

Ketamine Affects Prediction Errors about Statistical Regularities: A Computational Single-Trial Analysis of the Mismatch Negativity

 Lilian A. Weber,^{1*}  Andreea O. Diaconescu,^{1,2,3*}  Christoph Mathys,^{1,4,5}  André Schmidt,² Michael Kometer,⁶  Franz Vollenweider,⁶ and  Klaas E. Stephan^{1,7,8}

¹Translational Neuromodeling Unit (TNU), Institute for Biomedical Engineering, University of Zurich and ETH Zurich, 8032 Zurich, Switzerland, ²Department of Psychiatry, University of Basel, 4001 Basel, Switzerland, ³Krembil Centre for Neuroinformatics, Centre for Addiction and Mental Health, University of Toronto, Ontario M6J 1H4, Canada, ⁴Interacting Minds Centre, Aarhus University, DK-8000 Aarhus, Denmark, ⁵Scuola Internazionale Superiore di Studi Avanzati (SISSA), 34136 Trieste, Italy, ⁶Neuropharmacology and Brain Imaging, University Hospital of Psychiatry, University of Zurich, 8032 Zurich, Switzerland, ⁷Wellcome Centre for Human Neuroimaging, University College London, London WC1N 3AR, United Kingdom, and ⁸Max Planck Institute for Metabolism Research, 50931 Cologne, Germany

The auditory mismatch negativity (MMN) is significantly reduced in schizophrenia. Notably, a similar MMN reduction can be achieved with NMDA receptor (NMDAR) antagonists. Both phenomena have been interpreted as reflecting an impairment of predictive coding or, more generally, the “Bayesian brain” notion that the brain continuously updates a hierarchical model to infer the causes of its sensory inputs. Specifically, neurobiological interpretations of predictive coding view perceptual inference as an NMDAR-dependent process of minimizing hierarchical precision-weighted prediction errors (PEs), and disturbances of this putative process play a key role in hierarchical Bayesian theories of schizophrenia. Here, we provide empirical evidence for this theory, demonstrating the existence of multiple, hierarchically related PEs in a “roving MMN” paradigm. We applied a hierarchical Bayesian model to single-trial EEG data from healthy human volunteers of either sex who received the NMDAR antagonist S-ketamine in a placebo-controlled, double-blind, within-subject fashion. Using an unrestricted analysis of the entire time-sensor space, our trial-by-trial analysis indicated that low-level PEs (about stimulus transitions) are expressed early (102–207 ms poststimulus), while high-level PEs (about transition probability) are reflected by later components (152–199 and 215–277 ms) of single-trial responses. Furthermore, we find that ketamine significantly diminished the expression of high-level PE responses, implying that NMDAR antagonism disrupts the inference on abstract statistical regularities. Our findings suggest that NMDAR dysfunction impairs hierarchical Bayesian inference about the world’s statistical structure. Beyond the relevance of this finding for schizophrenia, our results illustrate the potential of computational single-trial analyses for assessing potential pathophysiological mechanisms.

Key words: HGF; hierarchical Gaussian filter; mismatch negativity; NMDA receptor; predictive coding; schizophrenia

Significance Statement

The NMDA receptor antagonist ketamine induces psychosis-like experiences in healthy individuals, consistent with the notion that the pathophysiology of schizophrenia involves NMDAR dysfunction. On the cognitive level, the stark misrepresentations of reality during psychosis suggest a dysfunction at high levels of belief hierarchies, where general and stable features of the environment are represented. Here, we investigate physiological indices of altered perception under ketamine—the reduction of the auditory “mismatch negativity”—based on their algorithmic interpretation as hierarchical belief updates. We find that ketamine indeed impacts cortical signaling of higher-level belief updates about environmental volatility. This finding bridges the physiological and computational concepts of NMDAR dysfunction and offers a novel mechanistic perspective on a central element of pathophysiological theories of schizophrenia.

Introduction

The auditory mismatch negativity (MMN), an electrophysiological response to rule violations in auditory input streams, has long been interpreted as evidence that the brain learns the statistical structure of its environment and predicts future sensory inputs (Paavilainen et al., 1999; Näätänen et al., 2005; Winkler, 2007). In psychiatry, reductions in MMN amplitude are among the most robust electrophysiological abnormalities in individuals with schizophrenia (Umbricht and Krljes, 2005; Todd et al., 2013; Erickson et al., 2016; Avissar et al., 2018). Physiologically, following primate work showing that MMN depends on intact NMDA receptor (NMDAR) signaling (Javitt et al., 1996), human electroencephalographic (EEG) studies demonstrated significant reductions of MMN responses under the NMDAR antagonist ketamine (Umbricht et al., 2000; Heekeren et al., 2008; Schmidt et al., 2012).

The robust impairment of MMN in schizophrenia, and the fact that similar MMN reduction can be achieved with NMDAR antagonists like ketamine, are in line with the long-standing notion that the pathophysiology of schizophrenia involves NMDAR dysfunction (Olney and Farber, 1995; Goff and Coyle, 2001). More specifically, aberrant NMDAR-mediated signaling plays a central role for explaining perceptual abnormalities and positive symptoms in schizophrenia from a “predictive coding” view (Stephan et al., 2006, 2009; Corlett et al., 2011, 2016; Adams et al., 2013; Friston et al., 2016; Sterzer et al., 2018). According to predictive coding and related “Bayesian brain” theories, the brain continuously updates a hierarchical model of its environment to infer the causes of sensory inputs and predict future events (Dayan et al., 1995; Rao and Ballard, 1999; Friston, 2010; Doya et al., 2011).

The auditory MMN is believed to reflect model updating during perceptual inference within the auditory hierarchy (Winkler, 2007; Garrido et al., 2009; Lieder et al., 2013b). In predictive coding, each hierarchical level provides predictions about the state of the level below and, in turn, receives a prediction error (PE) signal reflecting the discrepancy between predicted and actual state of the level below. It is thought that predictions are communicated by descending (backward) connections, while PEs are signaled by ascending (forward) connections (Friston, 2005); furthermore, glutamatergic signaling was found to mainly occur via AMPA receptors at ascending connections and via NMDARs at descending connections (Self et al., 2012). Critically, ascending PE signals are weighted by the relative precision of bottom-up input compared with predictions (priors) from higher levels. The MMN, which is a difference waveform, is commonly interpreted as the difference in precision-weighted PEs between surprising and predictable events.

This predictive coding perspective, which views the MMN as a reflection of perceptual inference in the auditory cortical hierarchy, makes the following two major predictions:

First, hierarchically related precision-weighted PEs should underlie the MMN (Lieder et al., 2013b), particularly when the stimulus stream exhibits considerable volatility (Mathys et al., 2011; Iglesias et al., 2013; Diaconescu et al., 2014). Trial-by-trial changes in MMN responses should reflect the temporal dynamics of Bayesian belief updating and the PEs involved (Mars et al., 2008; Ostwald et al., 2012; Lieder et al., 2013a; Kolossa et al., 2015; Jepma et al., 2016; Stefanics et al., 2018).

Second, the expression of precision-weighted PEs should be sensitive to NMDAR manipulations. As described above, blocking NMDARs should lead to a reduction of top-down (predictive) signaling, resulting in less constrained inference about the causes of sensory inputs. This, in turn, should render predictable and less predictable stimuli more similar in how surprising they are and thus alter the bottom-up signaling of PEs (Corlett et al., 2007, 2011, 2016).

Here, we examine these predictions and present the first computational single-trial EEG analysis of auditory MMN data under pharmacological manipulations of NMDAR function (S-ketamine vs placebo) in healthy volunteers. While analyses of the same dataset have been published previously (Schmidt et al., 2012, 2013), these used standard event-related potential (ERP) and connectivity methods operating on trial averages. Neither study used a computational trial-by-trial model and could thus not examine the trialwise expression of different PEs or their changes under NMDAR blockade.

Materials and Methods

Here, we reanalyze a previously published study (Schmidt et al., 2012) that administered S-ketamine to healthy volunteers. Details on participants, drug administration, and data acquisition have been provided previously (Schmidt et al., 2012, 2013); the interested reader is referred to these articles for more information. Here, we only briefly summarize these aspects and focus on the model-based EEG analysis.

Participants

Nineteen healthy individuals (12 males; mean age, 26 ± 5.09 years) gave informed written consent and participated in the study, which was approved by the Ethics Committee of the University Hospital of Psychiatry (Zurich, Switzerland). The use of psychoactive drugs was approved by the Swiss Federal Health Office, Department of Pharmacology and Narcotics (Bern, Switzerland). Inclusion criteria included physical and mental health, and the absence of a history of drug dependence and of present drug use. For detailed examinations before inclusion and additional questionnaire assessments, see Schmidt et al. (2012).

Experimental procedure and paradigm

The two sessions (placebo and S-ketamine) that all participants underwent in a counterbalanced fashion were separated by at least 2 weeks. Both participants and the experimenter interacting with them were blind to the drug order. S-ketamine was administered using an indwelling catheter that was placed in the antecubital vein of the nondominant arm. An initial bolus injection of 10 mg over 5 min was followed, after a 1 min break, by a continuous infusion with 0.006 mg/kg/min over 80 min. The initial dose was reduced by 10% every 10 min to keep S-ketamine plasma levels fairly constant (Feng et al., 1995; Vollenweider et al., 1997). The procedure in the placebo session was equivalent for administering an infusion of physiological sodium chloride solution and 5% glucose. Each participant was kept under constant supervision until all drug effects had worn off, and was then released into the custody of a partner or immediate relative.

EEG activity was recorded during an auditory “roving” oddball paradigm, originally developed by Cowan et al. (1993) and subsequently modified by Baldeweg et al. (2004). E-prime software (Schneider et al.,

Received Dec. 31, 2019; revised May 12, 2020; accepted June 5, 2020.

Author contributions: F.V. and K.E.S. designed research; A.S. and M.K. performed research; L.A.W. and C.M. contributed unpublished reagents/analytic tools; L.A.W. and A.O.D. analyzed data; L.A.W., A.O.D., and K.E.S. wrote the paper.

*L.A.W. and A.O.D. contributed equally to this work.

This study has been published as a preprint on bioRxiv (<https://doi.org/10.1101/528372>). The authors declare no competing financial interests.

This study was supported by the University of Zurich (K.E.S.); the René and Susanne Braginsky Foundation (K.E.S.); Swiss National Science Foundation Ambizione Grant P200P3_167952 and the Krembil Foundation (A.O.D.); the Swiss Neuromatrix (M.K., F.V.); and the Heffer Research Institute (F.V.).

Correspondence should be addressed to Lilian A. Weber at weber@biomed.ee.ethz.ch.

<https://doi.org/10.1523/JNEUROSCI.3069-19.2020>

Copyright © 2020 the authors

2002) was used to generate acoustic stimuli that were presented binaurally through headphones.

The stimuli consisted of seamlessly connected trains of pure sinusoidal tones (70 ms duration, 500 ms interstimulus interval) with a roving frequency structure using seven different frequencies from 500 to 800 Hz in steps of 50 Hz. Within each stimulus train, all tones were of one frequency and were followed by a train of tones of a different frequency. The number of times the same tone was presented within one stimulus train varied pseudorandomly between 1 and 11 such that 5% of all stimulus trains consisted of 1–2 identical tones, 7.5% of all stimulus trains consisted of 3–4 identical stimuli, and 87.5% of all stimulus trains consisted of 5–11 identical stimuli. For each participant and each session, a different sequence of tones was generated online.

Following the suggestion that MMN assessment is optimal when the participant's attention is directed away from the auditory domain (Näätänen, 2000), participants performed a distracting visual task and were instructed to ignore the sounds. Whenever a fixation cross changed its luminance, which occurred pseudorandomly every 2–5 s (not coinciding with auditory changes), participants had to press a button. One experimental session lasted ~15 min.

Data processing

The EEG was recorded at a sampling rate of 512 Hz using a BioSemi system with 64 scalp electrodes. Preprocessing and data analysis was performed using SPM12 (<http://www.fil.ion.ucl.ac.uk/spm/>). Continuous EEG recordings were referenced to the average, high-pass filtered using a Butterworth filter with a cutoff frequency of 0.5 Hz, downsampled to 256 Hz, and low-pass filtered using a Butterworth filter with a cutoff frequency of 30 Hz. The data were epoched into 500 ms segments around tone onsets, using a prestimulus baseline of 100 ms.

We rejected all trials overlapping with eye blink events, as detected by a thresholding routine on the vertical EOG channel, which was created from subtracting the activity of two additional electrodes that were attached infraorbitally and supraorbitally to the left eye. Finally, an artifact rejection procedure was applied using a thresholding approach on all EEG channels to detect problematic trials or channels. Trials in which the signal recorded at any of the channels exceeded 80 μ V relative to the prestimulus baseline were removed from subsequent analysis, and channels in which >20% of trials had to be rejected were marked as bad and subsequently interpolated for sensor-level statistics.

Bad channels occurred for three participants in the placebo session (with five, one, and one bad channel(s), respectively) and for two participants in the ketamine session (with one and two bad channels, respectively). Additionally, we had to mark two channels (F1 and C2) in five datasets as bad (two in placebo, three in ketamine) due to incorrect cabling. To exclude the possibility that our main results were driven by the interpolation of missing channel data, we performed all statistical analyses on the group level once without participant 14, who lost five channels due to bad signal quality, and once without the four participants affected by the cabling errors. The SPM results reported in this article remained significant unless stated otherwise.

The average total number of artifact-free trials was 1211 (SD, 201) in the placebo and 1464.6 (SD, 211.2) in the ketamine condition. The number of artifact-free trials was thus significantly lower in the placebo sessions. However, the resulting nonsphericity was accommodated by our second-level statistical tests (paired *t* tests to assess group differences; see the “Experimental design and statistical analysis” section). Note that we did not define categorical events like standard and deviant trials, but instead included all tones in our trial-by-trial analysis.

Computational model

In what follows, we briefly outline our perceptual model before describing the analysis steps used to apply this model to single-trial EEG data. In terms of notation, we denote scalars by lower case italics (e.g., *x*), vectors by lower case bold letters (e.g., ***x***), and matrices by upper case bold letters (e.g., ***X***). Trial numbers are indexed by the superscript (*k*) [e.g., $x^{(k)}$].

Perceptual model: the hierarchical Gaussian filter. To describe a participant's perceptual inference and learning during this roving MMN

paradigm, we use a multivariate version of the hierarchical Gaussian filter (HGF), a generic Bayesian model introduced by Mathys et al. (2011) that has been applied in various contexts, such as associative learning (Iglesias et al., 2013; Weinhhammer et al., 2018), social learning (Diaconescu et al., 2014, 2017), spatial attention (Vossel et al., 2014b), visual mismatch negativity (Stefanics et al., 2018), visual discrimination (Aukstulewicz et al., 2017), and sensorimotor learning (Palmer et al., 2019).

In the present task, participants were exposed to a tone sequence with seven different tones. Our modeling approach assumes that in this context, an agent infers the following two hidden states in the world: (1) the current (probabilistic) “laws” underlying the observed tone statistics—in our case, a matrix ***X***₂ of pairwise transition probabilities between all tones; and (2) the current level of environmental volatility (i.e., how quickly the inferred laws seem to change). This is represented in our model by the volatility x_3 , which is the degree to which the transition probabilities in ***X***₂ change from trial to trial. The rationale for tracking this quantity is that agents should learn faster (i.e., update their beliefs about the statistical laws in the environment according to prediction errors) if they experience the current environment to be changing rather than stable. Figure 1 shows a visualization of the corresponding generative model.

On each trial, the agent updates her beliefs about these two environmental states, given the new sensory input (i.e., tone). We denote these updated (posterior) beliefs in the following by their mean μ and their precision (or certainty) π (the inverse of variance, or uncertainty, σ). In the HGF, the general form of the update of the posterior mean at hierarchical level *i* on trial *k* is as follows:

$$\Delta\mu_i^{(k)} \propto \frac{\hat{\pi}_{i-1}^{(k)}}{\pi_i^{(k)}} \delta_{i-1}^{(k)}. \quad (1)$$

Here, $\delta_{i-1}^{(k)}$ denotes the PE about the state on the level below, which is weighted by a ratio of precisions: $\hat{\pi}_{i-1}^{(k)}$ is the precision of the prediction about the level below (*i* − 1), while $\pi_i^{(k)}$ is the precision of the current belief at the current hierarchical level *i*. The intuition behind this is that belief updates of an agent should be more strongly driven by PEs when the precision of predictions about the input is high relative to the precision of beliefs in the current estimate (e.g., when the environment is currently perceived as being volatile). For the specific update equations for the two levels of our model, as well as a detailed derivation of these equations, the interested reader is referred to the paper by Mathys et al. (2011).

Usually, in the HGF, subject-specific perceptual parameters describe the individual learning style and computational trajectories (e.g., beliefs and PEs) of an agent. Specifically, the three parameters κ , ω , and ϑ determine the strength of the coupling between the second and third level (κ), the tonic volatility estimate on the second level (ω ; i.e., the part of the learning rate that is independent of online volatility estimation), and the (tonic) volatility estimate on the third level (ϑ ; i.e., the speed of change in volatility). Here, we fixed the coupling parameter κ to 1 because the scale of x_3 is arbitrary in our setting (Mathys et al., 2014). This effectively eliminates this parameter from the model. An additional four free parameters determine the starting values of an individual's beliefs at the beginning of the task ($\mu_2^{(0)}$, $\sigma_2^{(0)}$, $\mu_3^{(0)}$, and $\sigma_3^{(0)}$). Since the current paradigm does not involve behavioral responses to the tones, and thus the model could not be fitted to behavior, we used the parameters (volatility estimates on both hierarchical levels and starting values of the beliefs) of a surprise-minimizing Bayesian observer for all participants, similar to a previous application of the HGF to single-trial (visual) MMN responses (Stefanics et al., 2018). This ideal observer was defined as the parameter values that result in minimal overall surprise about the stimulus sequence encountered. To estimate these parameters, we used the MATLAB function `tapas_fitModel` from the HGF toolbox (version 4.0), distributed as part of TAPAS (release version 1.6) with the `tapas_bayes_optimal_whatworld` function as a pseudoresponse model. For this estimation, we used the default priors for this model in the HGF toolbox, after verifying that all parameters that had a visible impact on belief trajectories were being estimated. The prior settings were as follows: $\omega =$

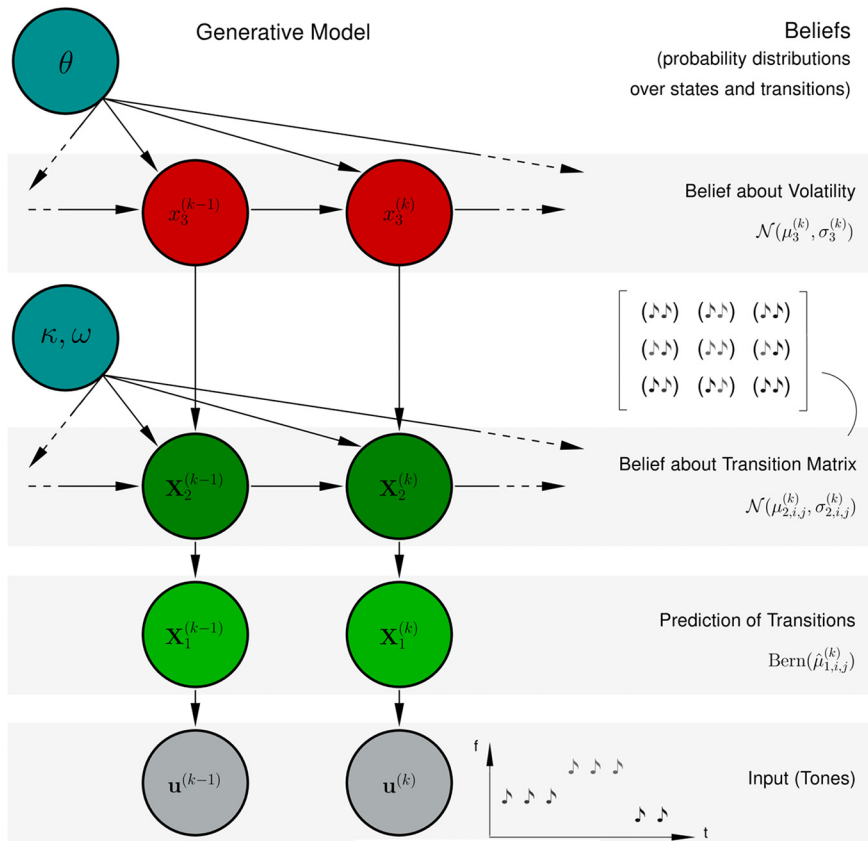


Figure 1. The perceptual model: a multivariate version of the binary three-level HGF. The agent infers on two continuous quantities: the transition tendencies from one tone (frequency) to another, stored in the transition matrix x_2 , and the (common) volatility of these tendencies, x_3 . To use this model, the agent only has to follow simple one-step update rules for its beliefs (parameterized by their mean μ and variance σ) about these quantities.

-6 (var = 25), $\vartheta = 0.05$ (var = 0.088), and $\sigma_3^{(0)} = 0.1$ (var = 1 in log space), with $\kappa = \mu_3^{(0)} = \sigma_{2ij}^{(0)} = 1$ and $\mu_{2ij}^{(0)} = \frac{1}{7}$ fixed (var = 0). Note that the scale of x_3 is arbitrary in our context (see above), and therefore the choice of κ and $\mu_3^{(0)}$ is arbitrary, too. We chose the initial value of $\mu_{2ij}^{(0)}$, the belief about tone transitions, to be neutral (i.e., all transitions were equally likely a priori), with sufficient uncertainty ($\sigma_{2ij}^{(0)}$) to enable learning. Because of the subject- and session-specific stimulus sequences (they were generated, under identical probabilities across participants, on the spot during each session), this procedure resulted in slightly different parameter values for the ideal observer for each recording session. There were no significant differences in these parameter values between placebo and ketamine conditions (low-level tonic volatility estimate ω : placebo mean = -10.04 , SD = 0.2; ketamine mean = -10.06 , SD = 0.28, $p = 0.83$; high-level tonic volatility estimate ϑ : placebo mean = 0.045, SD = 0.003; ketamine mean = 0.046, SD = 0.003, $p = 0.33$; and starting value of the high-level uncertainty $\sigma_3^{(0)}$: placebo mean = 0.10, SD = 0.00; ketamine mean = 0.10, SD = 0.00; $p = 0.67$, p values refer to two-sided paired t tests). The resulting belief and PE trajectories for a representative session are depicted in Figure 2.

Computational quantities: the precision-weighted prediction errors. The MMN has been interpreted as a precision-weighted PE (or model update signal) during auditory perceptual inference and statistical learning (Näätänen et al., 2001; Winkler, 2007; Garrido et al., 2008, 2009; Wacongne et al., 2012; Lieder et al., 2013a,b). In our model, two hierarchical levels are updated in response to new auditory inputs, as follows: the current estimate of the transition probabilities (μ_2), and the current estimate of environmental volatility (μ_3). The corresponding precision-weighted PEs driving these updates are hierarchically related and are computed sequentially, as follows: the agent first needs to update μ_2 (using the low-level PE about μ_1) before evaluating its high-level PE with respect to μ_2 , which is then used to update μ_3 . Following previous

notation (Iglesias et al., 2013), we denote these precision-weighted PEs by ε_2 and ε_3 .

We address the following questions in this article: (1) whether these precision-weighted PEs are reflected by trial-by-trial variations in the amplitude of evoked responses; (2) whether their hierarchical relation in the model is mirrored by a corresponding temporal relation in their electrophysiological correlates; and (3) whether NMDAR antagonism by S-ketamine alters the electrophysiological expression of these PEs.

Experimental design and statistical analysis

We examined manifestations of our two computational quantities (ε_2 and ε_3) in the event-related EEG responses for each trial in a time window from 100 to 400 ms poststimulus. We focused on this time window to model learning-induced modulations of both the MMN and the P300 waveforms.

The data from each trial in each session were converted into scalp images for all 64 channels and 91 time points using a voxel size of $4.25 \times 5.38 \times 3.33$ ms. The images were constructed using linear interpolation for removed bad channels and spatial smoothing with a Gaussian kernel (FWHM: 16×16 mm) in accordance with the assumptions of random field theory (Worsley et al., 1996; Kiebel and Friston, 2004) to accommodate for between-subject spatial variability in channel space.

Our vectors of precision-weighted PEs served as regressors in a general linear model (GLM) of trialwise EEG signals for each participant and each session separately, after removing the entries of trials that had been rejected during EEG preprocessing. We did not orthogonalize the regressors. Figure 3 summarizes the analysis steps for the model-based GLM.

Random-effects group analysis across all 19 participants was performed using a standard summary statistics approach (Penny and Holmes, 2007). We used one-sample t tests as second-level models, separately for each drug condition, and used F tests to simultaneously examine positive and negative relations of EEG amplitudes with the trajectories of our computational quantities. To examine differences between the two drug conditions, we tested for reduced responses under ketamine using a paired t test.

For all analyses, we report any results that survived familywise error (FWE) correction, based on Gaussian random field theory (Kilner and Friston, 2010), across the entire volume (time \times sensor space) at the cluster level ($p < 0.05$) with a cluster defining threshold (CDT) of $p < 0.001$ (Flandin and Friston, 2019). Notably, all reported results also survive whole-volume correction at the peak level ($p < 0.05$).

Region-of-interest analysis

To relate our trialwise analysis approach to the conventional ERP analysis presented in a previous report of our dataset (Schmidt et al., 2012), we performed a region of interest (ROI) analysis in a subset of frontal sensors, using our model-based PE estimates. Specifically, we defined “standards” and “deviants” as those trials with the 10% lowest and 10% highest precision-weighted PEs, respectively, according to our model. We averaged the respective responses across trials and across three frontal sensors [Fz, F3, and F4 (same sensors as in Schmidt et al., 2012)], separately for standard and deviant trials, and computed a difference wave (deviant – standard, or high PE – low PE) for each participant and each recording session (placebo, ketamine). Next, we extracted the peak of

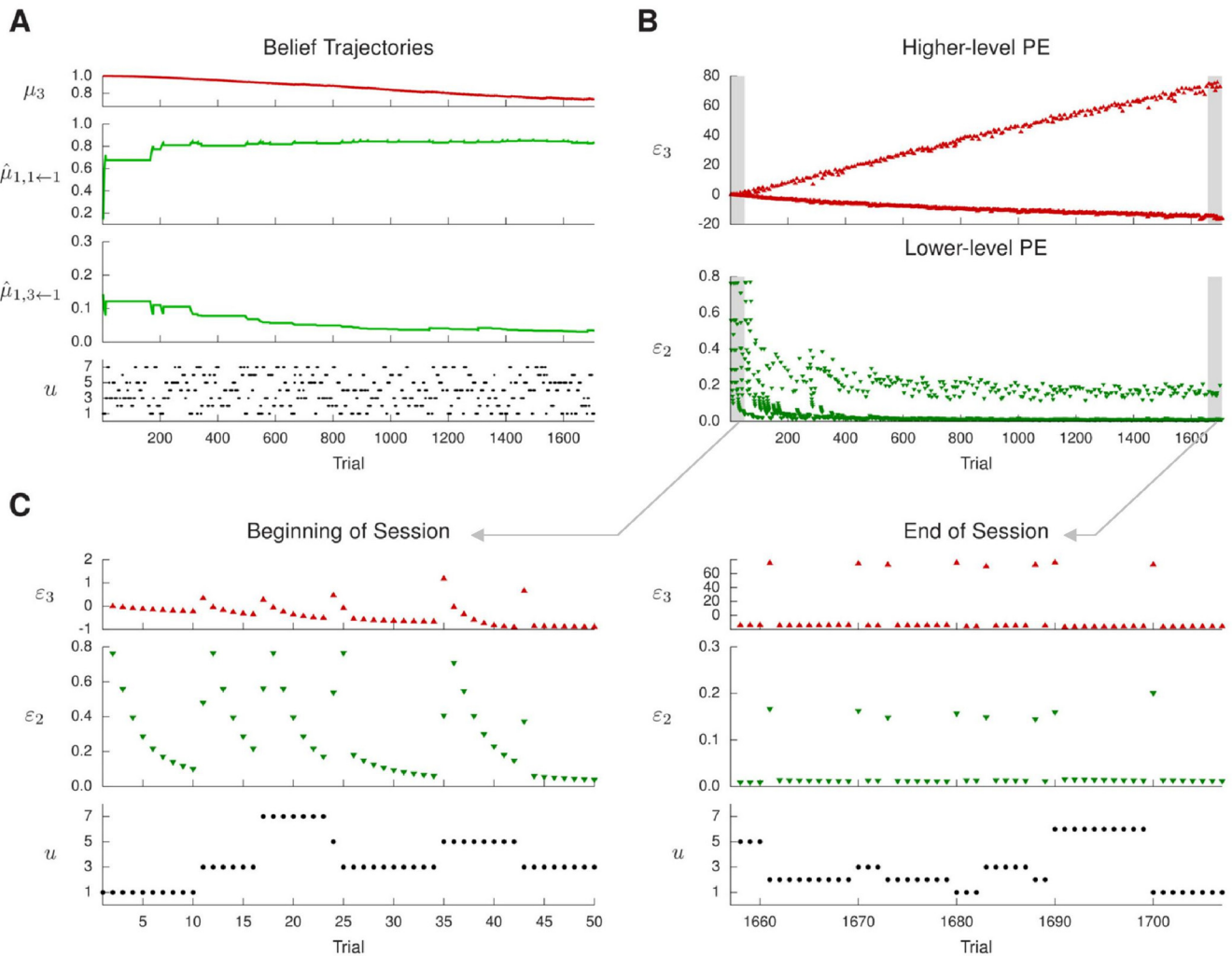


Figure 2. Simulation results: example trajectories from a representative participant in the placebo session. **A**, Time course over the experimental session (1707 trials) for the simulated belief of the agent about volatility (μ_3 , in red), her trial-by-trial prediction of the repetition probability of tone 1 ($\hat{\mu}_{1,1,1}$, in green), her trial-by-trial prediction of the transition probability from tone 1 to tone 3 ($\hat{\mu}_{1,3,1}$, in green), and the tone sequence she was exposed to (u , black dots). After an initial learning period, the agent has learned that repetitions are likely, and transitions are unlikely. **B**, Trial-by-trial values of the two precision-weighted PEs used as regressors in the GLM of the EEG signal. The initial learning period is characterized by high low-level PEs about stimulus occurrences (ε_2) for all tone events; later the deviant trials (tone transitions) separate clearly from the standard trials (repetitions). As the volatility estimate of the agent decreases, the higher-level PE (ε_3) in response to deviants separates more and more from the one in standard trials. **C**, The two PE trajectories in more detail for the 50 first tones of the session (left) and the 50 last tones of the session (right). At first, all tones are similarly surprising (ε_2 in the beginning of the session), but in the end, because the estimates of repetition and transition probabilities of the agent are stable and accurate, repetitions elicit almost no PEs anymore.

this difference wave between 100 and 200 ms (MMN), where deviant ERPs are more negative than standards, and tested whether this peak mismatch effect was significantly different between placebo and ketamine sessions using a paired *t* test. This procedure exactly mimics the analysis presented in the study by Schmidt et al. (2012), except for the definition of standard and deviant trials, which was based on our model here. Additionally, we ran the same test for a later time window (200–300 ms), where amplitudes of ERPs to deviant tones are more positive than standards (P3a). We separately performed these ROI analyses for our two precision-weighted PEs (ε_2 and ε_3), resulting in the following four tests: (1) placebo vs. ketamine for mismatch based on lower-level PEs in the early time window, and (2) in the late time window, and (3) for mismatch based on higher-level PEs in the early time window and (4) in the late time window.

Results

For each computational quantity of interest, our model-based EEG analysis proceeded in the following two steps: first, we performed whole-volume (spatiotemporal) analyses to search for

representations of our quantities in single-trial EEG responses; and second, we examined whether these electrophysiological representations of trialwise PEs differed significantly between ketamine and placebo.

Low-level precision-weighted prediction errors

By fitting computational trajectories to participants' single-trial EEG data, we found that under placebo there was a significant trial-by-trial relation between $\varepsilon_2^{(k)}$ (the precision-weighted transition PE) and EEG activity between 102 and 207 ms poststimulus, peaking at 121 ms at central channels (peak, $F_{(1,18)} = 70.0$; whole-volume cluster-level FWE corrected, $p = 2.8 \times 10^{-8}$; with a CDT of $p < 0.001$; Fig. 4). This time window includes the typical time when the negativity of the roving MMN is observed (Cowan et al., 1993; Baldeweg et al., 2004; Garrido et al., 2008). Critically, more negative EEG amplitudes in this cluster corresponded to higher PE values (i.e., more surprising events; Fig. 4). This suggests that the MMN typically observed in roving MMN

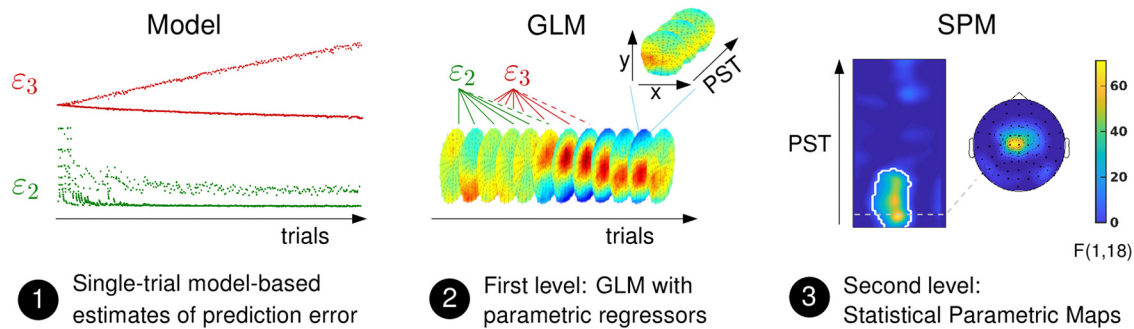


Figure 3. Sketch of the analysis pipeline for the model-based analysis. First, we simulated the beliefs of an agent using our hierarchical Bayesian model, which provided us with an estimate of precision-weighted PE on two hierarchical levels (ε_2 and ε_3) for each trial in each session. At first glance, it may seem that the ε_3 regressor simply amounts to a drift-like signal. This, however, is not the case; the design of our experiment, with prolonged trains of identical stimuli that exchange each other, leads to separate monotonic changes in log-volatility estimates for standard and deviant trials, with jump-like transitions between them (Fig. 2). Second, we used these estimates as parametric regressors in a general linear model (GLM) of the single-trial EEG signal in each session of each participant separately (peristimulus time window of 100–400 ms after tone onset) and computed the first-level statistics. Third, the β values for each quantity and each participant in each session entered the second-level analysis. We performed random-effects group analysis across all 19 participants separately for each drug condition in one-sample t tests and used F tests to examine correlations of EEG amplitudes with our computational quantities of interest, resulting in thresholded F statistical parametric maps (SPM) across within-trial time and sensor space. PST, Peristimulus time.

paradigms reflects the difference in low-level precision-weighted PEs about stimulus transitions between the subsets of trials labeled as standards and deviants by the experimenter.

Under ketamine infusion, we found a similar activation pattern, with significant clusters of activity at frontocentral electrodes between 107 and 188 ms (peak, $F_{(1,18)} = 63.4$; at 137 ms, $p = 3.1 \times 10^{-8}$), and at left temporal channels between 105 and 188 ms, peaking at 141 ms poststimulus (peak, $F_{(1,18)} = 76.4$; $p = 6.3 \times 10^{-6}$; Fig. 5).

High-level precision-weighted prediction errors

In the placebo condition, we found a significant trial-by-trial relation between $\varepsilon_3^{(k)}$ (the precision-weighted PE that serves to update volatility estimates) and EEG activity, both in an early time window (between 152 and 199 ms, peaking at 184 ms at right temporal channels; peak: $F_{(1,18)} = 58.8$, $p = 0.004$; and between 145 and 188 ms, peaking at 180 ms at frontal channels; peak: $F_{(1,18)} = 31.6$, $p = 0.009$) and in a later time window (between 215 and 277 ms, peaking at 266 ms poststimulus; peak: $F_{(1,18)} = 35.1$, $p = 0.002$; Fig. 4), where high-level PEs correlated with an increased central positivity corresponding to the P3a component of the auditory-evoked potential (Polich, 2007).

Under ketamine, we found a similar relationship of EEG amplitudes with the higher-level PE in the early time window (148–211 ms, peaking at 160 ms at right temporal channels; peak: $F_{(1,18)} = 35.0$, $p = 0.04$; and 156–215 ms, peaking at 207 ms at frontocentral channels; peak: $F_{(1,18)} = 25.2$, $p = 0.008$), but the later cluster occurred only much later (297–398 ms, peaking at 375 ms at right temporal channels; peak: $F_{(1,18)} = 32.3$, $p = 0.021$; and 324–398 ms, peaking at 398 ms at frontocentral channels; peak: $F_{(1,18)} = 35.5$, $p = 0.001$; Fig. 5). While the timing of this late effect is reminiscent of the P3b component, high-level PEs in this cluster correlated with a frontocentral negativity (Fig. 5) instead of a parietal positivity, as would be characteristic for P3b (Polich, 2007; Watson et al., 2009).

Effects of ketamine on PE representations

We tested for drug differences in activity elicited by precision-weighted PEs using paired t tests at the second level. We found no significant differences in activation by $\varepsilon_2^{(k)}$ in the ketamine condition compared with the placebo condition. In contrast, the activation by $\varepsilon_3^{(k)}$, the higher-level PE informing volatility estimates, was significantly reduced under ketamine compared with

placebo in a time window between 207 and 250 ms poststimulus, peaking at 223 ms across frontocentral channels (peak: $t_{(18)} = 5.95$, $p = 0.005$; Fig. 6). That is, the trial-by-trial relation between EEG signal and the higher-level PE was significantly more pronounced under placebo than under ketamine in this time window.

To relate this result to the previously reported effect of ketamine on MMN amplitudes between 100 and 200 ms in a frontal ROI in our dataset (Schmidt et al., 2012), we repeated the same ROI analysis performed there, but with standard and deviant events defined as the 10% least surprising trials (lowest precision-weighted PE) and 10% most surprising trials (highest precision-weighted PE), respectively, according to our trialwise estimates of low- and high-level PEs. We found that both in an early (100–200 ms) and in a late (200–300 ms) time window, ketamine significantly reduced the mismatch effect in a frontal ROI composed of sensors Fz, F3, and F4, but only when using the trial definition based on the higher-level PE (two-sided paired t tests: $t_{\text{early}(18)} = -3.57$, $p = 0.002$; $t_{\text{late}(18)} = 2.56$, $p = 0.02$; Fig. 6).

Discussion

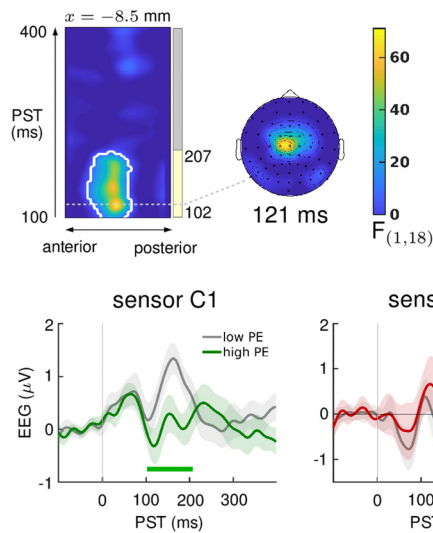
Current theories of schizophrenia conceptualize psychotic symptoms as disturbed hierarchical Bayesian inference, characterized by an imbalance in the relative weight (precision) assigned to prior beliefs (or predictions) and new sensory information that elicits PEs (Adams et al., 2013; Corlett et al., 2016; Sterzer et al., 2018). Neurobiologically, this disturbance is thought to result from alterations of NMDAR-dependent synaptic plasticity and to be reflected by abnormalities in perceptual paradigms, such as the auditory MMN (Stephan et al., 2006, 2009; Friston et al., 2016). Based on a computational single-trial analysis of the MMN under ketamine, the current results are supportive of the following two major predictions: (1) multiple and hierarchically related precision-weighted PEs should underlie the MMN; and (2) the expression of precision-weighted PEs should be sensitive to NMDAR manipulations.

Multiple, hierarchically related prediction errors underlie the MMN

The auditory MMN has been interpreted as reflecting model updates in an auditory processing hierarchy (Garrido et al., 2008, 2009; Lieder et al., 2013a). In our Bayesian learning model, levels

Placebo

Effect of ε_2 (Lower-level PE)



Effect of ε_3 (Higher-level PE)

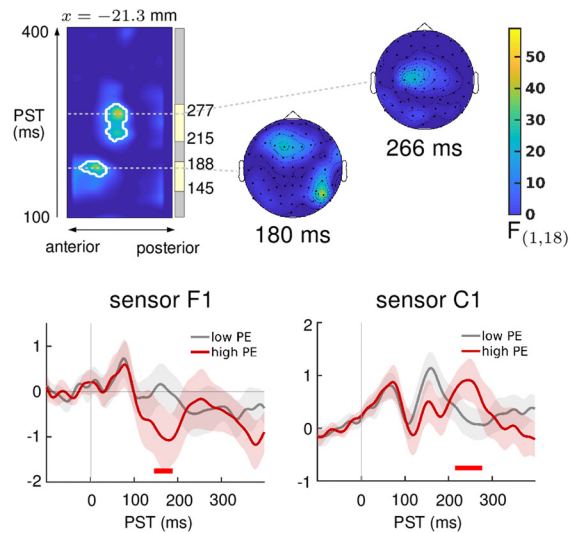


Figure 4. Results of the model-based EEG analysis in the placebo condition: effects of the high-level and the low-level PE. Top, The left side of each panel shows an F map across the scalp dimension y (from posterior to anterior, x -axis) and across peristimulus time (y -axis) at the spatial x -location indicated above the map. Significant F values ($p < 0.05$, whole-volume FWE corrected at the cluster level with a cluster-defining threshold of $p < 0.001$) are marked by white contours. Time windows of significant correlations are indicated by the yellow bars next to the colored clusters of significant F values. The scalp maps next to the F maps always show the F map at the indicated peristimulus time point, corresponding to the peak of that cluster, across a 2D representation of the sensor layout. We found significant correlations of the EEG signal with our two computational quantities across frontocentral and temporal channels. For the lower-level PE, ε_2 , the correlation peaked at 121 ms poststimulus at central channels; for the higher-level PE, ε_3 , it peaked at 180 ms at frontal channels, at 184 ms at temporal channels (data not shown here), and at 266 ms poststimulus at left central channels. Bottom, Average EEG responses to the 10% highest and the 10% lowest PE values at exemplary sensors within significant clusters. High values in low-level PEs correlated with an increased negativity between 102 and 207 ms poststimulus (sensor C1). High values in high-level PEs correlated with both an increased positivity (sensor P8) and an increased negativity between 145 and 188 ms poststimulus (sensor F1), and an increased central positivity between 215 and 277 ms poststimulus (sensor C1).

of a belief hierarchy are updated in response to two different precision-weighted PE signals (Mathys et al., 2011): a low-level PE that quantifies the mismatch between expected and actual tone transitions, and a higher-level PE that quantifies the change in estimated uncertainty about transition probabilities and is used to update estimates of environmental volatility. Effects of volatility on mismatch signals have been reported previously (Summerfield et al., 2011; Todd et al., 2014; Dzafic et al., 2018).

Notably, in the present study, the observed timing of low-level and high-level precision-weighted PE responses under placebo coincided with the timing of MMN and P3a components, respectively, previously shown to reflect related, but dissociable stages of automatic deviance processing (Rinne et al., 2006; Lecaigard et al., 2015). Furthermore, the temporal succession of these two PE signatures mirrored the temporal order that is predicted by the computational model.

Ketamine interferes with high-level belief updates

We found that ketamine changed the electrophysiological expression of the higher-level (but not lower-level) PE. Other authors have reported ketamine-induced changes of the deviant-related negativity at an earlier time corresponding to our lower-level PE representation and the classical MMN latency (Umbricht et al., 2000; Schmidt et al., 2012). One difficulty for comparing these reports to the current results is that the timing of ketamine effects in previously reported ERP analyses strongly depended on the type of MMN paradigm, the definition of standards, and the choice of electrodes and time windows (Oranje et al., 2000; Umbricht et al., 2000; Heekeren et al., 2008; Roser et al., 2011; Schmidt et al., 2012, 2013). For example, using classical averaging-based ERP analysis restricted to the early

MMN time window (100–200 ms after tone onset) and a subset of frontocentral and temporal channels, Schmidt et al. (2012) found an attenuation of early MMN amplitudes in frontal channels under ketamine in the same dataset used here. When repeating this ROI analysis here, but using a trial definition based on model-based estimates of PE, we found that mismatch effects were indeed attenuated by ketamine both in the early and a later time window (200–300 ms) in the frontal ROI for the higher-level PE, but not for the lower-level PE. This is consistent with our main single-trial analysis, where the high-level PE also showed an effect in frontal sensors within the early time window (Figs. 4, 5). However, this analysis, which considers all sensors and time points under multiple-comparison correction, locates the dominant effect of ketamine in the later time window of the P3a. This is also consistent with another set of ERP results from the same dataset (Schmidt et al., 2013) where, across all sensors and time points, a significant drug effect was found exclusively in a time window (220–240 ms) that was later than the classical MMN latency, and with literature on how ketamine attenuates later ERP components such as the P3 (Oranje et al., 2000; Watson et al., 2009; Rosburg and Schmidt, 2018).

Our finding that ketamine altered high-level PEs can also be compared with previous dynamic causal modeling (DCM) studies that examined the effects of ketamine during auditory roving MMN paradigms. While these studies (which used different approaches to modeling the input stream) gave different answers, both localized the effect of ketamine at higher levels of the auditory hierarchy. One study found that the effect of ketamine was best explained by changes of inhibition within frontal sources (Rosch et al., 2019). Previous DCM analyses of our

Ketamine

Effect of ε_2 (Lower-level PE)

Effect of ε_3 (Higher-level PE)

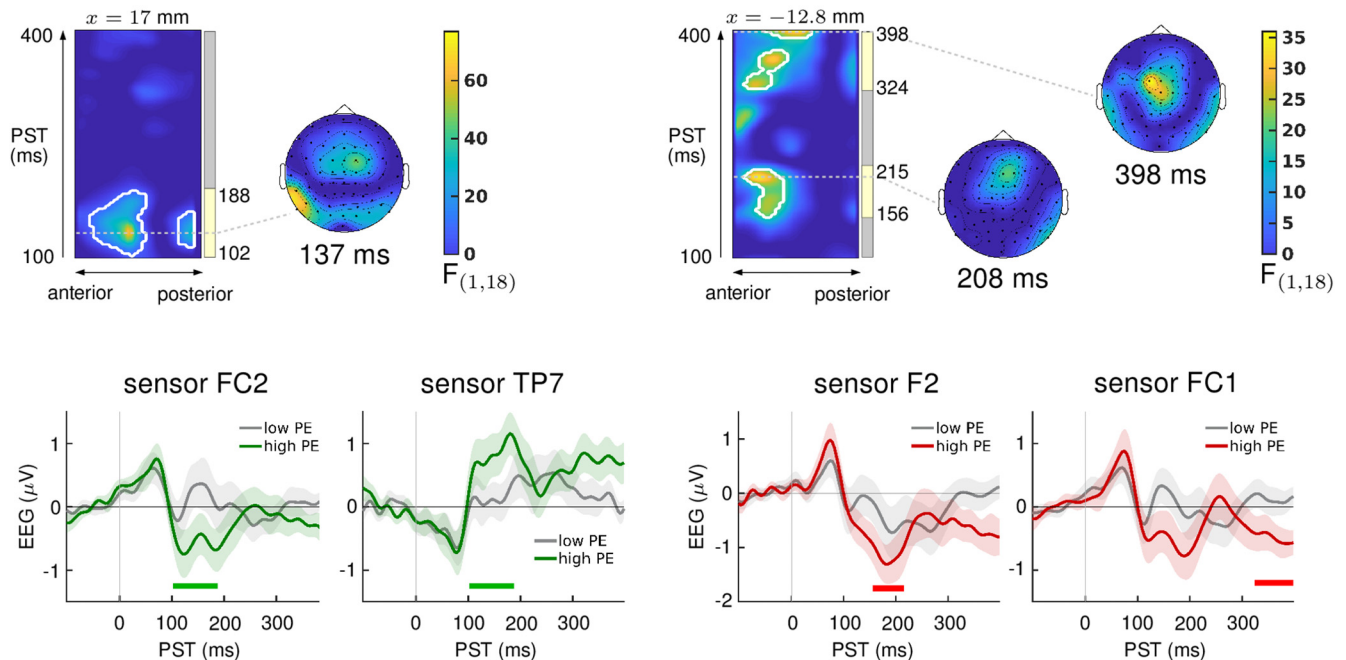


Figure 5. Results of the model-based EEG analysis in the ketamine condition: effects of the high-level and the low-level PE. Top, The left side of each panel shows an F map across the scalp dimension y (from posterior to anterior, x -axis) and across peristimulus time (y -axis) at the spatial x -location indicated above the map. Significant F values ($p < 0.05$, whole-volume FWE-corrected at the cluster level with a cluster-defining threshold of $p < 0.001$) are marked by white contours. Time windows of significant correlations are indicated by the yellow bars next to the colored clusters of significant F values. The scalp maps next to the F maps always show the F map at the indicated peristimulus time point, corresponding to the peak of that cluster, across a 2D representation of the sensor layout. We found significant correlations of the EEG signal with our two computational quantities across frontocentral and temporal channels. For the lower-level PE, ε_2 , the correlation peaked at 137 ms at frontocentral channels, and at 141 ms at left temporal channels (data not shown here). For the higher-level PE, ε_3 , it peaked at 208 and 398 ms after stimulus onset at frontocentral channels, and at 160 and 375 ms at right temporal channels (data not shown here). Bottom, Average EEG responses to the 10% highest and the 10% lowest PE values at exemplary sensors within significant clusters. High values in low-level PEs correlated with an increased negativity between 102 and 188 ms poststimulus (sensor FC2), and the reverse effect occurred in temporal channels between 105 and 188 ms (sensor TP7). High values in high-level PEs correlated with an increased negativity between 156 and 215 ms poststimulus (sensor F2), and an increased late frontocentral negativity between 324 and 398 ms poststimulus (sensor FC1).

dataset (Schmidt et al., 2013) suggested reduced bottom-up connectivity from auditory cortex (A1) to superior temporal gyrus (STG) under ketamine. Assuming that low-level and high-level PEs are computed at lower and higher levels of the auditory hierarchy (e.g., A1 and STG), respectively, this is compatible with disturbed computation of higher-level PEs in STG due to impaired message passing from A1.

It is important to note that our results do not allow for a unique interpretation of ketamine effects in computational terms. If one assumes a strictly monotonic relation between EEG amplitude and PEs, our finding suggests that ketamine reduces learning (smaller PEs) about environmental volatility. Because volatility estimates are a direct function of the PEs used to update them (Eq. 1), this can both lead to inflated estimates of volatility (slowed representation of stability after periods of inconstancy) or diminished ones (in the opposite case), depending on context. A previous study using ketamine found reduced stabilization of an internal model of environmental regularities during instrumental learning (Vinckier et al., 2016). One may be tempted to interpret this as an overestimation of volatility under ketamine; however, the previous model derived from a different computational concept, making direct comparisons problematic. Interestingly, an overestimation of volatility has been observed in patients with schizophrenia (Kaplan et al., 2016; Deserno et al., 2020) and individuals at risk for psychosis (Cole et al., 2020).

Limitations

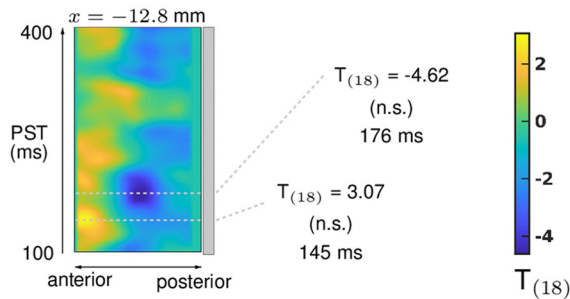
The HGF parameters allow for the expression of individual differences in learning (with potential relations to neuromodulatory mechanisms; Mathys et al., 2011; Vossel et al., 2014a). A main limitation of our approach is that we cannot infer on subject-specific learning styles, simply because the MMN paradigm does not provide behavioral responses to which the model could be fitted. Similar to Stefanics et al. (2018), we therefore used the parameters of a surprise-minimizing Bayesian observer for each of the tone sequences and simulated belief trajectories accordingly. An important future extension of HGF applications to MMN paradigms would be the formulation of a forward model from belief updates to EEG signals. This would allow for estimating subject-specific model parameters from single-trial EEG data directly.

A second limitation concerns the relatively small sample size ($N = 19$). This renders it difficult to interpret negative results, such as the lack of ketamine effects on low-level PEs. This will need to be addressed in future studies with larger samples and/or meta-analyses.

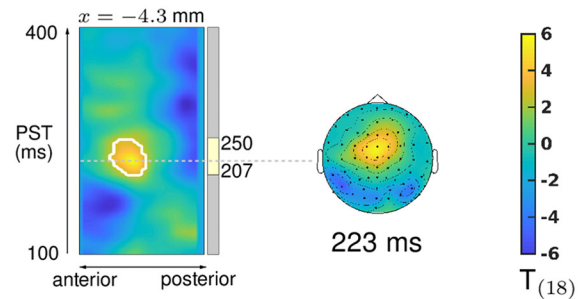
Finally, the particular roving paradigm used in this study was not optimized for investigating the effects of volatility, as the probabilities governing the auditory input stream are quite stable: repetitions are more likely than tone transitions, throughout the tone sequence. However, as the (subjective) inferred level of

Placebo > Ketamine

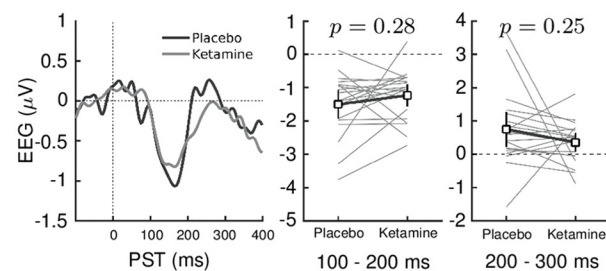
Effect on ε_2 (Lower-level PE)



Effect on ε_3 (Higher-level PE)



High PE – Low PE (ε_2) in frontal ROI (Fz, F3, F4)



High PE – Low PE (ε_3) in frontal ROI (Fz, F3, F4)

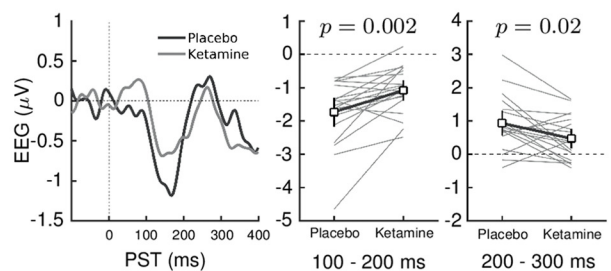


Figure 6. Drug effect on the representation of the lower-level PE (ε_2 , left) and the higher-level PE (ε_3 , right). Top, The left side of each panel shows a t map for the paired t test (placebo – ketamine) across the scalp dimension y (from posterior to anterior, x -axis) and across peristimulus time (y -axis) at the spatial x -location indicated above the map. Significant t values ($p < 0.05$, whole-volume FWE corrected at the cluster level with a cluster-defining threshold of $p < 0.001$) are marked by white contours. Time windows of significant correlations are indicated by the yellow bars next to the colored clusters of significant t values. The scalp map next to the t map shows the t map at the indicated peristimulus time point, corresponding to the peak of that cluster, across a 2D representation of the sensor layout. The effect of the higher-level PE was stronger under placebo compared with ketamine between 207 and 250 ms poststimulus, peaking at 223 ms at frontocentral channels. No significant drug effects were found for the lower-level PE. Bottom, Results of the ROI analysis. Left plots show the average difference wave (mismatch effect) for the contrast deviants (high PE) – standard (low PE) trials. Middle and right plots show peak mismatch effects (EEG amplitude differences) per participant in the placebo and the ketamine sessions (gray lines). Session averages are marked by small squares, with error bars showing ± 1.96 times the SEM. Middle plots, Peak mismatch effects between 100 and 200 ms. Right plots, Peak mismatch effects between 200 and 300 ms. The effect of ketamine was not significant when using the lower-level PE for the definition of standards and deviants. In contrast, ketamine significantly attenuated mismatch effects in both time windows for the higher-level PE in the frontal ROI. Please note that in this case we are not interested in the interaction (and are thus not committing the common error of confusing a difference in significance with a significant difference) but do wish to test both PE types separately to enable a comparison with previous results by Schmidt et al. (2012).

volatility determines an individual's learning rate, participants still need to infer the adequate level of volatility as they perform the task. This learning process is reflected by monotonic changes in log-volatility estimates in the belief trajectories of the surprise minimizing agent (Fig. 2). It is important to note, however, that standard and deviant trials affect log-volatility estimates differentially, and the resulting PE trajectory used as a regressor here does not simply correspond to a drift-like signal but properly reflects trial-by-trial belief updates. Still, future work that follows up on our current findings would benefit from using mismatch paradigms designed to include marked changes of volatility across time (Summerfield et al., 2011; Todd et al., 2014; Dzafic et al., 2018).

Conclusion and outlook

This study presents evidence for the role of hierarchically related PEs in the auditory MMN. While ketamine-induced reductions of MMN have been reported previously, our study enables two new insights by taking an explicitly computational perspective

and analyzing trial-by-trial belief updates. First, we offer an interpretation of two mismatch-related ERP components, the MMN and the P3a, in terms of hierarchically related PEs that are expressed trial-by-trial and reflect the updating of a hierarchical model of the statistical structure of the environment. Additionally, a reduced expression of the higher-level PE under infusion of S-ketamine suggests a disturbance of high-level inference about environmental volatility by perturbation of NMDA receptors (Coull et al., 2011).

Our results are clinically relevant as they support a bridge between physiology (NMDAR function) and computation (hierarchical Bayesian inference), as proposed by predictive coding theories of schizophrenia. By linking physiological indices of abnormal perceptual inference to their algorithmic interpretation in terms of hierarchically related PEs, the present work provides a starting point for future attempts to understand individual alterations of MMN in schizophrenia mechanistically. We hope that this will eventually contribute to the development of computational assays for improved differential diagnosis and treatment prediction in schizophrenia (Stephan et al., 2015).

References

- Adams RA, Stephan KE, Brown HR, Frith CD, Friston KJ (2013) The computational anatomy of psychosis. *Front Psychiatry* 4:47.
- Auksztulewicz R, Friston KJ, Nobre AC (2017) Task relevance modulates the behavioural and neural effects of sensory predictions. *PLoS Biol* 15:e2003143.
- Avissar M, Xie S, Vail B, Lopez-Calderon J, Wang Y, Javitt DC (2018) Meta-analysis of mismatch negativity to simple versus complex deviants in schizophrenia. *Schizophr Res* 191:25–34.
- Baldeweg T, Klugman A, Gruzelić J, Hirsch SR (2004) Mismatch negativity potentials and cognitive impairment in schizophrenia. *Schizophr Res* 69:203–217.
- Cole DM, Diaconescu AO, Pfeiffer UJ, Brodersen KH, Mathys CD, Julkowsky D, Ruhrmann S, Schilbach L, Tittgemeyer M, Vogetley K, Stephan KE (2020) Atypical processing of uncertainty in individuals at risk for psychosis. *Neuroimage Clin* 26:102239.
- Corlett PR, Honey GD, Fletcher PC (2007) From prediction error to psychosis: ketamine as a pharmacological model of delusions. *J Psychopharmacol* 21:238–252.
- Corlett PR, Honey GD, Krystal JH, Fletcher PC (2011) Glutamatergic model psychoses: prediction error, learning, and inference. *Neuropsychopharmacology* 36:294–315.
- Corlett PR, Honey GD, Fletcher PC (2016) Prediction error, ketamine and psychosis: an updated model. *J Psychopharmacol* 30:1145–1155.
- Coull JT, Morgan H, Cambridge VC, Moore JW, Giorlando F, Adapa R, Corlett PR, Fletcher PC (2011) Ketamine perturbs perception of the flow of time in healthy volunteers. *Psychopharmacology (Berl)* 218:543–556.
- Cowan N, Winkler I, Teder W, Näätänen R (1993) Memory prerequisites of mismatch negativity in the auditory event-related potential (ERP). *J Exp Psychol Learn Mem Cogn* 19:909–921.
- Dayan P, Hinton GE, Neal RM, Zemel RS (1995) The Helmholtz machine. *Neural Comput* 7:889–904.
- Deserno L, Boehme R, Mathys C, Katthagen T, Kaminski J, Stephan KE, Heinz A, Schlagenhauf F (2020) Volatility estimates increase choice switching and relate to prefrontal activity in schizophrenia. *Biol Psychiatry Cogn Neurosci Neuroimaging* 5:173–183.
- Diaconescu AO, Mathys CD, Weber LAE, Daunizeau J, Kasper L, Lomakina EI, Fehr E, Stephan KE (2014) Inferring on the intentions of others by hierarchical Bayesian learning. *PLoS Comput Biol* 10:e1003810.
- Diaconescu AO, Mathys CD, Weber LAE, Kasper L, Mauer J, Stephan KE (2017) Hierarchical prediction errors in midbrain and septum during social learning. *Soc Cogn Affect Neurosci* 12:618–634.
- Doya K, Ishii S, Pouget A, Rao RPN (2011) Bayesian brain: probabilistic approaches to neural coding. Cambridge, MA: MIT.
- Dzafic I, Randeniya R, Garrido MI (2018) Reduced top-down connectivity as an underlying mechanism for psychotic experiences in healthy people. *bioRxiv*. Advance online publication. Retrieved April 16, 2018. doi: 10.1101/296988.
- Erickson MA, Ruffle A, Gold JM (2016) A meta-analysis of mismatch negativity in schizophrenia: from clinical risk to disease specificity and progression. *Biol Psychiatry* 79:980–987.
- Feng N, Vollenweider FX, Minder EI, Rentsch K, Grampp T, Vonderschmitt DJ (1995) Development of a gas chromatography-mass spectrometry method for determination of ketamine in plasma and its application to human samples. *Ther Drug Monit* 17:95–100.
- Flandin G, Friston KJ (2019) Analysis of family-wise error rates in statistical parametric mapping using random field theory. *Hum Brain Mapp* 40:2052–2054.
- Friston KJ (2005) A theory of cortical responses A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci* 360:815–836.
- Friston KJ (2010) The free-energy principle: a unified brain theory? *Nat Rev Neurosci* 11:127–138.
- Friston KJ, Brown HR, Siemerkus J, Stephan KE (2016) The dysconnection hypothesis (2016). *Schizophr Res* 176:83–94.
- Garrido MI, Friston KJ, Kiebel SJ, Stephan KE, Baldeweg T, Kilner JM (2008) The functional anatomy of the MMN: a DCM study of the roving paradigm. *Neuroimage* 42:936–944.
- Garrido MI, Kilner JM, Stephan KE, Friston KJ (2009) The mismatch negativity: a review of underlying mechanisms. *Clin Neurophysiol* 120:453–463.
- Goff DC, Coyle JT (2001) The emerging role of glutamate in the pathophysiology and treatment of schizophrenia. *Am J Psychiatry* 158:1367–1377.
- Heekeren K, Daumann J, Neukirch A, Stock C, Kawohl W, Norra C, Waberski TD, Gouzoulis-Mayfrank E (2008) Mismatch negativity generation in the human 5HT_{2A} agonist and NMDA antagonist model of psychosis. *Psychopharmacology (Berl)* 199:77–88.
- Iglesias S, Mathys C, Brodersen KH, Kasper L, Piccirelli M, den Ouden HEM, Stephan KE (2013) Hierarchical Prediction Errors in Midbrain and Basal Forebrain during Sensory Learning. *Neuron* 80:519–530.
- Javitt DC, Steinschneider M, Schroeder CE, Arezzo JC (1996) Role of cortical N-methyl-D-aspartate receptors in auditory sensory memory and mismatch negativity generation: implications for schizophrenia. *Proc Natl Acad Sci U S A* 93:11962–11967.
- Jepma M, Murphy PR, Nassar MR, Rangel-Gomez M, Meeter M, Nieuwenhuis S (2016) Catecholaminergic regulation of learning rate in a dynamic environment. *PLoS Comput Biol* 12:e1005171.
- Kaplan CM, Saha D, Molina JL, Hockeimer WD, Postell EM, Apud JA, Weinberger DR, Tan HY (2016) Estimating changing contexts in schizophrenia. *Brain* 139:2082–2095.
- Kiebel SJ, Friston KJ (2004) Statistical parametric mapping for event-related potentials: I. Generic considerations. *Neuroimage* 22:492–502.
- Kilner JM, Friston KJ (2010) Topological inference for EEG and MEG. *Ann Appl Stat* 4:1272–1290.
- Kolossa A, Kopp B, Fingscheidt T (2015) A computational analysis of the neural bases of Bayesian inference. *Neuroimage* 106:222–237.
- Lecaignard F, Bertrand O, Gimenez G, Mattout J, Caclin A (2015) Implicit learning of predictable sound sequences modulates human brain responses at different levels of the auditory hierarchy. *Front Hum Neurosci* 9:505.
- Lieder F, Daunizeau J, Garrido MI, Friston KJ, Stephan KE (2013a) Modelling trial-by-trial changes in the mismatch negativity. *PLoS Comput Biol* 9:e1002911.
- Lieder F, Stephan KE, Daunizeau J, Garrido MI, Friston KJ (2013b) A neuro-computational model of the mismatch negativity. *PLoS Comput Biol* 9:e1003288.
- Mars RB, Debener S, Gladwin TE, Harrison LM, Haggard P, Rothwell JC, Bestmann S (2008) Trial-by-trial fluctuations in the event-related electroencephalogram reflect dynamic changes in the degree of surprise. *J Neurosci* 28:12539–12545.
- Mathys CD, Daunizeau J, Friston KJ, Stephan KE (2011) A Bayesian foundation for individual learning under uncertainty. *Front Hum Neurosci* 5:39.
- Mathys CD, Lomakina EI, Daunizeau J, Iglesias S, Brodersen KH, Friston KJ, Stephan KE (2014) Uncertainty in perception and the hierarchical Gaussian filter. *Front Hum Neurosci* 8:825.
- Näätänen R (2000) Mismatch negativity (MMN): perspectives for application. *Int J Psychophysiol* 37:3–10.
- Näätänen R, Tervaniemi M, Sussman E, Paavilainen P, Winkler I (2001) “Primitive intelligence” in the auditory cortex. *Trends Neurosci* 24:283–288.
- Näätänen R, Jacobsen T, Winkler I (2005) Memory-based or afferent processes in mismatch negativity (MMN): a review of the evidence. *Psychophysiology* 42:25–32.
- Olney JW, Farber NB (1995) Glutamate receptor dysfunction and schizophrenia. *Arch Gen Psychiatry* 52:998–1007.
- Oranje B, van Berckel BNM, Kemner C, van Ree JM, Kahn RS, Verbaten MN (2000) The effects of a sub-anaesthetic dose of ketamine on human selective attention. *Neuropsychopharmacology* 22:293–302.
- Ostwald D, Spitzer B, Guggenmos M, Schmidt TT, Kiebel SJ, Blankenburg F (2012) Evidence for neural encoding of Bayesian surprise in human somatosensation. *Neuroimage* 62:177–188.
- Paavilainen P, Jaramillo M, Näätänen R, Winkler I (1999) Neuronal populations in the human brain extracting invariant relationships from acoustic variance. *Neurosci Lett* 265:179–182.
- Palmer CE, Auksztulewicz R, Ondobaka S, Kilner JM (2019) Sensorimotor beta power reflects the precision-weighting afforded to sensory prediction errors. *Neuroimage* 200:59–71.
- Penny WD, Holmes AJ (2007) Random effects analysis. In: *Statistical parametric mapping: the analysis of functional brain images* (Penny WD, Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE, eds), pp 156–165. London: Academic.
- Polich J (2007) Updating P300: an integrative theory of P3a and P3b. *Clin Neurophysiol* 118:2128–2148.

- Rao RPN, Ballard DH (1999) Predictive Coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* 2:79–87.
- Rinne T, Särkkä A, Degerman A, Schröger E, Alho K (2006) Two separate mechanisms underlie auditory change detection and involuntary control of attention. *Brain Res* 1077:135–143.
- Rosburg T, Schmidt A (2018) Potential mechanisms for the ketamine-induced reduction of P3b amplitudes. *Front Behav Neurosci* 12:308.
- Rosch RE, Auksztulewicz R, Leung PD, Friston KJ, Baldeweg T (2019) Selective prefrontal disinhibition in a roving auditory oddball paradigm under N-methyl-D-aspartate receptor blockade. *Biol Psychiatry Cogn Neurosci Neuroimaging* 4:140–150.
- Roser P, Haussleiter IS, Chong H-J, Maier C, Kawohl W, Norra C, Juckel G (2011) Inhibition of cerebral type 1 cannabinoid receptors is associated with impaired auditory mismatch negativity generation in the ketamine model of schizophrenia. *Psychopharmacology (Berl)* 218:611–620.
- Schmidt A, Bachmann R, Kometer M, Csomor PA, Stephan KE, Seifritz E, Vollenweider FX (2012) Mismatch negativity encoding of prediction errors predicts S-ketamine-induced cognitive impairments. *Neuropsychopharmacology* 37:865–875.
- Schmidt A, Diaconescu AO, Kometer M, Friston KJ, Stephan KE, Vollenweider FX (2013) Modeling ketamine effects on synaptic plasticity during the mismatch negativity. *Cereb Cortex* 23:2394–2406.
- Schneider W, Eschman A, Zuccolotto A (2002) E-Prime reference guide. Pittsburgh: Psychology Software Tools Inc.
- Self MW, Kooijmans RN, Supèr H, Lamme VA, Roelfsema PR (2012) Different glutamate receptors convey feedforward and recurrent processing in macaque V1. *Proc Natl Acad Sci U S A* 109:11031–11036.
- Stefanics G, Heinzle J, Horváth AA, Stephan KE (2018) Visual mismatch and predictive coding: a computational single-trial ERP study. *J Neurosci* 38:4020–4030.
- Stephan KE, Baldeweg T, Friston KJ (2006) Synaptic plasticity and dysconnection in schizophrenia. *Biol Psychiatry* 59:929–939.
- Stephan KE, Friston KJ, Frith CD (2009) Dysconnection in schizophrenia: from abnormal synaptic plasticity to failures of self-monitoring. *Schizophr Bull* 35:509–527.
- Stephan KE, Iglesias S, Heinzle J, Diaconescu AO (2015) Translational perspectives for computational neuroimaging. *Neuron* 87:716–732.
- Sterzer P, Adams RA, Fletcher P, Frith C, Lawrie SM, Muckli L, Petrovic P, Uhlhaas P, Voss M, Corlett PR (2018) The predictive coding account of psychosis. *Biol Psychiatry* 84:634–643.
- Summerfield C, Wyart V, Johnen VM, de Gardelle V (2011) Human scalp electroencephalography reveals that repetition suppression varies with expectation. *Front Hum Neurosci* 5:67.
- Todd J, Harms L, Schall U, Michie PT (2013) Mismatch negativity: translating the potential. *Front Psychiatry* 4:171.
- Todd J, Heathcote A, Whitson LR, Mullens D, Provost A, Winkler I (2014) Mismatch negativity (MMN) to pitch change is susceptible to order-dependent bias. *Front Neurosci* 8:180.
- Umbricht D, Krljes S (2005) Mismatch negativity in schizophrenia: a meta-analysis. *Schizophr Res* 76:1–23.
- Umbricht D, Schmid L, Koller R, Vollenweider FX, Hell D, Javitt DC (2000) Ketamine-induced deficits in auditory and visual context-dependent processing in healthy volunteers: implications for models of cognitive deficits in schizophrenia. *Arch Gen Psychiatry* 57:1139–1147.
- Vinckier F, Gaillard R, Palminteri S, Rigoux L, Salvador A, Fornito A, Adapa R, Krebs MO, Pessiglione M, Fletcher PC (2016) Confidence and psychosis: a neuro-computational account of contingency learning disruption by NMDA blockade. *Mol Psychiatry* 21:946–955.
- Vollenweider FX, Leenders KL, Oye I, Hell D, Angst J (1997) Differential psychopathology and patterns of cerebral glucose utilisation produced by (S)- and (R)-ketamine in healthy volunteers using positron emission tomography (PET). *Eur Neuropsychopharmacol* 7:25–38.
- Vossel S, Bauer M, Mathys CD, Adams RA, Dolan RJ, Stephan KE, Friston KJ (2014a) Cholinergic stimulation enhances Bayesian belief updating in the deployment of spatial attention. *J Neurosci* 34:15735–15742.
- Vossel S, Mathys CD, Daunizeau J, Bauer M, Driver J, Friston KJ, Stephan KE (2014b) Spatial attention, precision, and Bayesian inference: a study of saccadic response speed. *Cereb Cortex* 24:1436–1450.
- Wacongne C, Changeux J-P, Dehaene S (2012) A neuronal model of predictive coding accounting for the mismatch negativity. *J Neurosci* 32:3665–3678.
- Watson TD, Petrakis IL, Edgecombe J, Perrino A, Krystal JH, Mathalon DH (2009) Modulation of the cortical processing of novel and target stimuli by drugs affecting glutamate and GABA neurotransmission. *Int J Neuropsychopharmacol* 12:357–370.
- Weilnhammer VA, Stuke H, Sterzer P, Schmack K (2018) The neural correlates of hierarchical predictions for perceptual decisions. *J Neurosci* 38:5008–5021.
- Winkler I (2007) Interpreting the mismatch negativity. *J Psychophysiol* 21:147–163.
- Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC (1996) A unified statistical approach for determining significant signals in images of cerebral activation. *Hum Brain Mapp* 4:58–73.