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¶], Christoph Mathys[3], André Schmidt[2], Michael Kometer[4], Franz Vollenweider[4], Klaas E. Stephan[1,5,6]

[1] Translational Neuromodeling Unit (TNU), Institute for Biomedical Engineering, University of Zurich and ETH Zurich, Zurich, Switzerland
[2] Dept. of Psychiatry, University of Basel, Basel, Switzerland
[3] SISSA, Trieste, Italy
[4] Neuropharmacology and Brain Imaging, University Hospital of Psychiatry, University of Zurich, Zurich, Switzerland
[5] Wellcome Centre for Human Neuroimaging, University College London, London, United Kingdom
[6] Max Planck Institute for Metabolism Research, Cologne, Germany.

[¶]These authors contributed equally to this work.

* Corresponding author:
Email: weber@biomed.ee.ethz.ch

Keywords: EEG, NMDA receptor, predictive coding, hierarchical Bayesian inference, schizophrenia, neuromodeling

Abstract

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, predictive coding views perceptual inference as an NMDAR-dependent process of minimizing hierarchical precision-weighted prediction errors (PEs). Disturbances of this putative process play a key role in hierarchical Bayesian theories of schizophrenia.

Here, we provide empirical evidence for this clinical theory, demonstrating the existence of multiple, hierarchically related PEs in a “roving MMN” paradigm. We applied a computational model, the Hierarchical Gaussian Filter (HGF), to single-trial EEG data from healthy volunteers that 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 computational trial-by-trial analysis indicated that low-level PEs (about stimulus transitions) are expressed early (102-207ms post-stimulus), while high-level PEs (about transition probability) are reflected by later components (152-199ms, 215-277ms) of single-trial responses. Furthermore, we find that ketamine significantly diminished the expression of high-level PE responses, implying that NMDAR antagonism disrupts inference on abstract statistical regularities.

Our findings are consistent with long-standing notions that NMDAR dysfunction may cause positive symptoms in schizophrenia by impairing hierarchical Bayesian inference about the world’s statistical structure. Beyond their relevance for schizophrenia, our results illustrate the potential of computational single-trial analyses for assessing potential disease mechanisms.

Introduction

The auditory mismatch negativity (MMN), an electrophysiological response to rule violations in auditory input streams, has long served as an empirical demonstration that the brain learns the statistical structure of its environment and predicts future sensory inputs (1–3). It plays an important role in psychiatric research, as it fulfils several criteria for a translational biomarker for schizophrenia (4, 5). Most importantly, a reduction in MMN amplitude is one of the most robust electrophysiological abnormalities in patients with schizophrenia (4–7).

Physiologically, MMN has been shown to depend on intact NMDA (N-methyl-D-aspartic acid) receptor signalling. Following an initial study in monkeys (8), human EEG studies (9–11) using the NMDA receptor (NMDAR) antagonist ketamine also found a significant reduction of MMN responses, although the results show non-trivial variations with ketamine dose, paradigm type and trial definition (12–14). From a neuropharmacological perspective, this renders the MMN paradigm an interesting potential readout of NMDAR function.

The robust impairment of MMN in schizophrenia, and the fact that a 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, leading to both cognitive and perceptual abnormalities and positive symptoms (15–21). This has been interpreted as an impairment of perceptual inference under a predictive coding view. In this “Bayesian brain” framework, the brain continuously updates a hierarchical model of its environment to infer the causes of its sensory inputs and predict future events (22–25).

The auditory MMN is believed to reflect such a model update during perceptual inference within the auditory processing hierarchy (3, 26, 27). In particular, in predictive coding, each level of a cortical hierarchy provides predictions about the state of the level below and, in turn, receives a prediction error (PE) signal that reflects the discrepancy between the prediction and the actual state of the level below; this PE signal then serves to update the prediction. This updating process rests on hierarchical message passing between cortical regions, until PEs are minimized on all levels of the hierarchy. While predictions are thought to be communicated by descending (backward) connections, drawing predominantly on glutamatergic NMDAR signaling, sensory PEs are signalled by ascending (forward) connections mainly via glutamatergic AMPA receptors (28). Critically, these ascending PE signals are weighted by the relative precision of bottom-up (sensory) input compared to predictions (priors) from higher levels. The MMN, which is a difference waveform, is then commonly interpreted as the difference in precision-weighted PEs between surprising events (‘deviants’) and more predictable events (‘standards’). Physiologically, the precision-weighting could be plausibly implemented by changes in postsynaptic gain, mediated by neuromodulatory or GABAergic influences (29).

Critically, the predictive coding perspective, which understands the MMN as a reflection of perceptual inference in the auditory cortical hierarchy, makes two important predictions:

First, multiple, hierarchically related precision-weighted PEs should underlie the MMN (27). For example, multiple precision-weighted PEs arise when considering volatility effects during learning (30–32). Volatility determines the learning rate, and even when the real volatility is low or absent in a cognitive paradigm, participants still need to infer the adequate level of volatility as they perform the task. Moreover, in perceptual learning paradigms like MMN, trial-by-trial changes in evoked responses (as measured by EEG) carry information about the temporal dynamics of this learning process (33–38). A suitable model for incorporating volatility in trial-

by-trial Bayesian belief updates is the Hierarchical Gaussian Filter (HGF) (30, 39), which quantifies the trajectories of hierarchically related PEs.

Second, the expression of precision-weighted PEs should be sensitive to NMDAR manipulations. According to the framework outlined above, a blockade of NMDARs would lead to a reduction of top-down (predictive) signalling, resulting in less constrained low-level inference about the causes of sensory inputs, and potentially aberrant bottom-up (PE) signalling (18, 19, 40). This could render all events equally surprising and thus blur differences between standard and deviant events on which the MMN rests. Such aberrant hierarchical Bayesian inference due to disturbances in the relative weighting of prior beliefs and prediction errors on multiple (sensory and cognitive) levels is at the heart of current computational accounts of schizophrenia (15–17, 21, 41, 42).

Here, we use a computational single-trial EEG analysis to test these predictions. In particular, we apply the HGF to data from a previously published MMN study (10) that administered an acute dose of S-ketamine to healthy volunteers, using a placebo-controlled, within-subject design. In brief, our analyses suggest that (i) hierarchically related precision-weighted PEs underlie the MMN and are expressed sequentially, corresponding to early and late components of evoked auditory responses; and (ii) ketamine affects the representation of high-level PEs that serve to update estimates of environmental volatility.

Methods and Materials

Details on the participants, drug administration, and data acquisition have previously been provided in (10) and (13) and the interested reader is referred to these papers for more information. Here, we focus on the model-based EEG analysis and only briefly summarize the other aspects.

PARTICIPANTS

19 healthy subjects (twelve 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. The use of psychoactive drugs was approved by the Swiss Federal Health Office, Department of Pharmacology and Narcotics (DPN), Bern, Switzerland. For further examinations prior to inclusion and additional questionnaire assessments, see (10).

EXPERIMENTAL PROCEDURE AND DATA PREPROCESSING

The two sessions (placebo and S-ketamine) that all subjects underwent in a counterbalanced fashion were separated by at least two weeks. Both subjects and the experimenter interacting with them were blind to the drug order. For details on the procedure and administration of S-ketamine, please see Supplementary Material.

Electroencephalographic (EEG) activity was recorded during an auditory "roving" oddball paradigm, originally developed by (43) and subsequently modified by (44). The EEG was recorded at a sampling rate of 512 Hz using a Biosemi system with 64 scalp electrodes. Pre-processing and data analysis was performed using SPM12 (http://www.fil.ion.ucl.ac.uk/spm/) and included high- and lowpass filtering and rejection of trials contaminated by eye blinks, as well as bad channels. For details on the paradigm and preprocessing, the reader is referred to the Supplementary Material. 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 non-sphericity was accommodated by our second-level statistical tests (paired t-tests to assess group differences), see Methods 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.

MODEL-BASED ANALYSIS

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

Perceptual Model: The Hierarchical Gaussian Filter (HGF)

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 (30) that has been applied in various contexts, such as associative learning (31, 45), social learning (32, 46), spatial attention (47), or visual discrimination (48).

In the present task, participants were exposed to a tone sequence with 7 different tones. Our modeling approach assumes that in this context, an agent infers two hidden states in the world: (i) a matrix $\mathbf{X}_2$ of pair-wise transition probabilities between all tones and (ii) the volatility x_3 (i.e., the degree to which these probabilities change from trial to trial). Figure 1 shows a visualization of the corresponding generative model.

On each trial, the agent updates his/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:

$$\Delta\mu_i^{(k)} \propto \frac{\hat{\pi}_{i-1}^{(k)}}{\pi_i^{(k)}} \delta_{i-1}^{(k)}. \tag{6}$$

Here, $\delta_{i-1}^{(k)}$ denotes the PE about (the value or volatility of) 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 an agent's belief updates 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.

The specific update equations for the two levels of our model are given in the Supplementary Material. For a detailed derivation of the update equations and the updates of the precisions, the interested reader is referred to (30). Usually, in the HGF, subject-specific perceptual parameters describe the individual learning style of an agent. 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 (learning rates on both hierarchical levels, and starting values of the beliefs) of a surprise-minimizing Bayesian observer for all participants (for details, see Supplementary Material). The resulting PE trajectories (Figure 2) were subsequently used as regressors in a general linear model (GLM).

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 (27, 49, 26, 36, 50, 3, 51). In our model, two hierarchical levels are updated in response to new auditory inputs: the current estimate of the transition probabilities ($\boldsymbol{\mu}_2$), and the current estimate of environmental volatility (μ_3). The

corresponding precision-weighted PEs driving these updates are hierarchically related and are computed sequentially: the agent first needs to update $\boldsymbol{\mu_2}$ (using the low-level PE about $\boldsymbol{\mu_1}$) before evaluating its high-level PE with respect to $\boldsymbol{\mu_2}$, which is then used to update μ_3.

The questions we address in this paper are whether these precision-weighted PEs, which we denote by ε_2 and ε_3,

(i) are reflected by trial-by-trial variations in the amplitude of evoked responses;
(ii) their hierarchical relation in the model is mirrored by a corresponding temporal relation in their electrophysiological correlates;
(iii) whether NMDAR antagonism by S-ketamine alters the electrophysiological expression of these PEs.

Single-trial EEG analysis: The General Linear Model

We looked for 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 100ms to 400ms post-stimulus. We focused on this time window in order 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 mm × 5.38 mm × 3.33 ms. The images were constructed using linear interpolation for removed bad channels and smoothing to accommodate for between-subject spatial variability in channel space.

Our vectors of precision-weighted PEs served as regressors in a GLM of trial-wise EEG signals for each subject and each session separately, correcting for multiple comparisons over the entire time-sensor matrix, using Gaussian random field theory (52). We did not orthogonalise the regressors. Figure 2 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 (53). We employed 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 family-wise error (FWE) correction, based on Gaussian random field theory, across the entire volume (time×sensor space) at the cluster level ($p<0.05$) with a cluster defining threshold (CDT) of $p<0.001$ (54). Notably, all reported results also survive whole-volume correction at the peak-level ($p<0.05$); the associated p-values are not included in the main text but listed in the tables.

Results

For each computational quantity of interest, our model-based EEG analysis proceeded in two steps: first, we performed whole-volume (spatiotemporal) analyses to search for representations of our quantities in single-trial EEG responses; second, we examined whether these electrophysiological representations of trial-wise 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 102ms and 207ms post-stimulus, peaking at 121ms at central channels (whole-volume cluster-level FWE corrected, p=2.8e-08, with a cluster-defining threshold (CDT) of p<0.001; Figure 3; Table 1). This time window includes the typical time when the negativity of the roving MMN is observed (43, 44, 49). This suggests that the MMN typically observed in roving MMN 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 fronto-central electrodes between 10ms and 188ms (p=3.1e-08), and at left temporal channels between 105ms and 188ms, peaking at 141ms post-stimulus (p=6.3e-06; Table 2).

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 about transition probabilities that serves to update log volatility estimates) and EEG activity, both in an early time window (152ms to 199ms, peaking at 184ms at right temporal channels, p=0.004, and from 145ms to 188ms, peaking at 180ms at frontal channels, p=0.009) and in a later time window (between 215ms and 277ms, peaking at 266ms post-stimulus, p=0.002; Figure 3; Table 1), where high-level prediction errors correlated with an increased central positivity corresponding to the P3a component of the auditory evoked potential (55).

Under ketamine, we found a similar relationship of EEG amplitudes with the higher-level PE in the early time window (148ms to 211ms, peaking at 160ms at left temporal channels, p=0.04, and 156ms to 215ms, peaking at 207ms at fronto-central channels, p=0.008), but the later cluster occurred only much later (297ms to 398ms, peaking at 375ms at left temporal channels, p=0.021, and 324ms to 398ms, peaking at 398ms at fronto-central channels, p=0.001; Table 2). While the timing of this late effect is reminiscent of the P3b component, its scalp distribution was not centered on parietal channels, as would be characteristic for P3b (55, 56), but instead looked very similar to the earlier cluster, with a peak at fronto-central channels.

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 compared to the placebo condition. In contrast, the activation by $\varepsilon_3^{(k)}$, the higher-level PE about transition probabilities, was significantly reduced under ketamine as compared to placebo in a time window between 207ms and 250ms post stimulus, peaking at 223ms across

fronto-central channels ($p = 0.005$; Figure 4; Table 3). 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.

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 (17, 18, 21). Neurobiologically, this disturbance of hierarchical Bayesian inference is thought to result from alterations of NMDAR-dependent synaptic plasticity and to be reflected by abnormalities in perceptual paradigms, such as the auditory mismatch negativity (MMN) (15, 16, 20).

Here, we provide supporting evidence for two important predictions of this framework using EEG data from a roving auditory MMN paradigm: First, we looked for a signature of hierarchically related PE responses, as predicted by our computational model, in trial-wise EEG amplitudes and tested whether any such activity would show the expected temporal relationship. We found that low-level precision-weighted PE signals, that update an agent's beliefs about tone transition probabilities, occurred first and correlated with the negative amplitude of evoked responses between 102ms and 207ms post-stimulus across fronto-central channels. High-level precision-weighted PEs about transition probabilities (which update estimates of volatility) correlated with EEG amplitudes in these channels somewhat later, between 152ms and 199ms post-stimulus, and additionally correlated with an increased positivity between 215ms and 277ms across central channels.

Second, we examined how ketamine, a non-competitive NMDAR antagonist, impacted on the electrophysiological expression of these precision-weighted PEs. While we found no significant differences in activity related to low-level PEs under ketamine compared to placebo, activation by higher-level PEs was significantly reduced under ketamine in a time window between 207ms and 250ms post-stimulus, peaking at 223ms across fronto-central channels. This suggests that inhibition of NMDA receptors by ketamine disturbed higher-level updates of volatility in the brain's inference hierarchy.

MULTIPLE, HIERARCHICALLY RELATED PREDICTION ERRORS UNDERLIE THE MMN

The auditory MMN has been interpreted as reflecting model updates in an auditory processing hierarchy (26, 36, 49). In our Bayesian learning model, levels of a belief hierarchy are updated in response to different precision-weighted PE signals (30): a low-level PE that quantifies the mismatch between expected and actual tone transitions and a higher-level PE that reflects how much estimates of transition probability must be updated in the light of new input and which is used to update estimates of environmental volatility. Effects of volatility on mismatch signals have been reported previously (57–59).

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 (60, 61). Furthermore, the temporal succession of these two PE signatures mirrored the temporal order as predicted by the computational model.

KETAMINE INTERFERES WITH HIGH-LEVEL BELIEF UPDATES

We found that ketamine changed the electrophysiological expression of the higher-level PE while leaving the representation of the earlier, stimulus-related transition PE intact. This is

consistent with earlier ERP results from the same dataset (13) – where a significant drug effect was found exclusively in a time window (220-240ms after tone onset) that was later than the classical MMN latency – and with literature on how ketamine attenuates later ERP components such as the P3 (12, 56, 62).

Our finding that ketamine altered high-level PEs can also be compared to 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 (63). Previous DCM analyses of our own dataset (13) suggested reduced bottom-up connectivity from auditory cortex (A1) to superior temporal gyrus (STG) under ketamine, compatible with disturbed computation of higher-level PEs in STG by impairing message passing from A1.

Interestingly, in our study, the high-level PE showed an effect under ketamine both in early and late time windows (Table 2). While the late effect corresponded to the P3a under placebo, it occurred later under ketamine, around the typical time of the P3b and other ERP components related to conscious processing and context updating (55, 64, 65). Speculatively, this could reflect a less automatic, stimulus-driven processing (as typically associated with P3a) (55, 56, 65) of volatility under ketamine.

Our results are physiological and 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 about environmental volatility. Depending on context, this can both lead to inflated estimates of volatility (slowed representation of stability after periods of inconstancy) or diminished ones (in the opposite case). A previous study using ketamine found reduced stabilization of an internal model of environmental regularities during instrumental learning (66). 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.

One would further expect that a misrepresentation of environmental volatility also impacts on the lower-level PE via its precision weight – e.g., higher estimates of volatility lead to higher belief uncertainty and thus to larger model updates. Indeed, other authors have reported ketamine-induced modulations of the deviant-related negativity in an earlier time window corresponding to our lower-level PE representation and the classical MMN latency (9). However, the timing of the ketamine effect in previously reported ERP analyses strongly depended on the type of auditory mismatch paradigm, the exact trial definition, and the choice of electrodes (9, 11–14).

VOLATILITY AND ABERRANT SALIENCE IN SCHIZOPHRENIA

The stark misrepresentations of reality that are at the core of positive symptoms in psychosis suggest a dysfunction at higher levels of belief hierarchies, where general and stable features of the environment are represented (17, 21, 67). Aberrant estimation of volatility might be one domain-general mechanism that could lead to dysfunctional high-level beliefs about many entities in the world. Specifically, a tendency to overestimating the likelihood of meaningful changes in the environment would lead to inferences being less constrained by prior knowledge in general and to chronically enhanced PE signaling (30), potentially necessitating a compensatory formation of very precise beliefs at higher hierarchical levels. This putative

imbalance between the precision of priors and PEs has been suggested as a core mechanism for the formation of psychosis, where a misattribution of relevance (precision) to normally inconspicuous events may lead to a perceived strangeness of the world and a strong demand for an explanation that requires new and precise high-level beliefs (17, 21, 41). This hypothetical shift in precision-weighting, with relatively enhanced precision of sensory PEs due to weak priors, represents a Bayesian perspective on the concept of “aberrant salience” (67). Interestingly, recent evidence supports overestimation of environmental volatility in schizophrenia (68, 69).

LIMITATIONS

The main limitation of our approach is that we are not able to infer upon subject-specific learning styles by fitting the HGF to subjects’ behavior. A key feature of the HGF is the fact that the model parameters allow for the expression of individual differences in learning (where these parameters are potentially linked to specific neuromodulatory mechanisms (30, 70)). Similar to (38), we were not able to exploit this feature here because of the nature of the classical auditory mismatch paradigm, where no behavioral responses to the tones are emitted. Instead, we used the parameters of a surprise-minimizing Bayesian observer for each of the tone sequences and simulated belief trajectories accordingly. An important extension of future applications of the HGF 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.

CONCLUSION AND OUTLOOK

In this study, we present first evidence for the role of hierarchically related PEs in the auditory MMN. 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 environment’s statistical structure. Additionally, we found a reduced expression of the higher-level PE under infusion of S-ketamine, indicating a disturbance of high-level inference about environmental volatility by perturbation of NMDA receptors.

Our results are clinically important 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 this will contribute to the development of computational assays for improved differential diagnosis and treatment prediction in schizophrenia (71–73).

ACKNOWLEDGEMENTS

This study was supported by the University of Zurich (KES), the René and Susanne Braginsky Foundation (KES), the SNF Ambizione PZ00P3_167952 (AOD), the Swiss Neuromatrix (MK, FXV), and the Hefter Research Institute (FXV).

DISCLOSURES

The authors report no biomedical financial interest or potential conflicts of interest.

References

1. 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.

2. 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.

3. Winkler I (2007): Interpreting the Mismatch Negativity. *J Psychophysiol*. 21: 147–163.

4. 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.

5. Todd J, Harms L, Schal l U, Michie PT (2013): Mismatch negativity: Translating the potential. *Front Psychiatry*. 4: 1–22.

6. Umbricht D, Krljes S (2005): Mismatch negativity in schizophrenia: a meta-analysis. *Schizophr Res*. 76: 1–23.

7. 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.

8. 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: 11962–7.

9. 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: 1139–47.

10. 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.

11. Heekeren K, Daumann J, Neukirch A, Stock C, Kawohl W, Norra C, *et al.* (2008): Mismatch negativity generation in the human 5HT2A agonist and NMDA antagonist model of psychosis. *Psychopharmacology (Berl)*. 199: 77–88.

12. 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.

13. 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.

14. 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.

15. Stephan KE, Friston KJ, Frith CD (2009): Dysconnection in Schizophrenia: From abnormal synaptic plasticity to failures of self-monitoring. *Schizophr Bull*. 35: 509–527.

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

17. Sterzer P, Adams RA, Fletcher P, Frith C, Lawrie SM, Muckli L, *et al.* (2018): The Predictive Coding Account of Psychosis. *Biol Psychiatry*. 84.

18. Corlett PR, Honey GD, Fletcher PC (2016): Prediction error, ketamine and psychosis: An updated model. *J Psychopharmacol*. 30: 1145–1155.

19. Corlett PR, Honey GD, Krystal JH, Fletcher PC (2011): Glutamatergic model psychoses: Prediction error, learning, and inference. *Neuropsychopharmacology*. 36: 294–315.

20. Stephan KE, Baldeweg T, Friston KJ (2006): Synaptic Plasticity and Dysconnection in Schizophrenia. *Biol Psychiatry*. 59: 929–939.

21. Adams RA, Stephan KE, Brown HR, Frith CD, Friston KJ (2013): The computational anatomy of psychosis. *Front Psychiatry*. 4: 1–26.

22. Dayan P, Hinton GE, Neal RM, Zemel RS (1995): The Helmholtz Machine. *Neural Comput*. 7: 889–904.

23. Doya K, Ishii S, Pouget A, Rao RPN (2011): *Bayesian brain: probabilistic approaches to neural coding*. Cambridge, Mass.: MIT Press.

24. Friston KJ (2010): The free-energy principle: a unified brain theory? *Nat Rev Neurosci*. 11: 127–38.

25. 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.

26. Garrido MI, Kilner JM, Stephan KE, Friston KJ (2009): The mismatch negativity: A review of underlying mechanisms. *Clin Neurophysiol*. 120: 453–463.

27. Lieder F, Stephan KE, Daunizeau J, Garrido MI, Friston KJ (2013): A Neurocomputational Model of the Mismatch Negativity. *PLoS Comput Biol*. 9: e1003288.

28. Self MW, Kooijmansa RN, Supèr H, Lammec VA, Roelfsema PR (2012): Different glutamate receptors convey feedforward and recurrent processing in macaque V1. *Proc Natl Acad Sci*. 109: 11031–11036.

29. Friston KJ (2009): The free-energy principle: a rough guide to the brain? *Trends Cogn Sci*. 13: 293–301.

30. Mathys CD, Daunizeau J, Friston KJ, Stephan KE (2011): A Bayesian foundation for individual learning under uncertainty. *Front Hum Neurosci*. 5: 1–20.

31. Iglesias S, Mathys CD, Brodersen KH, Kasper L, Piccirelli M, DenOuden HEM, Stephan KE (2013): Hierarchical Prediction Errors in Midbrain and Basal Forebrain during Sensory Learning. *Neuron*. 80: 519–530.

32. Diaconescu AO, Mathys CD, Weber LAE, Daunizeau J, Kasper L, Lomakina EI, *et al.* (2014): Inferring on the Intentions of Others by Hierarchical Bayesian Learning. *PLoS Comput Biol*. 10.

33. Kolossa A, Kopp B, Fingscheidt T (2015): A computational analysis of the neural bases of Bayesian inference. *Neuroimage*. 106: 222–237.

34. 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.

35. 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.

36. Lieder F, Daunizeau J, Garrido MI, Friston KJ, Stephan KE (2013): Modelling Trial-by-Trial Changes in the Mismatch Negativity. *PLoS Comput Biol*. 9.

37. 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.

38. 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.

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: 1–24.

40. Corlett PR, Honey GD, Fletcher PC (2007): From prediction error to psychosis: ketamine as a pharmacological model of delusions. *J Psychopharmacol*. 21: 238–252.

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

42. Corlett PR (2017): I Predict, Therefore I Am: Perturbed Predictive Coding Under Ketamine and in Schizophrenia. *Biol Psychiatry*. 81: 465–466.

43. Cowan N, Winkler I, Teder W, Naatanen R (1993): Memory prerequisites of mismatch negativity in the auditory event-related potential (ERP). *J Exp Psychol Learn Mem Cogn*. 19: 909–21.

44. Baldeweg T, Klugman A, Gruzelier J, Hirsch SR (2004): Mismatch negativity potentials and cognitive impairment in schizophrenia. *Schizophr Res*. 69: 203–17.

45. Weilnhammer VA, Stuke H, Sterzer P, Schmack K (2018): The Neural Correlates of Hierarchical Predictions for Perceptual Decisions. *J Neurosci*. 38: 5008–5021.

46. 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.

47. Vossel S, Mathys CD, Daunizeau J, Bauer M, Driver J, Friston KJ, Stephan KE (2014): Spatial attention, precision, and bayesian inference: A study of saccadic response speed. *Cereb Cortex*. 24: 1436–1450.

48. Auksztulewicz R, Friston KJ, Nobre AC (2017): Task relevance modulates the behavioural and neural effects of sensory predictions. *PLOS Biol*. 15: e2003143.

49. 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.

50. Wacongne C, Changeux J-P, Dehaene S (2012): A Neuronal Model of Predictive Coding Accounting for the Mismatch Negativity. *J Neurosci*. 32: 3665–3678.

51. Näätänen R, Tervaniemi M, Sussman E, Paavilainen P, Winkler I (2001): ‘Primitive intelligence’ in the auditory cortex. *Trends Neurosci*. 24: 283–288.

52. Kilner JM, Friston KJ (2010): Topological inference for EEG and MEG. *Ann Appl Stat*. 4.

53. Penny WD, Holmes AJ (2007): Chapter 12 : Random effects analysis. In: Penny WD, Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE, editors. *Stat Parametr Mapp Anal Funct Brain Images*. London: Academic Press, pp 156–165.

54. Flandin G, Friston KJ (2017): Analysis of family-wise error rates in statistical parametric mapping using random field theory. *Hum Brain Mapp*. . doi: 10.1002/hbm.23839.

55. Polich J (2007): Updating P300: An integrative theory of P3a and P3b. *Clin Neurophysiol*. 118: 2128–2148.

56. 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.

57. Dzafic I, Randeniya R, Garrido MI (2018): Reduced top-down connectivity as an underlying mechanism for psychotic experiences in healthy people. *bioRxiv*. 296988.

58. 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.

59. 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.

60. 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.

61. 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.

62. Rosburg T, Schmidt A (2018): Potential Mechanisms for the Ketamine-Induced Reduction of P3b Amplitudes. *Front Behav Neurosci*. 12: 308.

63. Rosch RE, Auksztulewicz R, Leung PD, Friston KJ, Baldeweg T (2018, August 13): Selective Prefrontal Disinhibition in a Roving Auditory Oddball Paradigm Under N-Methyl-D-Aspartate Receptor Blockade. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 0. doi: 10.1016/j.bpsc.2018.07.003.

64. Horváth J, Winkler I, Bendixen A (2008): Do N1/MMN, P3a, and RON form a strongly coupled chain reflecting the three stages of auditory distraction? *Biol Psychol*. 79: 139–147.

65. Bekinschtein TA, Dehaene S, Rohaut B, Tadel F, Cohen L, Naccache L (2009): Neural signature of the conscious processing of auditory regularities. *Proc Natl Acad Sci U S A*. 106: 1672–7.

66. Vinckier F, Gaillard R, Palminteri S, Rigoux L, Salvador A, Fornito A, *et al.* (2016): Confidence and psychosis: a neuro-computational account of contingency learning disruption by NMDA blockade. *Mol Psychiatry*. 21: 946–955.

67. 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.

68. Deserno L, Boehme R, Mathys CD, Katthagen T, Kaminski J, Stephan KE, *et al.* (2017): Overestimating environmental volatility increases switching behavior and is linked to activation of dorsolateral prefrontal cortex in schizophrenia. *bioRxiv*. 227967.

69. Kaplan CM, Saha D, Molina JL, Hockeimer WD, Postell EM, Apud JA, *et al.* (2016): Estimating changing contexts in schizophrenia. *Brain*. 139: 2082–2095.

70. Vossel S, Bauer M, Mathys CD, Adams RA, Dolan RJ, Stephan KE, Friston KJ (2014): Cholinergic Stimulation Enhances Bayesian Belief Updating in the Deployment of Spatial Attention. *J Neurosci*. 34: 15735–15742.

71. Adams RA, Huys QJM, Roiser JP (2016): Computational Psychiatry: towards a

mathematically informed understanding of mental illness. *J Neurol Neurosurg Psychiatry*. 87: 53–63.

72. Mathys CD (2016): How could we get nosology from computation? *Comput Psychiatry New Perspect Ment Illn*. MIT Press, pp 121–135.

73. Stephan KE, Iglesias S, Heinzle J, Diaconescu AO (2015): Translational Perspectives for Computational Neuroimaging. *Neuron*. 87: 716–732.

Figures

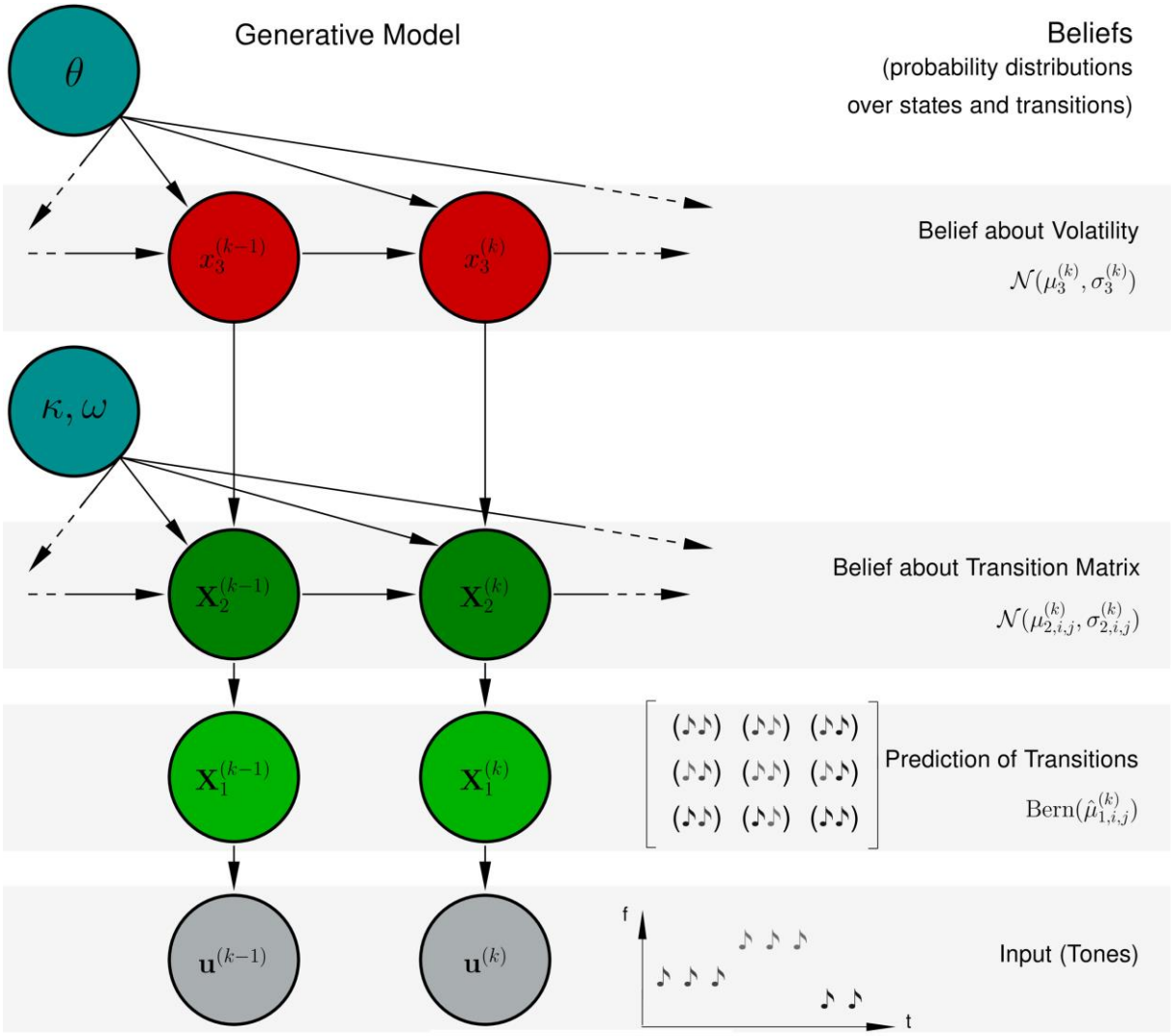

Figure 1. The Perceptual Model: A multivariate version of the binary three-level Hierarchical Gaussian Filter (HGF). The agent infers upon two continuous quantities: The transition tendencies from one tone (frequency) to another, stored in the transition matrix $\mathbf{X}_2$, and the (common) volatility of these tendencies, x_3. To employ 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 (updates described in the main text details described in the supplementary material).

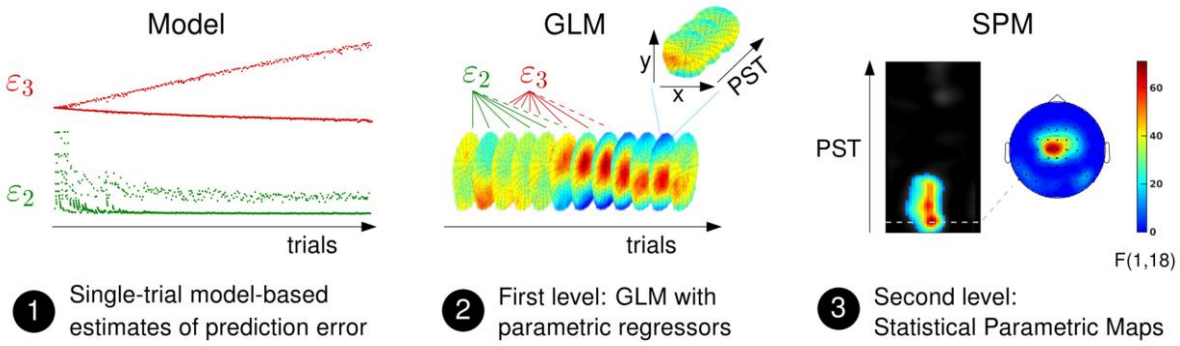

Figure 2. Sketch of the analysis pipeline for the model-based analysis. First, we simulated an agent's beliefs using our hierarchical Bayesian model, which provided us with an estimate of precision-weighted PE on two hierarchical levels (ε_2 and ε_3) for each trial in each session. At first glance, it may seem that the $\varepsilon_3$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 (see Figure S1C). Second, we used these estimates as parametric regressors in a GLM of the single-trial EEG signal in each session of each subject separately (peri-stimulus time window of 100 - 400 ms after tone onset) and computed the first level statistics. Third, the beta values for each quantity and each subject 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 maps across within-trial time and sensor space. PST = peri-stimulus time.

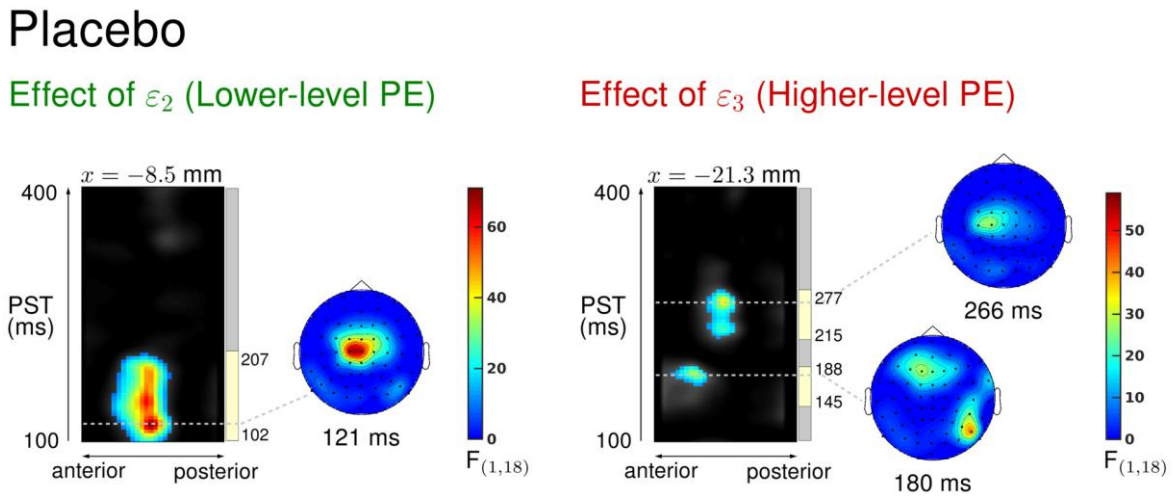

Figure 3. Results of the model-based EEG analysis in the placebo condition: effects of the high- and the low-level PE. The left side always shows an F-map across the scalp dimension *y* (from posterior to anterior, x-axis), and across peristimulus time (y-axis), at the spatial *x*-location indicated 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 color-coded according to the legend plotted on the right. 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 fronto-central and temporal channels. For the lower-level PE, ε_2, the correlation peaked at 121 ms post-stimulus at central channels, for the higher-level PE, ε_3, it peaked at 180 ms at frontal channels, at 184 ms at temporal channels (not shown here), and at 266 ms post-stimulus at left central channels.

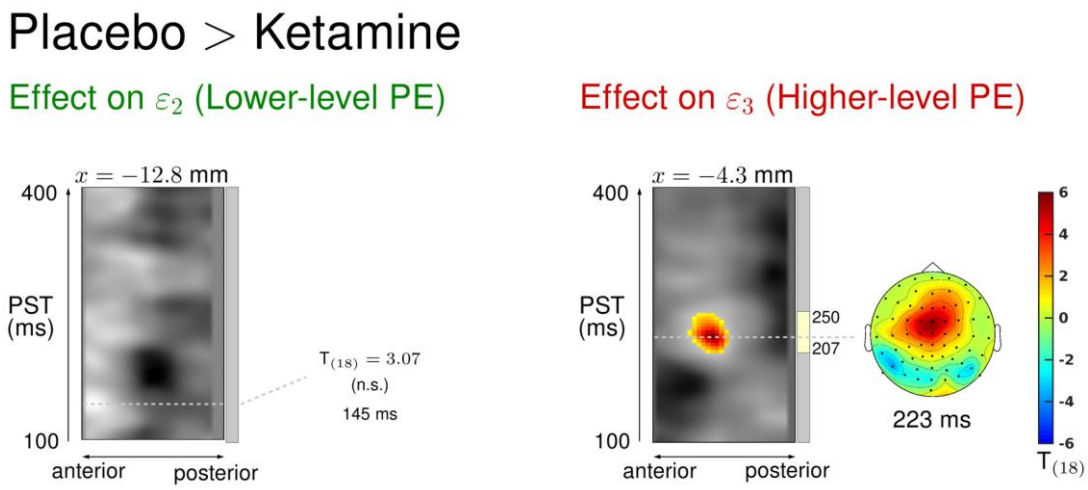

Figure 4. Drug effect on the representation of the lower-level PE (ε_2, left) and the higher-level PE (ε_3, right). The left side always shows a T-map for the paired T-test (Placebo – Ketamine) across the scalp dimension *y* (from posterior to anterior, x-axis), and across peristimulus time (y-axis), 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 color-coded according to the legend plotted on the right. Time-windows of significant correlations are indicated by the yellow bars next to the colored clusters of significant F 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 to ketamine between 207 ms and 250 ms post-stimulus, peaking at 223 ms at fronto-central channels. No significant drug effects were found for the lower-level PE.

Tables

Placebo	Peak Coordinates			peak-level			cluster-level	
	x [mm]	y [mm]	t [ms]	$F_{\{1,18\}}$	$Z_{\equiv}$	p_{FWE}	k_{E}	p_{FWE}
Higher-level PE (ε_3)	60	−46	184	58.79	4.91	0.003	681	0.004
	64	−36	184	56.94	4.87	0.004		
	−21	−9	266	35.13	4.2	0.045	852	0.002
	−8	−3	234	33.32	4.13	0.058		
	−21	24	180	31.64	4.06	0.073	541	0.009
	−17	67	152	17.58	3.27	0.533		
Lower-level PE (ε_2)	−8	−9	121	70.07	5.15	0.001	3746	0.000
	−4	−3	152	63.46	5.02	0.002		
	17	2	164	60.35	4.95	0.003		

Table 1. Significant representations of the two computational quantities under placebo. The table lists the peak coordinates, F values, corresponding Z values, whole-volume FWE-corrected p-values at the voxel level, cluster size (k_E) and FWE-corrected p-values at the cluster level of voxels showing significant correlations of EEG signal with the trajectory of one of the precision-weighted PEs ε_3 and ε_2 (F-test, $p < 0.05$ whole-volume FWE-corrected at the cluster-level with a cluster-defining threshold of $p < 0.001$).

Ketamine	Peak Coordinates			peak-level			cluster-level	
	x [mm]	y [mm]	t [ms]	$F_{\{1,18\}}$	$Z_{\equiv}$	p_{FWE}	k_{E}	p_{FWE}
Higher-level PE (ε_3)	−13	8	398	35.53	4.22	0.04	1205	0.001
	−8	8	363	26.04	3.79	0.151		
	−21	−3	340	21.71	3.55	0.285		
	60	−73	160	35.00	4.2	0.043	340	0.04
	60	−73	188	33.6	4.14	0.052		
	68	−19	172	29.34	3.96	0.094		
	60	−73	375	32.34	4.09	0.019	449	0.021
	60	−73	332	31.57	4.06	0.068		
	60	−73	348	30.23	4.00	0.082		
	13	40	207	25.23	3.75	0.17	625	0.008
	−8	2	180	22.05	3.57	0.271		
	17	29	164	18.99	3.37	0.42		
Lower-level PE (ε_2)	−60	−46	141	76.42	5.27	0.001	1819	0.000
	−60	−62	129	73.11	5.21	0.001		
	−8	−89	137	40.38	4.4	0.029		
	17	2	137	63.41	5.02	0.003	3117	0.000
	−17	−3	125	39.03	4.35	0.034		
	30	−9	172	37.65	4.3	0.041		

Table 2. Significant representations of the two computational quantities under ketamine. The table lists the peak coordinates, F values, corresponding Z values, whole-volume FWE-corrected p-values at the voxel level, cluster size (k_E) and FWE-corrected p-values at the cluster level of voxels showing significant correlations of EEG signal with the trajectory of one of the precision-weighted PEs ε_3 and ε_2 (F-test, $p < 0.05$ whole-volume FWE-corrected at the cluster-level with a cluster-defining threshold of $p < 0.001$).

Higher-level PE (ε_3)	Peak Coordinates			peak-level			cluster-level	
	x [mm]	y [mm]	t [ms]	$T_{\{18\}}$	$Z_{\equiv}$	p_{FWE}	k_{E}	p_{FWE}
Placebo>Ketamine	−4	2	223	5.95	4.37	0.025	685	0.005

Table 3. Significant drug differences for the representation of the higher-level PE ε_3 in Placebo vs. Ketamine condition. The table lists the peak coordinates, t values, corresponding Z values, FWE-corrected p-values at the voxel level, cluster size (k_E) and whole-volume FWE-corrected p-values at the cluster level of voxels showing significantly stronger representation of ε_3 under placebo compared to the ketamine condition (paired t-test, $p < 0.05$ whole-volume FWE-corrected at the cluster-level with a cluster-defining threshold of $p < 0.001$). No significant drug differences were found for the lower-level PE.