Behavioral/Cognitive

Delusions and the Role of Beliefs in Perceptual Inference

Katharina Schmack,[1] Ana Gòmez-Carrillo de Castro,[1] Marcus Rothkirch,[1] Maria Sekutowicz,[1] Hannes Rössler,[1] John-Dylan Haynes,[2,3] Andreas Heinz,[1,3] Predrag Petrovic,[4,5] and Philipp Sterzer[1,2,3]

[1]Department of Psychiatry and Psychotherapy, Charité Universitätsmedizin Berlin, 10117 Berlin, Germany, [2]Bernstein Center for Computational Neuroscience, Charité Universitätsmedizin Berlin, 10117 Berlin, Germany, [3]Graduate School of Mind and Brain, Humboldt-Universität zu Berlin, 10117 Berlin, Germany, [4]Stockholm Brain Institute, 17177 Stockholm, Sweden, and [5]Department of Clinical Neuroscience, Karolinska Institutet, 17177 Stockholm, Sweden

Delusions are unfounded yet tenacious beliefs and a symptom of psychotic disorder. Varying degrees of delusional ideation are also found in the healthy population. Here, we empirically validated a neurocognitive model that explains both the formation and the persistence of delusional beliefs in terms of altered perceptual inference. In a combined behavioral and functional neuroimaging study in healthy participants, we used ambiguous visual stimulation to probe the relationship between delusion-proneness and the effect of learned predictions on perception. Delusional ideation was associated with less perceptual stability, but a stronger belief-induced bias on perception, paralleled by enhanced functional connectivity between frontal areas that encoded beliefs and sensory areas that encoded perception. These findings suggest that weakened lower-level predictions that result in perceptual instability are implicated in the emergence of delusional beliefs. In contrast, stronger higher-level predictions that sculpt perception into conformity with beliefs might contribute to the tenacious persistence of delusional beliefs.

Introduction

Folk wisdom has it that "seeing is believing," but what we see is in turn also influenced by what we believe. Recent theoretical advances provide an elegant framework for explaining this reciprocal interaction of perception and beliefs in terms of Bayesian inference and learning (Mumford, 1992; Kersten et al., 2004; Friston, 2005). Originating from Helmholtz's idea of unconscious inference (von Helmholtz, 1867), perception can be described as an inferential process that combines sensory signals fed forward along the cortical hierarchy with endogenous predictions fed back from higher hierarchical levels. These predictions derive from an internal model that represents the knowledge and beliefs about the outer world, and enable a stable and unitary perceptual experience despite noisy and ambiguous sensory information. Whenever predictions are violated by sensory input, the resulting prediction error signal drives learning by updating the internal model's predictions.

This framework has paved the way for comprehensive models of psychopathology that conceptualize delusions, which are implausible yet fixed beliefs, from a perceptual perspective (Fletcher and Frith, 2009; Corlett et al., 2010). Delusions can be explained by altered integration of endogenous predictions with sensory information (Hemsley, 2005), based on aberrant prediction error signals (Heinz, 2002; Kapur, 2003) that drive maladaptive belief formation (Fletcher and Frith, 2009). In other words, imprecise predictions render sensory events surprising and salient, and the cognitive effort to make sense of such aberrant salience results in the formation of delusional beliefs. However, this framework fails to account for a key feature of delusions, namely their tenacious persistence despite contradicting evidence. This tenacity of delusions would be most plausibly explained by an excessive influence of delusional beliefs on the perceptual interpretation of the sensory evidence, which would equate to increased rather than diminished predictive signaling (Corlett et al., 2009). We accommodate this apparent contradiction between the explanations for the formation and the fixity of delusions within a new model that draws on the hierarchical structure of the outlined framework. According to our model, weakened predictive signaling within sensory processing stages results in unstable sensory representations. On the one hand, this leads to the experience of a changing and unpredictable outer world, in which sensory events become overly salient, yielding the emergence of delusional misinterpretations. On the other hand, faced with the instability of sensory representations, perceptual inference relies more on predictions from higher-level nonsensory brain circuits that encode beliefs. Thereby perception is sculpted into conformity with delusional beliefs, accounting for the tenacious maintenance of delusions.

Here, we empirically tested this model of delusions in two behavioral and one functional magnetic resonance imaging (fMRI) experiment. Based on the idea that delusions constitute an extreme expression of a continuously distributed phenotype (Meehl, 1962; Freeman, 2006), we studied the relationship between the tendency toward delusional ideation and perception in healthy in-

Received April 26, 2013; revised June 12, 2013; accepted June 14, 2013.

Author contributions: K.S., P.P., and P.S. designed research; K.S., A.G.-C.d.C., M.R., M.S., and H.R. performed research; J.-D.H. contributed unpublished reagents/analytic tools; K.S. analyzed data; K.S., A.H., P.P., and P.S. wrote the paper.

This work was supported by Deutsche Forschungsgemeinschaft (STE-1430/2-1). We thank Jordi Rubi for the construction of the fMRI-compatible dichoptic stimulation system.

The authors declare no competing financial interests.

Correspondence should be addressed to Katharina Schmack, Klinik für Psychiatrie und Psychotherapie—CCM, Charité Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany. E-mail: katharina.schmack@charite.de.

DOI:10.1523/JNEUROSCI.1778-13.2013

 0270-6474/13/3313701-12$15.00/0

dividuals. Using ambiguous visual stimuli to maximize the need for perceptual inference, we tested two main hypotheses: (1) the tendency toward delusional ideation is associated with weakened sensory predictions, resulting in perceptual instability; and (2) the tendency toward delusional ideation is associated with a stronger influence of cognitive beliefs.

Materials and Methods

Participants

One hundred five healthy individuals (age 18–44, 53 female) participated in Behavioral Experiments 1 and 2; four additional participants were excluded from analysis due to technical problems in data collection and analysis (see below). An independent group of 20 participants (age 20–37 years, 11 female) underwent fMRI; two additional participants were excluded from analysis because of technical problems in data collection (see below).

All participants were naive regarding the purpose of the experiment, had normal or corrected-to-normal vision, and no Axis I psychiatric disorder (Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Axis I Disorders). Participants were paid €20 for their time and gave written informed consent after all the procedures and possible consequences were explained to them. The study was approved by the Ethics Committee of the Charité University Medicine Berlin.

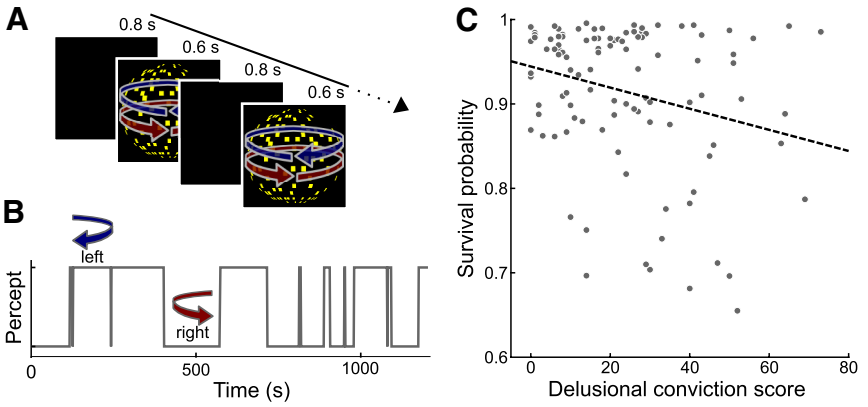

Figure 1. Delusional ideation and perceptual stability. ***A***, Schematic illustration of Behavioral Experiment 1. Sensory predictions were induced by repeated presentation of an ambiguous DK that can be perceived as a sphere rotating either leftward or rightward. The stimulus was presented repeatedly for 0.6 s interleaved by blank screens of 0.8 s duration. Upon each occurrence of the stimulus, participants reported the perceived rotation direction by button press. ***B***, Perceptual time course from one exemplary individual. Due to the stabilizing effect of endogenous predictions that are automatically built up during intermittent presentation of the ambiguous stimulus, participants tended to have the same percept across many successive presentation cycles. ***C***, Correlation between tendency toward delusional convictions and perceptual stability ($r = -0.26$, $p = 0.004$, product-moment correlation, p value based on 10,000 permutations). Tendency toward delusional convictions was measured with a validated questionnaire (Peters et al., 1999). Perceptual stability was calculated as the percept survival probability from one presentation cycle to the next. Higher values indicate higher perceptual stability. Each dot represents one participant. The dashed line illustrates the fitted regression line.

Measurement of delusional ideation

All participants completed the Peters et al. Delusion Inventory, a questionnaire designed to measure the tendency toward delusional ideation in the general population (Peters et al., 1999). The 40 items cover a wide range of delusion-like beliefs, including delusions of reference, control, and persecution. For each item, dimensional ratings assessing associated distress, preoccupation, and conviction are included in addition to a dichotomous absence–presence statement, resulting in four distinct subscores. To reduce the number of statistical tests, we adopted a two-step approach to relate the tendency toward delusional ideation to the behavioral effect of sensory predictions and higher-level beliefs in perceptual inference. First, an overall score representing the sum of the four subscores was correlated with the dependent variables derived from Behavioral Experiments 1 and 2 (survival probability and belief-induced bias, respectively; see below). Upon significance, the most predictive subscale(s) were identified in a stepwise regression analysis using forward selection and a criterion of $p < 0.05$.

Visual stimulation

In all experiments, visual stimuli were dot-kinematograms (DKs) that are perceived as a sphere rotating in depth around a vertical axis. Stimuli were presented using Matlab (MathWorks) and Cogent 2000 toolbox (http://www.vislab.ucl.ac.uk/cogent.php). To produce stereoscopic vision during the experimental induction of perceptual beliefs (see below), stimuli were presented dichoptically through a mirror stereoscope in Behavioral Experiments 1 and 2, or an MRI-compatible system (Schurger, 2009) in the fMRI experiment. Our DK stimulus was an orthographic projection of a sphere rotating around a vertical axis (Behavioral Experiments 1 and 2: diameter, 4.1° of visual angle; rotation speed, 0.167 revolutions/s; fMRI experiment: diameter, 5.1°, rotation speed, 0.056 revolutions/s). It consisted of 450 randomly distributed yellow square "dots" (Behavioral Experiments 1 and 2, maximum 0.2 × 0.2°; fMRI experiment, maximum 0.1 × 0.1°), moving coherently leftward or rightward on a black background with a central fixation cross and framed by a white square. Animation frames were updated every 40 ms, and a sequence of 8 s was looped repeatedly to produce continuous motion. Dot lifetime was 1 s on average.

Depending on the experiment and the experimental run, the rotation direction of the sphere was either ambiguous or unambiguous. The ambiguous sphere rotating around the vertical axis consisted of two identical DKs presented to each eye and evoked the two possible, mutually exclusive percepts of leftward or rightward rotation. To produce the unambiguous sphere used for the experimental induction of perceptual beliefs (see below), two slightly different DKs were displayed representing two different perspectives. The maximal offset between corresponding dots presented to the two eyes was 0.5° of visual angle. This interocular disparity minimizes the ambiguity of rotation direction. For naive observers, the ambiguous and nonambiguous stimuli are nearly indistinguishable. To mimic the spontaneous perceptual alternations that occur during continuous viewing of the ambiguous sphere, the unambiguous sphere alternated between both rotation directions with an overall rate comparable to the subject's switch rate but with different dominance times (80 vs 20% on average).

Experimental design

Behavioral Experiment 1. This experiment was aimed at measuring the influence of sensory predictions on perception to relate it to the tendency toward delusional ideation. Such endogenous predictions are automatically built up during repeated exposure to a stimulus and facilitate perceptual inference at each recurrence of the stimulus (Friston, 2005). In the case of ambiguous stimuli that are consistent with two possible, mutually exclusive perceptual interpretations, the incorporation of these predictions based on previous perceptual outcomes results in the stabilization of appearance. In other words, when an ambiguous stimulus is temporarily removed from view, the percept after reonset of the stimulus strongly tends to be the same as the last percept before stimulus removal (Orbach et al., 1963; Leopold et al., 2002). The stabilizing predictions are robust to interfering visual stimulation (Maier et al., 2003), encode low-level visual characteristics (Pearson and Brascamp, 2008), and can be disrupted by applying transcranial magnetic stimulation to functionally specialized visual areas (Brascamp et al., 2010), indicating that they are implemented at low levels of the cortical hierarchy, that is, within sensory cortices. Here, we used the survival probability of percepts across temporary stimulus removals to quantify in each individual participant the strength of lower-level sensory predictions in perceptual inference.

The ambiguous sphere stimulus (see Visual stimulation, above) was shown repeatedly for intervals of 0.6 s interleaved by blank screens of 0.8 s

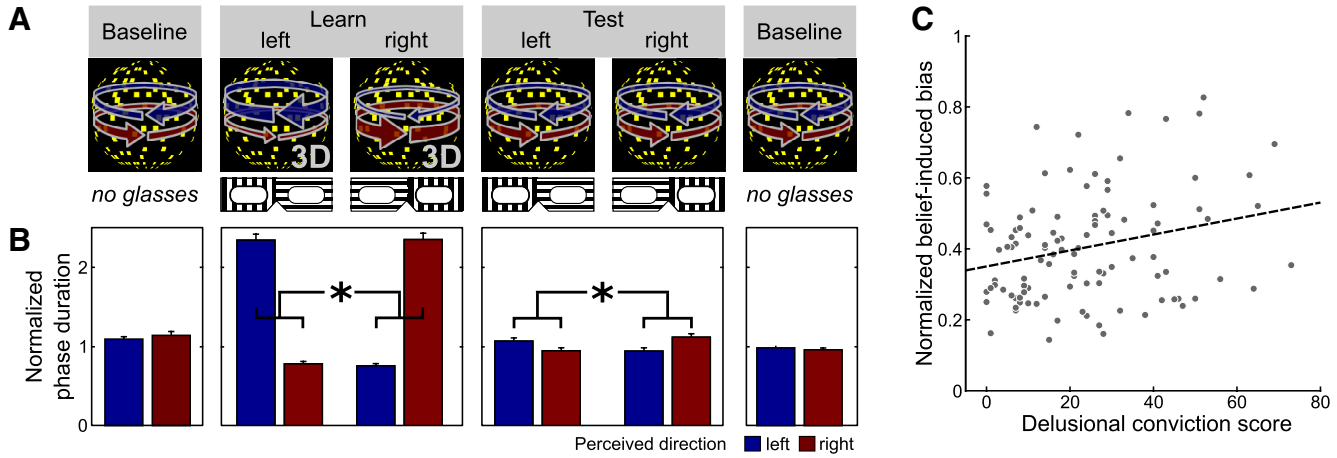

Figure 2. Delusional ideation and belief-induced bias. ***A***, Schematic illustration of Behavioral Experiment 2 and the fMRI experiment. Higher-level cognitive beliefs were induced by a placebo-like experimental manipulation. Participants viewed continuously a DK that is perceived as a rotating sphere and indicated changes in perceived rotation direction by button presses. In the initial and final baseline phases and in the test phase, the rotation direction of the sphere was ambiguous, yielding bistable perception alternating spontaneously between leftward and rightward rotation direction. In the learning phase, stereoscopic depth cues were surreptitiously added to the stimulus, which forced the sphere to rotate in one direction 80% of the time. In the learning and test phase, participants wore transparent glasses, which they believed to contain polarizing filters and to bias their perception toward one rotation direction, depending on the orientation of the glasses. ***B***, Effect of beliefs on reported perception (*$p < 0.001$, paired t test, p value based on 10,000 permutations). Bars show the mean phase duration of each percept normalized with respect to the mean phase duration in the baseline runs. Error bars denote SE. ***C***, Correlation between tendency toward delusional convictions and belief-induced bias ($r = 0.26$, $p = 0.004$, product-moment correlation, p value based on 10,000 permutations). Tendency toward delusional convictions was measured with a validated questionnaire (Peters et al., 1999). Belief-induced bias was calculated as the ratio of belief-congruent and belief-incongruent mean phase durations normalized with respect to the learning phase. Higher values indicate a stronger belief-induced bias. Each dot represents one participant. The dashed line illustrates the fitted regression line.

duration for an overall duration of 20 min (Fig. 1*A*). Participants indicated the rotation direction at each presentation cycle by button presses, choosing between leftward and rightward rotation. To quantify the strength of sensory predictions in perceptual inference, we calculated the survival probability of percepts from each stimulus presentation to the next. Thus, a higher survival probability indicates a more stable perception, reflecting stronger sensory predictions. One participant was excluded from further analysis because he did not report a single perceptual switch during the whole course of this experiment, rendering a true estimate of his survival probability impossible.

Behavioral Experiment 2. To test for the relation between the tendency toward delusional ideation and the effect of higher-level beliefs in perceptual inference, in Behavioral Experiment 2 we used a placebo-like manipulation to induce perceptual beliefs (Fig. 2*A*). Participants were informed that they would perform an experiment on depth perception. They now viewed the same sphere stimulus as in Behavioral Experiment 1 continuously and reported changes in perceived rotation direction while again maintaining visual fixation. During the initial baseline phase, the sphere was perceptually ambiguous, yielding bistable perception alternating spontaneously between rightward and leftward rotation. In a subsequent learning phase, we then induced perceptual beliefs using a placebo-like manipulation (Sterzer et al., 2008). Participants now wore glasses, which they believed to be polarizing filters. They were told that they would view the same ambiguous stimulus as in the baseline phase but that the polarizing glasses would bias their perception toward one rotation direction. These glasses would contain two different filters so that one eye would be only reached by horizontally polarized light and the other eye only by vertically polarized light. Due to stereoscopic vision induced by these glasses, participants' perception would be biased toward one rotation direction. The rotation direction would depend on which eye looked through which filter and the orientation of the glasses would be manually reversed between runs. In reality, the glasses were completely transparent, but the stimulus was rendered unambiguous by dichoptically induced disparity depth cues, which forced the sphere to rotate in one direction for 80% of the time. The glasses and the predominant rotation direction were reversed half way through the learning phase. In the following test phase, we then probed whether the perceptual beliefs induced by the glasses influenced perception of the ambiguous sphere. Participants maintained the belief that the glasses would bias their perception but were now again presented with the completely ambiguous stimulus used in the baseline phase. The experiment ended with a final baseline phase, in which participants viewed the ambiguous sphere again without the transparent glasses.

During each experimental run, the stimulus was presented continuously for the duration of the run (240 s). The experiment started with two initial baseline runs, during which the ambiguous sphere was presented, followed by two unambiguous learning runs, two ambiguous test runs, and two final ambiguous baseline runs. In the learning and test runs, subjects wore the transparent glasses, which they believed to contain polarizing filters. The actual (learning phase) or expected (test phase) dominant rotation directions in each run were contingent on the orientation of the transparent glasses and were switched between the runs so that there was one learning run per orientation, and one test run per orientation, respectively. Before each run, the experimenter made sure that participants were aware of the actual orientation of the glasses by asking the subjects which filter was on which eye. The association between orientations and rotation directions was counterbalanced across participants to control for potential systematic effects of the glasses on perception. The order of dominant directions across runs was pseudorandomized and balanced across participants, so that potential systematic effects of the dominant direction in the second learning run on the following test runs (e.g., adaptation) were cancelled out. Participants indicated perceptual transitions by button presses choosing between three response options: leftward rotation, rightward rotation, and uncertain perceptual state. As uncertain perceptual states only occurred at a negligible amount of the total time (mean, 3.2 ± 0.6% SEM), they were discarded from further analysis.

Debriefing after the experiment revealed that most participants (83 of 105) had not noticed a difference between the visual stimulus in the learning phase and in the test phase. Furthermore, the majority (16 of 22) of those who had noticed a difference, had attributed it to subjective factors, such as habituation or fatigue. Only six individuals reported that they had suspected a physical stimulus manipulation and had therefore not believed in an effect of the placebo glasses on their perception.

For each participant, we calculated the belief-induced bias by normalizing the ratio of belief-congruent and belief-incongruent mean phase duration with respect to the ratio from the learning phase. Higher values indicate a stronger belief-induced bias, hence a stronger effect of beliefs

in perceptual inference. Three participants of Behavioral Experiment 2 were excluded from further analysis because of button malfunction.

fMRI experiment. To examine the neural correlates of the effect of higher-level beliefs in perceptual inference, an independent group of participants underwent the placebo-like manipulation of beliefs used in Behavioral Experiment 2 during fMRI scanning. The design of the fMRI experiment closely resembled the design of Behavioral Experiment 2 with minor modifications in experimental run duration (212 s) and number (4 initial ambiguous baseline runs, 2 unambiguous learning runs, 4 ambiguous test runs, 4 final baseline runs). Similar to Behavioral Experiment 2, debriefing after the experiment suggested that all participants of the fMRI experiment had believed in an effect of the placebo glasses on their perception: the majority (15 of 20) reported that they had not noticed a difference between the visual stimulus in the learning phase and in the test phase; and those who had noticed a difference had attributed it to subjective factors, such as habituation or fatigue.

The fMRI experiment also included two additional localizer runs of 235 s each. These were aimed at the identification of areas in visual cortex responsive to visual motion. To identify the voxels that maximally responded to the stimulus used in the main experiment, the same moving ambiguous sphere stimulus was alternated with static images of the sphere stimulus. Each localizer run comprised five stimulation blocks of 13.5 s duration per condition (moving dots and stationary dots). The blocks were separated by rest periods of 9.0 s, in which only the fixation cross and the fusion square were shown.

fMRI scanning was performed on a 3 T scanner (TRIO, Siemens) equipped with a 12-channel head coil using a T2*-weighted two-dimensional gradient-echo echo-planar imaging sequence (TR, 2260 ms; TE, 25 ms; flip angle, 90°; matrix size, 64 × 64; FOV, 192 mm; voxel size, 3 × 3 × 3 mm). Thirty-eight slices parallel to the calcarine sulcus were collected, covering the whole brain (slice thickness, 2.5 mm; interslice gap, 0.5 mm). Data were acquired in 14 runs, each comprising 94 volumes. Additionally, two functional localizer runs of 104 volumes each were acquired. For anatomical reference, a structural image was collected using a T1-weighted three-dimensional magnetization prepared rapid gradient-echo sequence (TR, 1900 ms; TE, 2.52 ms; matrix size, 256 × 256; FOV, 256 mm; flip angle, 9°; voxel size, 1 × 1 × 1 mm).

Two participants were excluded from further analysis of the fMRI data because of excessive head movement (several shifts of >1.5 mm between 2 successive scans) and monocular vision associated with a lack of stereoscopic depth perception, which impeded the induction of perceptual beliefs during the learning phase. As in Behavioral Experiment 2, uncertain perceptual states only occurred at a negligible amount of time (mean, 1.6 ± 0.6% SEM), and were therefore discarded from further analysis.

Control for eye movements. Participants were instructed to maintain visual fixation. To control for an effect of fixation quality on perception, eye movements during the behavioral experiments were recorded using a video-based eye tracker (MK2 High-Speed Camera, Cambridge Research Systems; 100 Hz). To calculate a measure of fixation quality, the one-dimensional time courses of gaze positions were linearly detrended. Gaze positions were then transformed into a two-dimensional histogram of their Helmholtz coordinates of a resolution 1 × 1 arcmin and smoothed with a two-dimensional Gaussian kernel [full width at half maximum (FWHM), 6 arcmin]. Fixation quality was calculated as the area included by the contour line encompassing 95% of the eye positions. A smaller contour line area corresponds to better fixation, whereas a larger area is indicative of larger overall eye movements. For some participants, fewer than half of gaze positions could be successfully tracked. Such participants were excluded from these control analyses.

fMRI data analysis

Multivoxel pattern analysis overview. To test the hypothesis that experimentally induced beliefs altered the neural correlates of perception in visual cortex, we used multivoxel pattern analysis (MVPA), as described previously (Haynes and Rees, 2005; Brouwer and van Ee, 2007). To remove nonsteady-state effects caused by T1 saturation and stimulus-onset effects, the first four scans were discarded. Preprocessing was performed with SPM8 (www.fil.ion.ucl.ac.uk/spm). Images were slice-time corrected and spatially realigned to the first image. Importantly, no spatial normalization or spatial smoothing was applied at this stage. Then the raw blood-oxygen-level-dependent (BOLD) signal was extracted for every scan. The resulting signal time courses were linearly detrended and normalized to vectors of unit length. Each scan was then assigned to one of the two percepts (leftward rotation and rightward rotation) according to the perceptual dominance time course reported by the subjects. For convenience, perceptual states indicated as uncertain were treated as the preceding perceptual state, as they only occurred at an average of 1.6 ± 0.6% of the total time. To account for the hemodynamic delay (+4000 ms) and the time between a perceptual switch and the button press (−1000 ms) the perceptual time courses were shifted +3000 ms in time. The resulting label contained a 1 for each scan corresponding to a leftward rotation ("left") and −1 for each scan corresponding to a rightward rotation ("right").

Classification was performed with a linear support vector machine (SVM; Cortes and Vapnik, 1995) and the implementation LIBSVM (http://www.csie.ntu.edu.tw/%7ecjlin/libsvm/). A training subset of the fMRI and label data was used to train the classifier. The trained classifier then generated predictions of perceptual states for each time point from an independent test subset of fMRI data that was drawn from independent experimental runs (see below for details). Because of sampling inaccuracies and the stochastic nature of perceptual dominance time courses, the number of volumes within the training dataset labeled as "right" did not necessarily equal the number of volumes labeled as "left." Imbalanced training datasets are known to decrease the performance of an SVM classifier by biasing its predictions toward the label with more occurrences in the training data. We corrected for this bias by using different cost parameters c_r and c_l for each class (Veropoulos et al., 1999), after having confirmed the effectiveness of this correction in a surrogate dataset (see below). The sum of the cost parameters was fixed at 1. The class imbalance has also to be taken into consideration when assessing the performance of a classifier. Since standard accuracy measures do not account equally for positive and negative errors and thus tend to overestimate classifier performance when there is class imbalance in the data, we applied a balanced accuracy measure (Velez et al., 2007) to prevent such errors. We calculated the specific accuracy for each class ("left" and "right") by dividing the number of correctly predicted scans (e.g., predicted "right" and perceived "right") by the number of all scans in this class (e.g., perceived "right"). The arithmetic mean of the two class-specific accuracies gave the balanced prediction accuracy.

To test whether the application of different cost parameters c_r and c_l for each class improves classifier performance in an unbalanced dataset, we generated a multivariate surrogate dataset. To mimic the class imbalance observed in our real data, each participant's real class label derived from his or her perceptual time course in the baseline phase was used for data creation. For each of the 720 scans, a 2000-dimensional vector was randomly drawn from one of two multivariate normal distributions, depending on the class membership of the scan. The mean vectors of the two distributions were separated by an Euclidean distance of 2; the covariance matrices of both distributions were the unity matrix. The resulting 720 × 2000 (scans × voxels) surrogate data matrix was used to train and test two SVM classifiers: one "standard classifier" using the same cost parameter c for both classes and one "weighted classifier" using different cost parameters c_r and c_l for each class. Using a runwise cross-validation scheme, the balanced decoding accuracy was calculated for each classifier. This procedure was repeated 10 times and the decoding accuracies were averaged separately for each classifier across these repetitions, yielding two values for each participant. The comparison of these values showed a significant advantage for the "weighted classifier" compared with the "standard classifier" (66.39 vs 64.96%, $\mu = 1.43\%$, $\sigma = 0.27\%$, $t_{(19)} = 23.45$, $p < 0.001$, paired t test), confirming the effectiveness of the use of weighted cost parameters.

MVPA: searchlight analysis. Before our critical MVPA, we verified in a whole-brain "searchlight" (Kriegeskorte et al., 2006) approach that we could identify the neural correlates of perceived rotation direction during viewing of a constant physical stimulus (Fig. 3*A*). For each voxel v_i, we defined a small spherical cluster (radius, 10 mm) centered on v_i. The fMRI data from this cluster were then used for training and testing

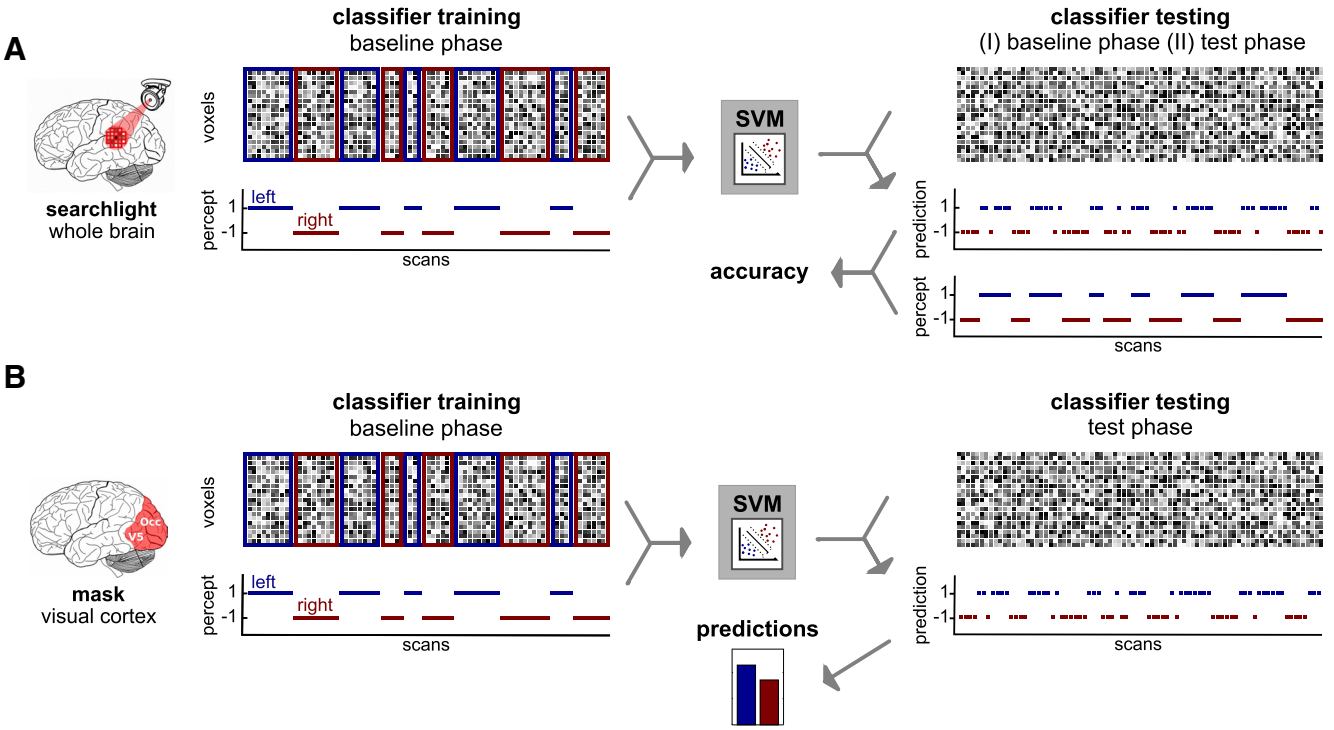

Figure 3. Schematic illustration of the MVPA. ***A***, MVPA searchlight analysis. For a given voxel in the brain, a local cluster of the surrounding voxels was defined. For each of the voxels in the cluster, the raw fMRI signal time courses were extracted, yielding a pattern activation matrix for each run, in which each line represents one voxel and each column one scan. Each column of this matrix was labeled according to the perceptual time course. The labeled pattern activation matrices from a subset of runs (I, 7 of 8 baseline runs; II, all baseline runs) were used to train a classifier (SVM). The trained classifier was then tested on an unlabeled pattern activation matrix from another independent subset of runs (I, remaining baseline run; II, all test runs) and predicted a perceptual state for each scan. Prediction accuracy was derived from the comparison of the predicted and the real perceptual time course. This procedure was repeated for every voxel in the brain, resulting in voxelwise accuracy maps. ***B***, MVPA in visual cortex. From a mask comprising occipital lobe and functionally defined hMT/V5, selection was made of the 2000 voxels that were most informative about the perceived rotation direction in the training phase as defined by a *t* test. For each of these voxels, the raw fMRI signal time courses were extracted, yielding a pattern activation matrix for each run, in which each line represents one voxel and each column one scan. Each column of this matrix was labeled according to the perceptual time course. These labeled pattern activation matrices from all baseline runs were used to train a classifier (SVM). The trained classifier was then tested on an unlabeled pattern activation matrix from the test runs and predicted a perceptual state for each scan. The summed predictions for each of both rotation directions were then used to calculate the predicted effect of beliefs.

the classifier. This procedure was repeated for every voxel in the brain, resulting in a voxelwise accuracy map for each subject. These maps were then spatially normalized to the Montreal Neurological Institute (MNI) template and spatially smoothed with a Gaussian kernel (FWHM, 8 mm). In a paired *t* test, these maps were compared with chance level (50%). The obtained *t* maps were used for statistical inference and display purposes. Results were considered statistically significant at $p < 0.05$, familywise error-corrected (FWE-corrected) for multiple comparisons across the whole brain or within a priori regions of interest in visual cortex (see below).

Two "searchlight" analyses were conducted. We first asked whether decoding of perceived rotation direction was possible in the baseline phase, in which participants viewed the ambiguous sphere in the absence of perceptual beliefs. The data from the baseline phase only were used in a runwise cross-validation scheme. We then probed whether rotation direction could also be decoded from scans acquired during the test phase, when perception was biased by the participant's beliefs. The classifier was trained on all eight runs from the baseline phase and tested on all four runs from the test phase.

MVPA: visual cortex. The main purpose of MVPA was to quantify the effect of beliefs on perception at the neural level (Fig. 3*B*). Analysis was spatially restricted to visual cortex, as we were mainly interested in decoding from sensory stimulus representations. For this purpose, a mask that consisted of occipital cortex and the human motion complex hMT/V5 on the occipitotemporal junction was generated for each participant, as the analyses were conducted with spatially unnormalized data. First, an anatomical mask for the occipital lobe from the Wake Forest University (WFU) PickAtlas (http://www.nitrc.org/projects/wfupickatlas/) was warped to the participants' brains by applying the inverse transformation matrix resulting from the spatial normalization of the structural scan to the MNI template. Second, hMT/V5 was identified on the basis of the contrast "motion > stationary" from the localizer runs for each participant separately. A bilateral cluster on the occipitotemporal junction was identified at a threshold of $p < 0.05$ (uncorrected) and manually delineated in MRIcron (http://www.cabiatl.com/mricro/mricron/index.html). The voxels for the MVPA were selected from these masks of visual cortex, which contained between 3928 and 5663 voxels.

First, we compared the prediction accuracy between the baseline phase and the test phase. The data used for training and testing in the searchlight analyses differed in amount and proximity in time, which could imply differences in prediction accuracy due to generalization problems. To correct for this phenomenon, we matched training and test data as closely as possible: the first four baseline runs were used for training a classifier, which was then tested separately on the last four baseline runs and the four test runs. The absolute decoding accuracies were then compared in a paired *t* test. We then tested whether the content-specific representations of perceived rotation directions in visual cortex were biased by perceptual beliefs. In this case, a classifier that predicts the reported perceptual state should predict more expected percepts than unexpected percepts. To minimize the risk of spurious findings, classifier performance was optimized with regard to the prediction accuracy using the baseline runs only. First, different voxel selection methods were compared. Different numbers of voxels (200, 500, 1000, 2000, and 3000 voxels) were selected by different procedures (most activated voxels during the localizer runs, most motion-sensitive voxels during the localizer runs, and voxels with the highest *t* values in a *t* test comparing rightward rotation to leftward rotation in the training dataset). The best performance was achieved with 2000 voxels selected by a *t* test comparing

rightward rotation to leftward rotation. Another issue that was addressed in the optimization procedure was the temporal shift of the class label of each scan ("left" or "right"; see above). Because both hemodynamic delay and time between a perceptual switch and the button press might vary between participants, different labels with time shifts ranging from 0 to 6000 ms in steps of 500 ms were generated and used for classification. The shift that yielded the highest runwise cross-validation accuracy was then chosen for each participant. The optimized classifier reached a mean runwise cross-validation accuracy of 63.02% during the baseline. It is important to note that the optimization procedure was performed exclusively on the baseline runs and thus could not impose any systematic bias on our critical analysis, i.e., the decoding of rotation directions in the independent test runs. The optimal parameters were then used for a classifier that was trained on all eight baseline runs and tested on all four test runs.

The predictions of this classifier were used to calculate the predicted effect of beliefs. For this purpose, the proportion of predictions for each class (e.g., "right" and "left") was corrected according to the class-specific cross-validation accuracy from the independent baseline runs. The rationale behind this correction was that a classifier with a prediction accuracy <100% arithmetically underestimates the difference between the true proportions of the classes. The extent of this underestimation error scales as a function of class-specific prediction accuracies. As the accuracies in the test phase did not differ from the baseline phase (see above), in the test phase the relation between predicted and perceived percepts for one class (e.g., "left") can be approximated by $L_p = acc_l \cdot L_s + (1 - acc_r) \cdot (1 - L_s)$, where L_p is the predicted proportion of "left" percepts, L_s is the true proportion of "left" percepts, acc_l and acc_r are the class-specific accuracies for "left" and "right" derived from testing the classifier on the independent baseline dataset. Thus, the true proportion of percepts perceived as "left" (and analogously for "right") was calculated as $L_s = (L_p - 1 + acc_r)/(acc_l - acc_r)$. The difference of the expected and the unexpected proportion gave the predicted effect of beliefs. Again, it is important to note that the values for the class-specific accuracies that were used for this correction were known from previous testing of the classifier on the independent baseline dataset, so that nonindependence errors in our critical analysis of the test runs were precluded.

Statistical parametric mapping analysis of belief-related activity. Following the idea that higher-level nonsensory brain circuits may be involved in generating the observed influence of beliefs on visual information processing, we used a standard univariate statistical approach to test for an association of beliefs with brain activity. fMRI data from the experimental learning and test runs were analyzed using statistical parametric mapping (SPM8). After discarding the first four volumes, images were slice-time corrected and spatially realigned to the first volume. Each participant's structural T1 image was coregistered to an individual mean EPI image. Transformation parameters were derived from normalizing the coregistered structural image to a MNI template, and the derived parameters were then applied to normalize the EPI volumes. Normalized images were smoothed with a Gaussian kernel (FWHM, 8 mm).

Images from the test runs were analyzed in an event-related manner using the general linear model. On the first level, belief-congruent and belief-incongruent perceptual switches were modeled separately as regressors of interest. Other button presses indicating uncertainty were included as regressors of no interest. To account for the period between a perceptual switch and a button press, 1000 ms were subtracted from the times of the button presses. The evoked hemodynamic responses to the estimated perceptual switches were modeled as stick functions convolved with the canonical hemodynamic response function implemented in SPM8 and its first temporal derivative. Additional regressors of no interest were the six movement parameters as well as one constant term per experimental run. For each participant, a contrast image that compared the parameter estimates for belief-congruent and belief-incongruent perceptual switches was computed for the learning and the test phase separately and used for group level inference.

The effect of beliefs on BOLD signal was assessed in a second-level analysis. For the learning and the test phase separately, the individual contrast images ("belief-congruent > belief-incongruent") were correlated with the behavioral belief-induced bias calculated as the ratio of belief-congruent and belief-incongruent mean phase duration from the test phase normalized with respect to the ratio from the learn phase. This analysis focused on orbitofrontal cortex (OFC), as previous reports have consistently implicated this region in mediating the effect of beliefs and expectations on sensory processing (Petrovic et al., 2002, 2005; Wager et al., 2004; Bar et al., 2006; Kveraga et al., 2007; Summerfield and Koechlin, 2008). OFC was specified by a mask from a publication-based probabilistic MNI atlas (http://neuro.imm.dtu.dk/services/jerne/brede/WOROI_685.html; access date March 12, 2011) and used as a binary mask at the threshold of 90% probability (417 voxels). Results were considered statistically significant at $p < 0.05$, FWE-corrected for multiple comparisons across all voxels within the OFC mask.

SPM correlation analysis of delusional ideation and functional connectivity. To test our prediction that in individuals prone to delusions, compared with individuals less prone to delusions, perceptual inference relies more on predictions from higher-level brain circuits that encode beliefs, we investigated the correlation between the tendency toward delusional ideation and functional connectivity between OFC and other brain regions. For this purpose, we analyzed psychophysiological interactions (PPIs) in SPM8, again separately for the learning and the test phase. PPI is defined as the change in contribution of one brain area to another with the experimental or psychological context (Friston et al., 1997). It computes whole-brain connectivity on a voxel-by-voxel basis between the time series of a seed region and the time series of all other voxels, modulated by a context stimulus.

The seed region in OFC was defined using a sphere with a radius of 20 mm centered on the group maximum from the analyses of belief-related activity. After preprocessing (see above), fMRI time series were extracted from individual peak voxels within this sphere ("belief-congruent > belief-incongruent") and deconvolved to generate the neuronal signal for the seed region (Gitelman et al., 2003). The PPI was then defined as the element-by-element product of the neuronal time series and a vector coding for the effect of beliefs. A general linear model was set up for each subject, which included the following regressors of interest: the time series of the seed region (the physiological variable), the convolved stimulus stick function (the psychological variable: "belief-congruent > belief-incongruent") and the reconvolved interaction term (the psychophysiological variable). Additional regressors of no interest were the convolved stick functions of button presses indicating uncertainty, six head movement parameters, and one constant for each run. Contrasts of interest for the PPI were created.

In a second-level analysis, the individual contrast images for the PPI were correlated with the conviction score from the delusional questionnaire. Results were considered statistically significant at $p < 0.05$, FWE-corrected for multiple comparisons across the whole brain or within our region-of-interest in visual cortex, the motion-sensitive area hMT/V5. To define this region, the data from the localizer runs were used. The first four scans were discarded. In SPM8, the data were preprocessed applying slice-time correction, motion adjustment, spatial normalization to the MNI template, and spatial smoothing with a Gaussian Kernel (FWHM, $8 \times 8 \times 8$ mm). A first-level analysis then included two regressors of interest (motion and stationary) and the six movement parameters as regressors of no interest. For the identification of the motion-sensitive area hMT/V5, the individual contrast images "motion > stationary" were subjected to a second-level one-sample t test. Using MRIcron, a bilateral cluster on the occipitotemporal junction was identified and manually delineated (threshold $p < 0.001$ uncorrected, 409 voxels).

Regions-of-interest in visual cortex. To define regions-of-interest in visual cortex, the data from the localizer runs were used. The first four scans were discarded. In SPM8 the data were preprocessed applying slice-time correction, motion adjustment, spatial normalization to the MNI template, and spatial smoothing with a Gaussian kernel (FWHM, $8 \times 8 \times 8$ mm). A first-level analysis then included two regressors of interest (motion and stationary) and the six movement parameters as regressors of no interest. For the identification of the motion-sensitive area hMT/V5 the individual contrast images "motion > stationary" were subjected to a second-level one-sample t test. Using MRIcron, a bilateral cluster on the occipitotemporal junction was identified and manually delineated (threshold $p < 0.001$ uncorrected, 409 voxels). For the iden-

tification of more posterior occipital visual cortex regions (i.e., early visual areas) that responded to the stimulus, the individual contrast images "motion > baseline" were subjected to a second-level one-sample *t* test. Using MRIcron, a bilateral cluster in occipital cortex was identified and manually delineated (threshold $p < 0.001$ uncorrected, 841 voxels).

Results

Behavioral Experiment 1: delusional ideation is associated with perceptual instability

In Behavioral Experiment 1, we asked whether the tendency toward delusional ideation is associated with weakened lower-level sensory predictions in perceptual inference. The ambiguous sphere stimulus was presented intermittently and the survival probability of a percept from one stimulus presentation to the next was used to quantify in each individual participant the strength of lower-level sensory predictions in perceptual inference.

As reported previously (Orbach et al., 1963; Leopold et al., 2002; Maier et al., 2003; Pearson and Brascamp, 2008), the survival probability of percepts across temporary stimulus removals was high (91.33 ± 0.87% SEM; see Fig. 1*B* for an exemplary perceptual time course) but variable between participants. Across individuals, this tendency of percept stabilization was inversely correlated with the overall delusions questionnaire score ($r = -0.26$, $p = 0.008$, product-moment correlation, *p* value based on 10,000 permutations). The dimension "delusional conviction" alone accounted best for this correlation (Fig. 1*C*; $r = -0.26$, $p = 0.004$, stepwise linear forward regression, *p* value based on 10,000 permutations), showing that individuals with a stronger tendency toward delusional convictions exhibited less perceptual stabilization. To test whether interindividual differences in the feedforward sensory signal due to eye movements were related to the tendency of perceptual stabilization, we analyzed the eye-tracking data collected during the experiment. Valid eye-tracking data were obtained in 66 participants. Fixation quality was not significantly correlated with the tendency of perceptual stabilization expressed as the percept survival probability ($r = -0.02$, $p = 0.40$, product-moment correlation, *p* value based on 10,000 permutations), indicating that there was no influence of eye movements on perceptual stability. Thus, these findings confirm our first hypothesis that the tendency toward delusional ideation is associated with perceptual instability and suggest, in line with previous accounts (Hemsley, 1993, 2005; Heinz, 2002; Kapur, 2003; Corlett et al., 2009; Fletcher and Frith, 2009), that the formation of delusional beliefs might be related to an attenuated effect of sensory predictions in perceptual inference.

Behavioral Experiment 2: delusional ideation is associated with a stronger belief-induced bias on reported perception

Behavioral Experiment 2 tested whether the tendency toward delusional ideation is associated with a stronger effect of higher-level cognitive beliefs in perceptual inference. Beliefs were induced using a placebo-like manipulation (Sterzer et al., 2008). Participants first viewed the ambiguous sphere stimulus continuously and reported perceptual changes between leftward and rightward rotation. In a learning phase, beliefs regarding the dominant rotation of the sphere were induced with transparent glasses. Participants were made to believe that the glasses contained polarizing filters that would bias their perception toward one direction, while in fact the sphere stimulus was surreptitiously disambiguated with 3D cues to yield strong dominance of one rotation direction. In a subsequent test phase, we then probed the effect of the induced perceptual belief on perception of the ambiguous sphere.

As intended by the disambiguation of the stimulus, belief-congruent percepts in the learning phase were perceived for longer average phase durations than belief-incongruent percepts (difference of normalized percept durations $\mu = 1.65$, $\sigma = 0.75$, $t = 22.5$, $p < 0.001$, paired *t* test, *p* value based on 10,000 permutations). Importantly, in the following test phase, perception of the ambiguous sphere was also biased toward the belief-congruent rotation direction (difference of normalized percept durations $\mu = 0.14$, $\sigma = 0.35$, $t = 4.2$, $p < 0.001$, paired *t* test, *p* value based on 10,000 permutations; Fig. 2*B*), which in the absence of disambiguation 3D cues can be attributed only to the participants' beliefs. Strikingly, the strength of the belief-induced perceptual bias in the test phase was related to the tendency toward delusional ideation, as revealed by a positive correlation between the ratio of belief-congruent and belief-incongruent perceptual phase durations and the overall score from the delusion questionnaire ($r = 0.26$, $p = 0.005$, product-moment correlation, *p* value based on 10,000 permutations). Again, the dimension "delusional conviction" alone predicted the perceptual effect best ($r = 0.26$, $p = 0.004$, stepwise linear forward regression, *p* value based on 10,000 permutations; Fig. 2*C*), showing that in individuals with a stronger tendency toward delusional convictions, perception was more strongly biased by beliefs. Importantly, the strength of the belief-induced perceptual bias was negatively correlated with survival probability from Behavioral Experiment 1 (compare Fig. 1*A*, $r = -0.19$, $p = 0.03$, product-moment correlation, *p* value based on 10,000 permutations). This finding supports our idea that in those individuals who are more prone to delusional ideation and in whom sensory representations are less stable, perceptual inference rests more upon cognitive beliefs.

To investigate whether eye movements might account for the reported belief-induced perceptual bias, we analyzed the eye-tracking data collected during the experiment. Valid eye-tracking data for the entire experiment could be obtained from only 17 participants due to technical problems imposed by the transparent glasses that were used in the learning and test phase. The effect of beliefs was highly significant in this subsample (difference of normalized percept durations $\mu = 0.18$, $\sigma = 0.56$, $t = 2.0$, $p = 0.03$, *p* value based on 10,000 permutations), and fixation quality did not differ significantly between baseline and test phase (difference of area in square degrees $\mu = -1.53$, $\sigma = 8.48$, $t = -0.75$, $p = 0.25$, paired *t* test, *p* value based on 10,000 permutations), rendering it unlikely that the belief-induced perceptual bias was due to eye movements. When we restricted the analysis to the baseline phase only, for which eye movements could be recorded in 55 participants, we still did not find an effect of eye movements on perception: fixation quality in the baseline phase was not significantly correlated with mean percept duration ($r = 0.01$, $p = 0.40$, product-moment correlation, *p* value based on 10,000 permutations). Together and in line with previous reports (van Dam and van Ee, 2005), we did not find any evidence for an influence of eye movements on perception of the ambiguous stimulus.

Our behavioral data thus far show that in individuals with higher delusion scores and lower perceptual stability reported perception of the ambiguous stimulus was more strongly influenced by experimentally induced beliefs. An important question arises about whether these participants only reported belief-congruent percepts more readily, or whether their beliefs actually biased sensory processing of the stimulus via predictive signals from higher-order cortical areas, as predicted by our model of delusions. To address this issue, we examined the neural correlates of the perceptual effect of beliefs in a new cohort of

participants ($n = 20$) that underwent the same placebo-like manipulation as in Behavioral Experiment 2 (compare Fig. 2*A*), but now during fMRI scanning.

fMRI experiment: signal patterns in visual cortex reflect the enhanced effect of beliefs on perception in delusion-proneness

Behaviorally, perception of the ambiguous sphere in the test phase was again strongly biased by beliefs (difference of normalized percept durations $\mu = 0.42$, $\sigma = 0.81$, normalized durations; $t = 2.34$, $p < 0.001$, paired t test, p value based on 10,000 permutations; Fig. 4*A*). Despite the considerably smaller sample size, there was still a trend toward a significant association of the belief-induced perceptual bias with the tendency toward delusional convictions ($r = 0.34$, $p = 0.08$, product-moment correlation, p value based on 10,000 permutations; Fig. 4*B*), replicating our findings from Behavioral Experiment 2 in an independent sample.

In our analysis of the fMRI data, we first tested whether beliefs influenced the sensory processing of the ambiguous stimulus. We used MVPA, a highly sensitive analytical tool that has been successfully applied to decode conscious perception from fMRI signals even when the contents of awareness change during constant physical stimulation (Haynes and Rees, 2005; Kamitani and Tong, 2006; Brouwer and van Ee, 2007). In a whole-brain searchlight approach (Kriegeskorte et al., 2006; Fig. 3*A*), we verified that perceived rotation direction can be decoded from fMRI signal patterns. A SVM classifier (Cortes and Vapnik, 1995) was trained on a subset of fMRI data from the baseline phase to discriminate between perceived rightward and leftward rotation at each time point and then tested on an independent subset of fMRI data from the baseline phase and the test phase. Note that during these phases the ambiguous stimulus was constantly presented and that decoding thus related to purely perceptual changes. Voxelwise accuracy maps resulting from the searchlight analysis were computed and tested against chance level. When testing the classifier on fMRI data from the baseline phase (i.e., in the absence of beliefs), perceived rotation direction could be decoded with above-chance accuracy from visual cortex, including bilateral early occipital visual areas and extending anteriorly to the motion-sensitive area hMT/V5 [Fig. 5*A*, left occipital (−18, −91, 10), $t_{(19)} = 7.15$, $p < 0.001$; right occipital (18, −94, 10), $t_{(19)} = 4.37$, $p = 0.015$; left hMT/V5 (−48, −76, 10), $t_{(19)} = 6.02$, $p < 0.001$; all p values FWE-corrected within functionally defined visual cortex regions-of-interest; a cluster in right hMT/V5 did not survive correction for multiple testing (33, −79, 1), $t_{(19)} = 2.75$, $p = 0.006$, uncorrected]. Notably, when the same classifier was applied to the test phase, in which perception was strongly biased by beliefs, the regional pattern of decoding accuracy closely resembled the one resulting from the first analysis of the baseline phase only [Fig. 5*B*, left occipital (−30, −94, 7), $t_{(19)} = 5.51$, $p = 0.002$; right occipital (12, −97, 1), $t_{(19)} = 4.32$, $p = 0.020$; left hMT/V5 (−48, −70, 1), $t_{(19)} = 3.95$, $p = 0.024$; right hMT/V5 (51, −70, 4), $t_{(19)} = 3.59$, $p = 0.047$, all p values FWE-corrected within functionally defined visual cortex regions of interest). Decoding accuracy was also above chance in right precentral gyrus [(36, −7, 52), $t_{(18)} = 6.81$, $p = 0.02$, FWE-corrected within whole brain], most likely due to motor responses.

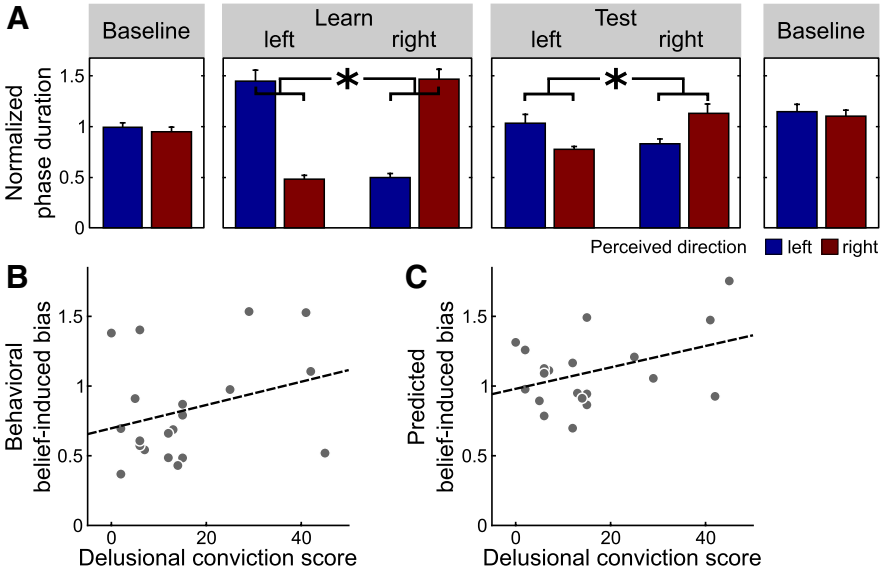

Figure 4. Delusional ideation and belief-induced bias in the fMRI experiment ($n = 20$). ***A***, Effect of beliefs on reported perception (*$p < 0.001$, paired t test, p value based on 10,000 permutations). Bars show the mean phase duration of each percept normalized with respect to the mean phase duration in the baseline runs. Error bars denote SE. ***B***, Correlation between tendency toward delusional convictions and belief-induced bias as reported by the participants ($r = 0.34$, $p = 0.08$, product-moment correlation, p value based on 10,000 permutations). Behavioral belief-induced bias was calculated as the ratio of reported belief-congruent and belief-incongruent mean phase durations in the test phase normalized with respect to the learning phase. Tendency toward delusional convictions was measured with a validated questionnaire (Peters et al., 1999). Each dot represents one participant. The dashed line illustrates the fitted regression line. ***C***, Correlation between tendency toward delusional convictions and belief-induced bias as predicted by MVPA in visual cortex ($r = 0.40$, $p = 0.04$, product-moment correlation, p value based on 10,000 permutations). An optimized classifier was trained on the baseline runs to predict perception from fMRI activation patterns in visual cortex (compare main text) and tested on the test and learning runs. Predicted belief-induced bias was then calculated as the ratio of predicted belief-congruent and belief-incongruent mean phase durations in the test phase normalized with respect to the learning phase. Each dot represents one participant. The dashed line illustrates the fitted regression line.

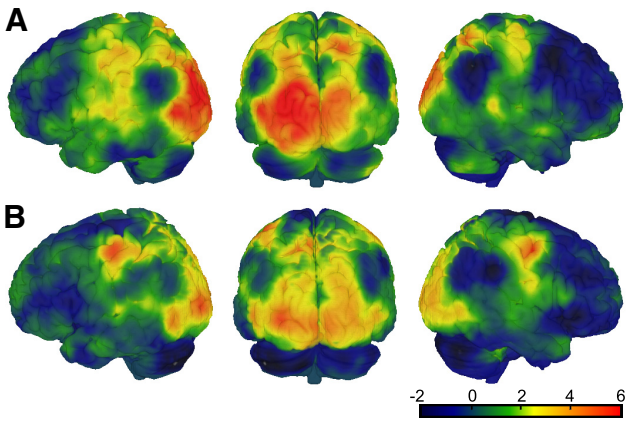

Figure 5. Whole-brain decoding of perceived rotation direction using an MVPA searchlight approach. Colors indicate t values from a voxelwise paired t test comparing decoding accuracy to chance level. ***A***, Decoding accuracy of a classifier trained and tested on the data from the baseline runs in a runwise cross-validation scheme. ***B***, Decoding accuracy of a classifier trained on the data from all baseline runs and tested on the data from all test runs. Please note that ***A*** and ***B*** are not directly comparable as the data used for training and testing differed in amount and proximity in time. The t value thresholds corresponding to $p < 0.05$, FWE-corrected across the whole brain are 5.62 (***A***) and 5.80 (***B***).

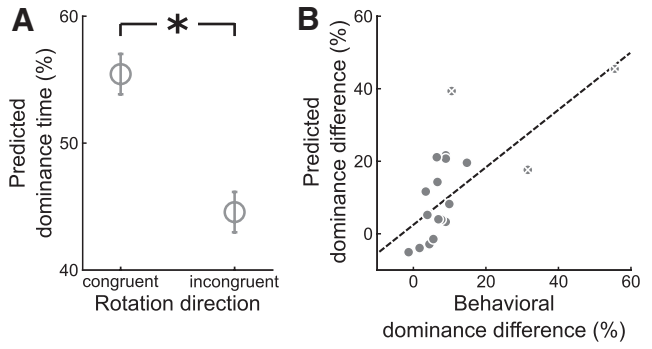

Figure 6. Effect of experimentally induced beliefs on neural correlates of perceived motion direction in visual cortex as revealed by MVPA. ***A***, Mean cumulative dominance times predicted by the classifier for the belief-congruent versus the belief-incongruent rotation direction during the test runs. Error bars denote SEM. *$p < 0.005$, two-tailed paired t test. ***B***, Correlation of the behavioral and the dominance difference predicted by the classifier ($r = 0.70$, $p < 0.001$, product-moment correlation, p value based on 10,000 permutations). Dominance difference was calculated as the difference between dominance times for the belief-congruent and the belief-incongruent rotation direction. Each gray dot represents one participant. White crosses mark three potential outliers that deviated >2 SDs from the group mean in the behavioral or predicted dominance difference. The dashed line illustrates the fitted regression line for the entire group. Exclusion of the three potential outliers yielded similar results.

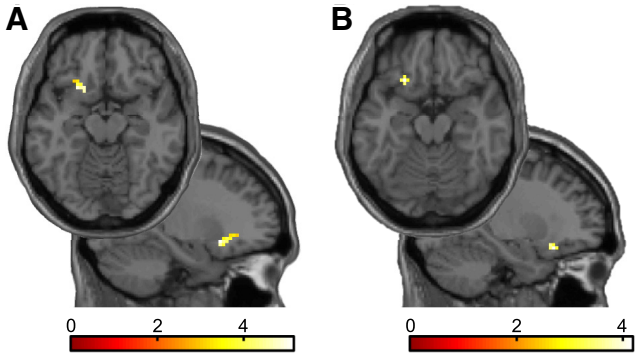

Figure 7. Belief-related neural activity in learn and test phase ($t = 5.14$, $p = 0.009$ and $t = 4.16$, $p = 0.045$, both FWE-corrected). ***A***, ***B***, Axial and sagittal slices show voxels in which the switch-related activity (belief-congruent vs belief-incongruent perceptual switches) in the learning phase (***A***) and in the test phase (***B***) predicted the behavioral belief-induced perceptual bias. For display purposes, t maps are thresholded at $p < 0.005$, $k > 5$ voxels.

Hence, we could establish that subjectively perceived rotation direction is represented in fMRI activation patterns in visual cortex regardless of the belief-induced perceptual bias. Most critical to our research question, however, we sought to quantify the effect of beliefs on perception at the neural level. We reasoned that, if experimentally induced beliefs altered the sensory processing of the ambiguous stimulus, this should be reflected in fMRI signal patterns in visual cortex and enable us to decode the belief-induced bias in perception using MVPA. We restricted the analysis to those brain regions that are primarily involved in the sensory processing of the stimulus, that is, visual cortex comprising occipital cortex and the functionally defined motion-sensitive area hMT/V5 (Fig. 3*B*). Again, an SVM classifier was trained on a subset of fMRI data from the baseline phase to discriminate between perceived rightward and leftward rotation at each time point. When testing the classifier separately on an independent subset of fMRI data from the baseline and test phase, there was no significant difference in decoding accuracy ($\mu = -0.2\%$, $\sigma = 5.0\%$, $t_{(19)} = -0.14$, $p = 0.89$, paired t test, p value based on 10,000 permutations). Consistent with this, the optimized classifier predicted significantly more expected than unexpected percepts during the test phase (Fig. 6*A*; 55.4 vs 44.6%, $\mu = 10.9\%$, $\sigma = 14.2\%$, $t_{(19)} = 3.42$, $p = 0.003$, paired t test, p value based on 10,000 permutations). Strikingly, a correlation analysis across participants showed that the dominance difference that we could decode from visual cortex correlated significantly with the dominance difference expressed in behavioral reports ($r = 0.70$, $p < 0.001$, product-moment correlation, p value based on 10,000 permutations; Fig. 6*B*). Excluding three potential outliers that differed >2 SDs from the group mean yielded similar results ($r = 0.66$, $p = 0.002$, product-moment correlation, p value based on 10,000 permutations). These findings strongly suggest that participants did not merely report what they believed but that information processing in visual cortex was altered in accord with their beliefs. Importantly, and corroborating at the neural level our behavioral finding of an association between the tendency toward delusional ideation and the belief-induced perceptual bias, the tendency toward delusional ideation correlated significantly with the belief-induced bias as decoded from visual cortex ($r = 0.40$, $p = 0.04$, product-moment correlation, p value based on 10,000 permutations; Fig. 4*C*; this analysis was performed with the dimension "delusional conviction," which had shown the strongest correlation effects in our behavioral experiments).

fMRI experiment: delusional ideation is associated with enhanced connectivity between orbitofrontal and visual cortex

We next asked which brain circuits may be involved in generating the observed influence of beliefs on visual information processing. Placebo studies suggest that the OFC acts to generate and maintain experimentally induced beliefs that in turn modulate activity in sensory brain areas (Petrovic et al., 2002, 2005; Wager et al., 2004). Moreover, feedback signals from the same region have been implicated in the expectation-mediated facilitation of perceptual decision making and object recognition (Bar et al., 2006; Kveraga et al., 2007; Summerfield and Koechlin, 2008). Thus we conjectured that perceptual beliefs might be associated with enhanced OFC activity. For the learning and the test phase separately, we tested whether fMRI activity evoked by the belief-congruent percept, compared with the belief-incongruent percept, was related to the strength of the belief-induced perceptual bias. In line with previous work on the placebo effect in pain and emotion perception (Petrovic et al., 2002, 2005; Wager et al., 2004), we indeed found the belief-induced perceptual bias to be associated with fMRI responses in the left OFC. Interestingly, this effect was observed both during the learning phase [Fig. 7*A*; $(-21, 17, -11)$, $t_{(18)} = 5.14$, $p = 0.009$, FWE-corrected within a priori-defined OFC region of interest] and the test phase [Fig. 7*B*; $(-24], 23, -17)$, $t_{(18)} = 4.16$, $p = 0.045$, FWE-corrected within OFC region of interest]. Across the whole brain, no other clusters showed a significant effect in either the learning or the test phase. Thus, these results indicate that OFC is critically involved in the generation and maintenance of perceptual beliefs.

One central claim of our model was that in individuals with delusions, perceptual inference relies more on predictions from higher-level brain circuits that encode beliefs. Consequently, we reasoned that the tendency toward delusional ideation should be associated with enhanced functional coupling between OFC and visual cortex. In a voxelwise connectivity analysis using a psychophysiological interaction approach (Friston et al., 1997), we tested for a correlation between the tendency toward delusional ideation and changes in functional connectivity between OFC and other brain regions during perception of belief-congruent

versus belief-incongruent percepts. This analysis was performed with the dimension "delusional conviction," which had shown the strongest correlation effects in our behavioral experiments. For the learning phase (compare Fig. 2*A*), in which the rotation direction was unambiguous and the need for perceptual inference thus minimized, this analysis did not show any significant effects, not even at a lenient statistical threshold of $p < 0.005$ (uncorrected). Strikingly, in the following test phase, where the stimulus was completely ambiguous, the tendency toward delusional convictions was associated with stronger belief-dependent connectivity between OFC and bilateral motion-sensitive area hMT/V5 in visual cortex [Fig. 8; (−33, −79, 1]), $t_{(18)} = 5.62$, $r = 0.49$, $p = 0.003$ and (39, −64, 10), $t_{(18)} = 4.24$, $r = 0.44$, $p = 0.035$, both FWE-corrected within functionally defined visual cortex region of interest]. In accord with our second hypothesis, these findings show that, faced with ambiguous sensory input, participants with high delusion scores exhibit enhanced functional coupling of higher-level areas in OFC with lower-level sensory areas in visual cortex, providing the neural basis for a stronger influence of beliefs on perceptual inference in delusion-prone individuals.

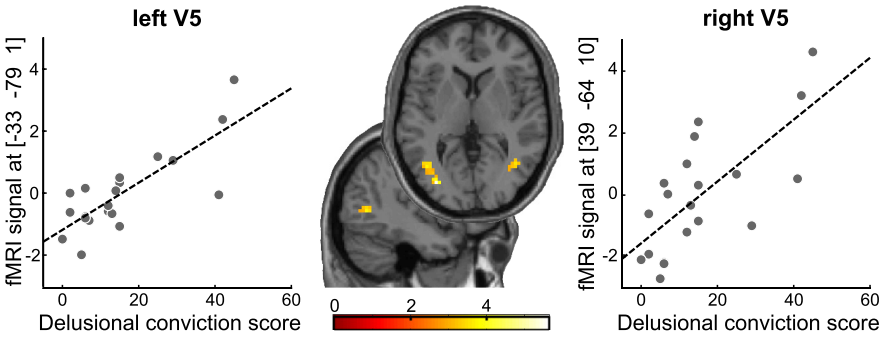

Figure 8. Correlation of delusions and belief-related connectivity between OFC and bilateral hMT/V5 in test phase (left $r = 0.49$, $p = 0.003$; right $r = 0.44$, $p = 0.035$, both FWE-corrected). Axial and sagittal slices show voxels in which the psychophysiological interaction [OFC × (belief-congruent − belief-incongruent perceptual switches)] correlated with the conviction score from a validated delusions questionnaire (Peters et al., 1999). For display purposes, *t* maps are thresholded at $p < 0.005$, $k > 5$ voxels.

Discussion

Our current study provides evidence along several lines for a comprehensive model of delusional beliefs based on a Bayesian framework of perceptual inference and belief formation. First, in our sample of healthy individuals, delusional ideation was associated with perceptual instability, indicating weakened sensory predictions. This finding is in line with a growing body of evidence indicating a link between delusions and weakened effects of endogenous predictions on action and perception (Blakemore et al., 2000; Schneider et al., 2002; Dakin et al., 2005; Lindner et al., 2005; Shergill et al., 2005; Uhlhaas et al., 2006; Dima et al., 2009; Synofzik et al., 2010; Teufel et al., 2010; Voss et al., 2010; Sanders et al., 2013). With recourse to previous accounts (Heinz, 2002; Kapur, 2003; Hemsley, 2005; Corlett et al., 2009, 2010; Fletcher and Frith, 2009), we therefore suggest that attenuated predictive signaling within sensory processing stages provides the starting point for maladaptive learning processes that yield the emergence of delusional beliefs. Second, to the best of our knowledge, this is the first experimental demonstration of an increased influence of beliefs on perception in delusion-prone individuals, which in contrast to previous findings (Blakemore et al., 2000; Schneider et al., 2002; Dakin et al., 2005; Hemsley, 2005; Lindner et al., 2005; Shergill et al., 2005; Uhlhaas et al., 2006; Dima et al., 2009; Synofzik et al., 2010; Teufel et al., 2010; Voss et al., 2010; Sanders et al., 2013) offers an explanation for the persistence of delusional beliefs. At the neural level, this was paralleled by an association of the tendency toward delusional ideation and enhanced functional connectivity between frontal areas and visual areas encoding perception. Importantly, the effect of beliefs on perception was reflected in fMRI signal patterns, showing for the first time that higher-level cognitive beliefs not only bias reported perception but alter information processing in visual cortex. Thus, we propose that excessive predictive signaling from higher-level cortical areas encoding beliefs to lower-level sensory areas shapes perception in accord with delusional beliefs, which may constitute the critical mechanism for the tenacious persistence of delusional beliefs. Third, weaker perceptual stability was directly associated with a stronger influence of beliefs on perception. Crucially, this establishes a link between the explanations for the formation and the fixity of delusional beliefs: when faced with unstable sensory representations that arise from weak predictions within lower-level sensory processing stages, perceptual inference seems to rely more strongly on predictions from higher-level nonsensory areas encoding beliefs.

Essentially, our findings are consistent with the idea that delusional beliefs result from an aberration in the signaling of prediction error (Heinz, 2002; Kapur, 2003; Corlett et al., 2009, 2010; Fletcher and Frith, 2009). More specifically, a lack of precision in the prediction error signal (Fletcher and Frith, 2009; Corlett et al., 2010) at lower levels of the cortical hierarchy might impede the appropriate implementation of reliable sensory predictions. This would be paralleled by unstable sensory representations, which is in line with our observation of an association between the tendency toward delusional ideation and decreased perceptual stability. From a conceptual perspective (Friston, 2005), the consequences of a decreased precision of the prediction error signal are twofold. On the one hand, as suggested previously (Heinz, 2002; Kapur, 2003; Corlett et al., 2009, 2010; Fletcher and Frith, 2009), the noise in the prediction error signal engages maladaptive learning by inappropriately updating higher-level predictions. The experience of expected and irrelevant sensory events as surprising and relevant initiates a search for explanation and leads to the formation of delusional beliefs. On the other hand, sensory prediction instability that results from imprecise prediction error signaling at lower levels biases perceptual inference toward higher-level predictions. Faced with sensory signals that are constantly indicated as chaotic and unpredictable, perception relies more strongly on internal sources of information, such as cognitive beliefs, which might be the critical mechanism for the persistence of delusional beliefs. This is in accord with our finding of an association between the tendency toward delusional ideation and a stronger belief-induced perceptual bias. It has been proposed that in the framework of perceptual inference, the role of attention can be conceptualized as optimizing the precision of sensory signals (Feldman and Friston, 2010). Similar to attention, other high-level predictions about the state of the world, such as beliefs or expectations, may also serve to optimize the precision of sensory signals. Along these lines, our findings suggest that exaggerated high-level predictions in delusion-prone individuals may represent an attempt to

optimize imprecise prediction error signaling at sensory processing levels, thereby also adding to the ongoing debate regarding the relationship between expectations and attention in perceptual inference (Summerfield and Egner, 2009). In summary, the imprecise prediction error signal at lower levels of the cortical hierarchy triggers two distinct adaptive processes accounting for the fundamental characteristics of delusional beliefs: the attempt to flexibly respond to the exaggerated noise in the sensory representations through the modification of higher-level predictions yields the formation of delusional beliefs, while the attempt to impose stability on unstable sensory representations through reliance on higher-level predictions results in the maintenance of delusional beliefs. Although it is conceivable that the two processes occur simultaneously, it is likely that additional situational factors, such as general arousal, have an influence on which of the two processes is preferred.

Delusions have been proposed to constitute an extreme expression of a continuously distributed phenotype (Meehl, 1962). In line with this continuity view of psychosis, varying degrees of delusional ideation are observed in the general, nonclinical population (Freeman, 2006). Subclinical delusional ideation is predictive of later psychosis (Chapman et al., 1994), and is associated with the same epidemiological and environmental risk factors as psychotic disorder (van Os et al., 2009), indicating that delusional ideation and clinical delusions can be explained in terms of similar underlying mechanisms. Therefore, we suggest that our findings linking altered perceptual inference with delusional ideation in healthy individuals indeed inform the understanding of delusions. Importantly, our approach enables us to safely conclude that the observed alterations of perceptual inference are not related to psychotropic medication or other consequences of psychotic disorder, but rather to the tendency toward unfounded beliefs itself. Our findings thus illustrate how interindividual differences in perception relate to interindividual differences in belief formation and maintenance and provide a promising starting point for future work on physiological and pathological beliefs.

While we did not study delusions in a clinical sense, our current findings offer a comprehensive and plausible explanation for the emergence and persistence of unfounded beliefs. They thus contribute substantially to the understanding of the mechanisms that jointly govern beliefs and perception both in health and in psychosis.

References

Bar M, Kassam KS, Ghuman AS, Boshyan J, Schmid AM, Schmidt AM, Dale AM, Hämäläinen MS, Marinkovic K, Schacter DL, Rosen BR, Halgren E (2006) Top-down facilitation of visual recognition. Proc Natl Acad Sci U S A 103:449–454. CrossRef Medline

Blakemore SJ, Smith J, Steel R, Johnstone CE, Frith CD (2000) The perception of self-produced sensory stimuli in patients with auditory hallucinations and passivity experiences: evidence for a breakdown in self-monitoring. Psychol Med 30:1131–1139. CrossRef Medline

Brascamp JW, Kanai R, Walsh V, van Ee R (2010) Human middle temporal cortex, perceptual bias, and perceptual memory for ambiguous three-dimensional motion. J Neurosci 30:760–766. CrossRef Medline

Brouwer GJ, van Ee R (2007) Visual cortex allows prediction of perceptual states during ambiguous structure-from-motion. J Neurosci 27:1015–1023. CrossRef Medline

Chapman LJ, Chapman JP, Kwapil TR, Eckblad M, Zinser MC (1994) Putatively psychosis-prone subjects 10 years later. J Abnorm Psychol 103:171–183. CrossRef Medline

Corlett PR, Frith CD, Fletcher PC (2009) From drugs to deprivation: a Bayesian framework for understanding models of psychosis. Psychopharmacology (Berl) 206:515–530. CrossRef Medline

Corlett PR, Taylor JR, Wang XJ, Fletcher PC, Krystal JH (2010) Toward a neurobiology of delusions. Prog Neurobiol 92:345–369. CrossRef Medline

Cortes C, Vapnik V (1995) Support-vector networks. Machine Learning 20:273–297. CrossRef

Dakin S, Carlin P, Hemsley D (2005) Weak suppression of visual context in chronic schizophrenia. Curr Biol 15:R822–R824. CrossRef Medline

Dima D, Roiser JP, Dietrich DE, Bonnemann C, Lanfermann H, Emrich HM, Dillo W (2009) Understanding why patients with schizophrenia do not perceive the hollow-mask illusion using dynamic causal modelling. Neuroimage 46:1180–1186. CrossRef Medline

Feldman H, Friston KJ (2010) Attention, uncertainty, and free-energy. Front Hum Neurosci 4:215. CrossRef Medline

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

Freeman D (2006) Delusions in the nonclinical population. Curr Psychiatry Rep 8:191–204. CrossRef Medline

Friston K (2005) A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci 360:815–836. CrossRef Medline

Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997) Psychophysiological and modulatory interactions in neuroimaging. Neuroimage 6:218–229. CrossRef Medline

Gitelman DR, Penny WD, Ashburner J, Friston KJ (2003) Modeling regional and psychophysiologic interactions in fMRI: the importance of hemodynamic deconvolution. Neuroimage 19:200–207. CrossRef Medline

Haynes JD, Rees G (2005) Predicting the stream of consciousness from activity in human visual cortex. Curr Biol 15:1301–1307. CrossRef Medline

Heinz A (2002) Dopaminergic dysfunction in alcoholism and schizophrenia—psychopathological and behavioral correlates. Eur Psychiatry 17:9–16. CrossRef Medline

Hemsley DR (1993) A simple (or simplistic?) cognitive model for schizophrenia. Behav Res Ther 31:633–645. CrossRef Medline

Hemsley DR (2005) The schizophrenic experience: taken out of context? Schizophr Bull 31:43–53. CrossRef Medline

Kamitani Y, Tong F (2006) Decoding seen and attended motion directions from activity in the human visual cortex. Curr Biol 16:1096–1102. CrossRef Medline

Kapur S (2003) Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. Am J Psychiatry 160:13–23. CrossRef Medline

Kersten D, Mamassian P, Yuille A (2004) Object perception as Bayesian inference. Annu Rev Psychol 55:271–304. CrossRef Medline

Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. Proc Natl Acad Sci U S A 103:3863–3868. CrossRef Medline

Kveraga K, Ghuman AS, Bar M (2007) Top-down predictions in the cognitive brain. Brain Cogn 65:145–168. CrossRef Medline

Leopold DA, Wilke M, Maier A, Logothetis NK (2002) Stable perception of visually ambiguous patterns. Nat Neurosci 5:605–609. CrossRef Medline

Lindner A, Thier P, Kircher TT, Haarmeier T, Leube DT (2005) Disorders of agency in schizophrenia correlate with an inability to compensate for the sensory consequences of actions. Curr Biol 15:1119–1124. CrossRef Medline

Maier A, Wilke M, Logothetis NK, Leopold DA (2003) Perception of temporally interleaved ambiguous patterns. Curr Biol 13:1076–1085. CrossRef Medline

Meehl PE (1962) Schizotaxia, schizotypy, schizophrenia. Am Psychol 17:827–838. CrossRef

Mumford D (1992) On the computational architecture of the neocortex. II. the role of cortico-cortical loops. Biol Cybern 66:241–251. CrossRef Medline

Orbach J, Ehrlich D, Heath HA (1963) Reversibility of the Necker cube. I. An examination of the concept of "satiation of orientation". Percept Mot Skills 17:439–458. CrossRef Medline

Pearson J, Brascamp J (2008) Sensory memory for ambiguous vision. Trends Cogn Sci 12:334–341. CrossRef Medline

Peters ER, Joseph SA, Garety PA (1999) Measurement of delusional ideation in the normal population: introducing the PDI (Peters et al. delusions inventory). Schizophr Bull 25:553–576. CrossRef Medline

Petrovic P, Kalso E, Petersson KM, Ingvar M (2002) Placebo and opioid

analgesia—imaging a shared neuronal network. Science 295:1737–1740. CrossRef Medline

Petrovic P, Dietrich T, Fransson P, Andersson J, Carlsson K, Ingvar M (2005) Placebo in emotional processing-induced expectations of anxiety relief activate a generalized modulatory network. Neuron 46:957–969. CrossRef Medline

Sanders LL, de Millas W, Heinz A, Kathmann N, Sterzer P (2013) Apparent motion perception in patients with paranoid schizophrenia. Eur Arch Psychiatry Clin Neurosci 263:233–239. CrossRef Medline

Schneider U, Borsutzky M, Seifert J, Leweke FM, Huber TJ, Rollnik JD, Emrich HM (2002) Reduced binocular depth inversion in schizophrenic patients. Schizophr Res 53:101–108. CrossRef Medline

Schurger A (2009) A very inexpensive MRI-compatible method for dichoptic visual stimulation. J Neurosci Methods 177:199–202. CrossRef Medline

Shergill SS, Samson G, Bays PM, Frith CD, Wolpert DM (2005) Evidence for sensory prediction deficits in schizophrenia. Am J Psychiatry 162: 2384–2386. CrossRef Medline

Sterzer P, Frith C, Petrovic P (2008) Believing is seeing: expectations alter visual awareness. Curr Biol 18:R697–R698. CrossRef Medline

Summerfield C, Egner T (2009) Expectation (and attention) in visual cognition. Trends Cogn Sci 13:403–409. CrossRef Medline

Summerfield C, Koechlin E (2008) A neural representation of prior information during perceptual inference. Neuron 59:336–347. CrossRef Medline

Synofzik M, Thier P, Leube DT, Schlotterbeck P, Lindner A (2010) Misattributions of agency in schizophrenia are based on imprecise predictions about the sensory consequences of one's actions. Brain 133:262–271. CrossRef Medline

Teufel C, Kingdon A, Ingram JN, Wolpert DM, Fletcher PC (2010) Deficits in sensory prediction are related to delusional ideation in healthy individuals. Neuropsychologia 48:4169–4172. CrossRef Medline

Uhlhaas PJ, Phillips WA, Mitchell G, Silverstein SM (2006) Perceptual grouping in disorganized schizophrenia. Psychiatry Res 145:105–117. CrossRef Medline

van Dam LC, van Ee R (2005) The role of (micro) saccades and blinks in perceptual bistability from slant rivalry. Vision Res 45:2417–2435. CrossRef Medline

van Os J, Linscott RJ, Myin-Germeys I, Delespaul P, Krabbendam L (2009) A systematic review and meta-analysis of the psychosis continuum: evidence for a psychosis proneness-persistence-impairment model of psychotic disorder. Psychol Med 39:179–195. CrossRef Medline

Velez DR, White BC, Motsinger AA, Bush WS, Ritchie MD, Williams SM, Moore JH (2007) A balanced accuracy function for epistasis modeling in imbalanced datasets using multifactor dimensionality reduction. Genet Epidemiol 31:306–315. CrossRef Medline

Veropoulos K, Campbell C, Cristianini N (1999) Controlling the sensitivity of support vector machines. pp 55–60. Proceedings of the International Joint Conference on Artificial Intelligence.

von Helmholtz H (1867) Handbuch der physiologischen Optik. Leipzig: Leopold Voss.

Voss M, Moore J, Hauser M, Gallinat J, Heinz A, Haggard P (2010) Altered awareness of action in schizophrenia: a specific deficit in predicting action consequences. Brain 133:3104–3112. CrossRef Medline

Wager TD, Rilling JK, Smith EE, Sokolik A, Casey KL, Davidson RJ, Kosslyn SM, Rose RM, Cohen JD (2004) Placebo-induced changes in FMRI in the anticipation and experience of pain. Science 303:1162–1167. CrossRef Medline