the preceding block. While controlling the effect of S-ketamine (red) vs placebo (blue), we found that external mode increased dissociative symptoms ( $1.05 \pm 0.54$ ,  $T(208.05) = 1.95$ ,  $p = 0.05$ ), but had no effect on wakefulness (Q1), subjective intoxication (Q2) or nervousness (Q3).

**C.** 5-ASC scores were elevated under S-ketamine (red) relative to placebo (blue;  $4.89 \pm 1.59$ ,  $T(27.14) = 3.08$ ,  $p = 9.33 \times 10^{-3}$ ).

**D.** Neither PDI, CAPS, nor 5-ASC scores were predictive of the probability of external mode (shown separately for S-ketamine in red and placebo in blue).

**E.** Stereodisparity thresholds were not predictive of the probability of external mode ( $-28.73 \pm 781.1$ ,  $z = -0.04$ ,  $p = 0.97$ ). Thresholds did not differ between S-ketamine (red) and placebo (blue;  $W = 102$ ,  $p = 0.66$ ).

**F.** Neither PDI, CAPS (patients in red and controls in blue), nor the PANSS items P1 (delusions) or P3 (hallucinations, patients only) predicted the probability of external mode.

**G.** In patients (red) and controls (blue), stereodisparity thresholds were not predictive of the probability of external mode ( $-1.88 \pm 2.05$ ,  $z = -0.92$ ,  $p = 1$ ). Thresholds did not differ between groups ( $V = 976$ ,  $p = 0.52$ ).

606 **7.6 Supplemental Table S1**

RESOURCE	SOURCE	IDENTIFIER
<b>Deposited data &amp; code</b>		
Analyzed data & custom code	<a href="https://github.com/veithweilnhammer/modes_ketamine_scz">https://github.com/veithweilnhammer/modes_ketamine_scz</a>	N/A
<b>Software</b>		
<b>Matlab</b>	<a href="https://www.mathworks.com/">https://www.mathworks.com/</a>	RRID:SCR_001622
Psychtoolbox 3	<a href="http://psychtoolbox.org/">http://psychtoolbox.org/</a>	RRID:SCR_002881
<b>R</b>	<a href="http://www.r-project.org/">http://www.r-project.org/</a>	RRID:SCR_001905
RStudio	<a href="https://www.rstudio.com/">https://www.rstudio.com/</a>	RRID:SCR_000432
lme4, afex, statConfR, ggplot2, ggridges, gridExtra, tidyr, plyr, readxl	<a href="http://cran.r-project.org/">http://cran.r-project.org/</a>	RRID:SCR_003005
<b>Python 3</b>	<a href="http://www.python.org/">http://www.python.org/</a>	RRID:SCR_008394
Jupyter Notebook	<a href="https://jupyter.org/">https://jupyter.org/</a>	RRID:SCR_018315
numpy	<a href="http://www.numpy.org">http://www.numpy.org</a>	RRID:SCR_008633
pandas	<a href="https://pandas.pydata.org">https://pandas.pydata.org</a>	RRID:SCR_018214
SSM	<a href="https://github.com/lindermanlab/ssm">https://github.com/lindermanlab/ssm</a>	N/A

607 **Supplemental Table S1. Key resources.**

608 **7.7 Supplemental Table S2**

Scale	Scope	Condition	mean $\pm$ s.e.m.
<b>PDI</b> <sup>27</sup>	Delusion proneness	Global	46.22 $\pm$ 7.19
<b>CAPS</b> <sup>28</sup>	Hallucination proneness	Global	23 $\pm$ 5.05
<b>BPRS</b> <sup>29</sup>	Screen for psychotic illness	Global	0.64 $\pm$ 0.27
<b>5D-ASC</b> <sup>31</sup>	Altered states of consciousness	S-ketamine	7.11 $\pm$ 1.59
		Placebo	2.2 $\pm$ 0.75
<b>CADSS</b> <sup>23</sup>	Dissociation	S-ketamine	7.8 $\pm$ 0.33
		Placebo	6.43 $\pm$ 0.17
<b>Q1</b>	Wakefulness	S-ketamine	0.41 $\pm$ 0.03
		Placebo	0.48 $\pm$ 0.03
<b>Q2</b>	Intoxication	S-ketamine	0.29 $\pm$ 0.03
		Placebo	0.09 $\pm$ 0.02
<b>Q3</b>	Nervousness	S-ketamine	0.17 $\pm$ 0.02
		Placebo	0.13 $\pm$ 0.03
<b>Stereovision</b>	Disparity thresholds	S-ketamine	2.89 $\times 10^{-3}$ $\pm$ 6.18 $\times 10^{-4}$
		Placebo	2.75 $\times 10^{-3}$ $\pm$ 4.39 $\times 10^{-4}$

609 **Supplemental Table S2. Psychometric data for the S-ketamine experiment.**

610 **7.8 Supplemental Table S3**

Scale	Scope	Condition	mean $\pm$ s.e.m.
<b>PDI</b> <sup>27</sup>	Delusion proneness	Patients	$138.83 \pm 16.64$
		Controls	$21.87 \pm 5.75$
<b>CAPS</b> <sup>28</sup>	Hallucination proneness	Patients	$65.17 \pm 10.56$
		Controls	$7.13 \pm 2.2$
<b>P1</b>	Delusions	Patients	$3.83 \pm 0.39$
<b>P3</b>	Delusions	Patients	$3.35 \pm 0.44$
<b>Stereovision</b>	Disparity thresholds	Patients	$2.82 \times 10^{-3} \pm$
			$5.13 \times 10^{-4}$
		Controls	$3.46 \times 10^{-3} \pm$ $7.14 \times 10^{-4}$

611 **Supplemental Table S3. Psychometric data for Scz-control-study.**

## References

1. Fletcher, P. C. *et al.* Perceiving is believing: A Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience* **10**, 48–58 (2009).
2. Adams, R. A. *et al.* The computational anatomy of psychosis. *Frontiers in psychiatry* **4**, 47 (2013).
3. Sterzer, P. *et al.* The Predictive Coding Account of Psychosis. *Biological Psychiatry* **84**, 634–643 (2018).
4. Corlett, P. R. *et al.* Glutamatergic Model Psychoses: Prediction Error, Learning, and Inference. *Neuropsychopharmacology* **36**, 294–315 (2011).
5. Stein, H. *et al.* Reduced serial dependence suggests deficits in synaptic potentiation in anti-NMDAR encephalitis and schizophrenia. *Nature Communications* **11**, 1–11 (2020).
6. Murray, J. D. *et al.* Linking Microcircuit Dysfunction to Cognitive Impairment: Effects of Disinhibition Associated with Schizophrenia in a Cortical Working Memory Model. *Cerebral Cortex* **24**, 859–872 (2014).
7. Catts, V. S. *et al.* A quantitative review of the postmortem evidence for decreased cortical N-methyl-D-aspartate receptor expression levels in schizophrenia: How can we link molecular abnormalities to mismatch negativity deficits? *Biological Psychology* **116**, 57–67 (2016).
8. Wang, M. *et al.* NMDA Receptors Subserve Persistent Neuronal Firing during Working Memory in Dorsolateral Prefrontal Cortex. *Neuron* **77**, 736–749 (2013).
9. Nakao, K. *et al.* Schizophrenia-Like Dopamine Release Abnormalities in a Mouse Model of NMDA Receptor Hypofunction. *Schizophrenia Bulletin* **45**, 138–147 (2019).
10. Self, M. W. *et al.* Different glutamate receptors convey feedforward and recurrent processing in macaque V1. *Proceedings of the National Academy of Sciences* **109**, 11031–11036 (2012).
11. Castro-Alamancos, M. A. *et al.* Short-term synaptic enhancement and long-term potentiation in neocortex. *Proceedings of the National Academy of Sciences* **93**, 1335–1339 (1996).
12. Oorschot, M. *et al.* Temporal dynamics of visual and auditory hallucinations in psychosis. *Schizophrenia Research* **140**, 77–82 (2012).
13. Hermans, K. *et al.* Temporal dynamics of suspiciousness and hallucinations in clinical high risk and first episode psychosis. *Psychiatry Research* **290**, 113039 (2020).

14. Laquitaine, S. *et al.* A Switching Observer for Human Perceptual Estimation. *Neuron* **97**, 462–474.e6 (2018).
15. Roy, N. A. *et al.* Extracting the dynamics of behavior in sensory decision-making experiments. *Neuron* **109**, 597–610.e6 (2021).
16. Ashwood, Z. C. *et al.* Mice alternate between discrete strategies during perceptual decision-making. *Nature Neuroscience* **25**, 201–212 (2022).
17. Weilhhammer, V. *et al.* Sensory processing in humans and mice fluctuates between external and internal modes. *PLOS Biology* **21**, e3002410 (2023).
18. Albert, S. *et al.* A hierarchical stochastic model for bistable perception. *PLOS Computational Biology* **13**, e1005856 (2017).
19. Weilhhammer, V. *et al.* Psychotic Experiences in Schizophrenia and Sensitivity to Sensory Evidence. *Schizophrenia bulletin* **46**, 927–936 (2020).
20. Weilhhammer, V. *et al.* An active role of inferior frontal cortex in conscious experience. *Current Biology* **31**, 2868–2880.e8 (2021).
21. Muckli, L. *et al.* Contextual Feedback to Superficial Layers of V1. *Current Biology* **25**, 2690–2695 (2015).
22. Weilhhammer, V. *et al.* The Neural Correlates of Hierarchical Predictions for Perceptual Decisions. *The Journal of Neuroscience* **38**, 5008–5021 (2018).
23. Mertens, Y. L. *et al.* The Clinician-Administered Dissociative States Scale (CADSS): Validation of the German Version. *Journal of Trauma & Dissociation* **23**, 366–384 (2022).
24. Schmack, K. *et al.* Striatal dopamine mediates hallucination-like perception in mice. *Science (New York, N.Y.)* **372**, eabf4740 (2021).
25. Longden, E. *et al.* The Relationship Between Dissociation and Symptoms of Psychosis: A Meta-analysis. *Schizophrenia Bulletin* **46**, 1104–1113 (2020).
26. Harrison, P. J. *et al.* GRIN2A (NR2A): A gene contributing to glutamatergic involvement in schizophrenia. *Molecular Psychiatry* **28**, 3568–3572 (2023).
27. Peters, E. R. *et al.* Measurement of delusional ideation in the normal population: Introducing the PDI (Peters et al. Delusions Inventory). *Schizophrenia bulletin* **25**, 553–76 (1999).
28. Bell, V. *et al.* The Cardiff Anomalous Perceptions Scale (CAPS): A New Validated Measure of Anomalous Perceptual Experience. *Schizophrenia Bulletin* **32**, 366–377 (2006).

- 669 29. Overall, J. E. *et al.* The Brief Psychiatric Rating Scale. *Psychological Reports* **10**,  
670 799–812 (1962).
- 671 30. Brainard, D. H. The Psychophysics Toolbox. *Spatial vision* **10**, 433–6 (1997).  
672
- 673 31. Dittrich, A. *et al.* 5D-ABZ: Fragebogen zur Erfassung Aussergewöhnlicher Bewusst-  
674 seinszustände. *Eine kurze Einführung. PSIN Plus, Zürich* (1999).
- 675 32. Kay, S. R. *et al.* The positive and negative syndrome scale (PANSS) for schizophrenia.  
676 *Schizophrenia bulletin* **13**, 261–76 (1987).
- 677 33. Linderman, S. *et al.* SSM: Bayesian Learning and Inference for State Space Models.  
678 (2020).