Check for updates

# Localizing hierarchical prediction errors and precisions during an oddball task with volatility: Computational insights and relationship with psychosocial functioning in healthy individuals

Colleen E. Charlton[a,*], Daniel J. Hauke[b,*], Michelle Wobmann[c], Renate de Bock[c], Christina Andreou[d], Stefan Borgwardt[d], Volker Roth[e], Andreea O. Diaconescu[a,f,g,h]

[a]Krembil Centre for Neuroinformatics, Centre for Addiction and Mental Health (CAMH), Toronto, ON, Canada

[b]Centre for Medical Image Computing, Department of Computer Science, University College London, London, United Kingdom

[c]Department of Psychiatry (UPK), University of Basel, Basel, Switzerland

[d]Department of Psychiatry and Psychotherapy, Translational Psychiatry, University of Lübeck, Lübeck, Germany

[e]Department of Mathematics and Computer Science, University of Basel, Basel, Switzerland

[f]Institute of Medical Sciences, University of Toronto, Toronto, ON, Canada

[g]Department of Psychiatry, University of Toronto, Toronto, ON, Canada

[h]Department of Psychology, University of Toronto, Toronto, ON, Canada

*Joint first author

Corresponding Author: Colleen Charlton (colleen.charlton@camh.ca; ccharlton@hotmail.ca)

## ABSTRACT

The auditory mismatch negativity (MMN) has been widely used to investigate deficits in early auditory information processing, particularly in psychosis. Predictive coding theories suggest that impairments in sensory learning may arise from disturbances in hierarchical message passing, likely due to aberrant precision-weighting of prediction errors (PEs). This study employed a modified auditory oddball paradigm with varying phases of stability and volatility to disentangle the impact of hierarchical PEs on auditory MMN generation in 43 healthy controls (HCs). Single-trial EEG data were modeled with a hierarchical Bayesian model of learning to identify neural correlates of low-level PEs about tones and high-level PEs about environmental volatility. Our analysis revealed a reduced expression of the auditory MMN in volatile compared to stable phases of the paradigm. Additionally, lower Global Functioning (GF): Social scores were associated with a reduced difference waveform at 332 ms after stimulus presentation across the entire MMN paradigm. Further analysis revealed that this association was present during the volatile phase but not the stable phase of the paradigm. Source reconstruction suggested that the association between the stable difference waveform and psychosocial functioning originated in the left superior temporal gyrus. Finally, we found significant EEG signatures of both low- and high-level PEs and precision ratios. Our findings highlight the value of computational models in understanding the neural mechanisms involved in early auditory information processing and their connection to psychosocial functioning.

**Keywords:** sensory learning, hierarchical Gaussian filter, predictive coding, EEG, functioning

## 1. INTRODUCTION

Deficits in early auditory information processing, as measured by event-related potentials (ERP), provide critical insights into the pathophysiology of psychosis and its connection to psychosocial functioning (Ro & Clark, 2009; Thomas et al., 2017). The mismatch negativity (MMN), an ERP component reflecting the brain's capacity to automatically detect and encode novel or unexpected auditory events, has been especially useful in studying automatic sensory processing and its relationship to

psychosocial functioning (Friston, 2005; Garrido, Kilner, Stephan, et al., 2009; Näätänen et al., 2007). Notably, reductions in the MMN, along with the subsequent P3a component, have been linked to reduced psychosocial functioning in healthy adults (Light et al., 2007) and across the psychosis spectrum (Carrión et al., 2015; Hermens et al., 2010; Murphy et al., 2020).

Recent advances in computational neuroscience, particularly within the predictive coding framework, have reframed our understanding of the MMN as a reflection of prediction error (PE) processing in the brain (Friston, 2003, 2005; Garrido, Kilner, Stephan, et al., 2009; Lieder, Daunizeau, et al., 2013; Rao & Ballard, 1999). The MMN is thought to reflect the brain's inability to suppress PEs in response to unpredictable sensory input. This framework emphasizes the hierarchical nature of information processing, where higher-level PEs encompass abstract, slowly changing information, while lower-level PEs focus on rapidly changing basic sensory features (Kiebel et al., 2008; Wacongne et al., 2011).

The hierarchical Gaussian filter (HGF), a hierarchical Bayesian model of learning, illustrates this concept by modeling the impact of PEs through precision-weighted updates (Mathys et al., 2011, 2014). In this framework, the precision ratio serves as a learning rate that influences the balance between prior beliefs and new sensory input. Modulation of the precision ratio is thought to involve neuromodulatory gain control mechanisms, possibly involving N-methyl-D-aspartate (NMDA) and acetylcholine receptors, which encode precision as changes in neuronal excitability (Kanai et al., 2015; Schöbi et al., 2021; Weber et al., 2022). In stable settings, high precision is assigned to our prior beliefs, minimizing the impact of sensory fluctuations. In volatile environments, the precision ratio (i.e., gain control) shifts to prioritize lower-level sensory inputs, ensuring effective adaptation to environmental changes.

While much research has focused on the detection of the MMN at the sensor level, this approach provides limited insight into the specific cortical regions involved in generating and modulating PEs. Based on PE accounts of the MMN, we predict that low-level PEs are generated in early sensory regions (Parras et al., 2017), whereas high-level PEs are computed in higher-level regions, including the inferior frontal gyrus (IFG) (Camalier et al., 2019). Localizing the neural sources of PEs is crucial for identifying how different cortical areas contribute to predictive processing and interact during auditory learning, enabling non-invasive analysis of brain dynamics (Asadzadeh et al., 2020).

Prior research has demonstrated the sensitivity of the MMN to environmental volatility, showing larger MMN amplitudes in stable conditions, where deviant tones are less expected, and smaller amplitudes in volatile settings where deviant occurrences are more frequent (Dzafic et al., 2020; Kotchoubey, 2014; Todd et al., 2011, 2014; Weber et al., 2022). Several of these studies used the local-global paradigm to show how changes in MMN amplitude are linked to variations in the regularity of auditory stimuli. This paradigm distinguishes between local (tone-to-tone transitions) and global (sequence-to-sequence transitions) regularities, suggesting that local-rule violations are processed in lower-order cortical areas, whereas global-rule violations are detected by higher-order areas (Fong et al., 2020).

Building on these findings, we employ a modified auditory oddball paradigm (Weber et al., 2022) with alternating stable and volatile periods to better distinguish learning about tone probabilities and environmental volatility. The manipulation of volatility is crucial as it allows us to explore how the brain dynamically adjusts its predictive processes in response to environmental changes. In classic oddball paradigms, low-level and high-level PEs are often correlated (Charlton et al., 2022). Our paradigm is deliberately designed to decorrelate low-level and high-level PEs to achieve a clearer separation between these processes. Additionally, manipulating volatility broadens the range of environmental dynamics participants must learn about, extending beyond the fixed deviant probabilities seen in the MMN paradigm. This approach allows us to examine how the brain flexibly adjusts its predictive mechanisms across a wider spectrum of true volatility, offering new insights into dynamic learning processes.

Unlike the local-global paradigm, which primarily examines "what" is changing (i.e., local or global deviance), our study uses the HGF, which examines "how fast" these auditory changes occur at different hierarchical levels. The HGF assumes a specific relationship between hierarchically-coupled states that evolve in time as Gaussian random walks, linked through their variances (Hauke, 2022). This approach allows for a more nuanced and dynamic perspective of change, beyond the static learning process often assumed in classical MMN analyses.

In our exploratory study, we employ a computational approach using the HGF to model single-trial electroencephalogram (EEG) data in a novel MMN paradigm. This approach diverges from traditional MMN analyses, which typically assess amplitude changes from a limited selection of trials. By modeling brain activity on a trial-by-trial basis with the HGF, we can capture intermediate learning signals for a more accurate depiction of the brain's continuous and dynamic learning processes. We investigate the individual components of precision-weighted PEs—precision ratios and unweighted PEs—and examine their

associations with psychosocial functioning in the general population. Furthermore, we use multiple sparse prior (MSP) source reconstruction (Friston et al., 2008) to localize the cortical generators of PEs and their precision weights, offering insight into the interactions between brain regions during predictive coding. By integrating the HGF model with MSP source reconstruction, our exploratory study aims to provide a more comprehensive understanding of MMN generation and its relationship with psychosocial functioning.

## 2. METHODS

### 2.1. Participants

A total of 43 healthy controls (HC) were recruited through online and public advertisements in Basel, Switzerland. HCs were recruited as part of a larger study investigating persecutory ideation in early psychosis, described in greater detail elsewhere (Hauke et al., 2024). Due to insufficient patient data, only results from HCs are presented here. Inclusion and exclusion criteria are detailed in the Supplementary Material. All participants provided written informed consent, and the study was conducted in accordance with the Declaration of Helsinki and approved by the local ethics committee (Ethikkommission Nordwest- und Zentralschweiz, no. 2017-01149).

### 2.2. Demographic variables and functioning assessment

Demographic and functioning data were collected during an interview within 5 days of EEG data collection. Functioning was assessed with the Global Functioning: Social (GF: Social) and the Global Functioning: Role (GF: Role) scales (Cornblatt et al., 2007). The former measures social relationships and interpersonal activities, and the latter assesses performance and functioning in work, education, and home activities depending on the age of the individual. A single score for each scale was provided by trained raters, with higher scores indicating better performance. The median score for GF: Social was 9 and for GF: Role was 9, with both scores ranging from 6 to 10. Demographic characteristics are summarized in Table 1.

### 2.3. Mismatch negativity paradigm

EEG data were collected during the mismatch negativity paradigm specifically designed to minimize the correlation between low-level and high-level precision-weighted prediction errors (for details, see Weber et al., 2022). As such, the task included periods of volatility, characterized by rapid changes in the probabilities of the tones, and

**Table 1.** Participant characteristics.

|  | Healthy controls |
| --- | --- |
| Age (mean [SD]) | 22.9 [6.8] |
| Years of education (mean [SD]) | 13.2 [3.3] |
| Working memory (mean [SD]) | 6.6 [2.1] |
| Sex (f/m) | 20/23 |
| Handedness (l/r) | 5/38 |
| Cannabis (y/n) | 27/16 |

Demographic characteristics of the study sample.

periods of stochasticity, where the tone probability averaged at 50%. Participants were presented with a series of tones delivered binaurally through Etymotics HF5 headphones, while they engaged in a visual distraction task to direct participants attention away from the tones (see Supplementary Material for more details). The auditory stimuli comprised two pure sinusoidal tones: a high (528 Hz) and a low (440 Hz) tone, each lasting 70 ms (including 5 ms fade-in/fade-out) with an inter-stimulus interval of 500 ms and totaling 1800 tones. Stimuli were presented at comfortable loudness determined by the participants. The task consisted of two types of phases: stable phases were defined as periods where the probability of hearing the same tone remained constant for at least 90 trials and volatile phases were all other phases (Fig. 1A). Auditory and visual stimuli were presented using PsychToolbox (PTB3, psychotoolbox.org; version 3.0.14) and Matlab (R2018a).

Participants' responses during the visual distraction task were analyzed by computing mean reaction times and hit rates, which are reported in the Supplementary Material. One subject had a hit-rate below 75%, which may suggest that the individual was attending the tones instead of the visual task. To address this, we reanalyzed the data excluding this subject, and found that all subsequent results still held, unless otherwise specified.

### 2.4. EEG recording and preprocessing

EEG data were collected using a 64-electrode cap (BioSemi MP150 System) with active electrodes and additional reference and ground electrodes. Electrooculograms were recorded via electrodes placed on the supraorbital and infraorbital ridges of the left eye and on the outer canthi of both eyes. Signals were digitised at 1024 Hz with a DC amplifier. EEG data were preprocessed and analyzed using SPM12 (http://www.fil.ion.ucl.ac.uk/spm/; version 7487) and Matlab (R2023a; version 9.14.0.2206163). Continuous EEG data were high-pass filtered (0.1 Hz), down-sampled to 256 Hz, and low-pass filtered (30 Hz) to ensure comparability with previous results (Weber et al., 2022). Data were epoched

**A.** Volatility Schedule

**B.** Hierarchical Bayesian Model
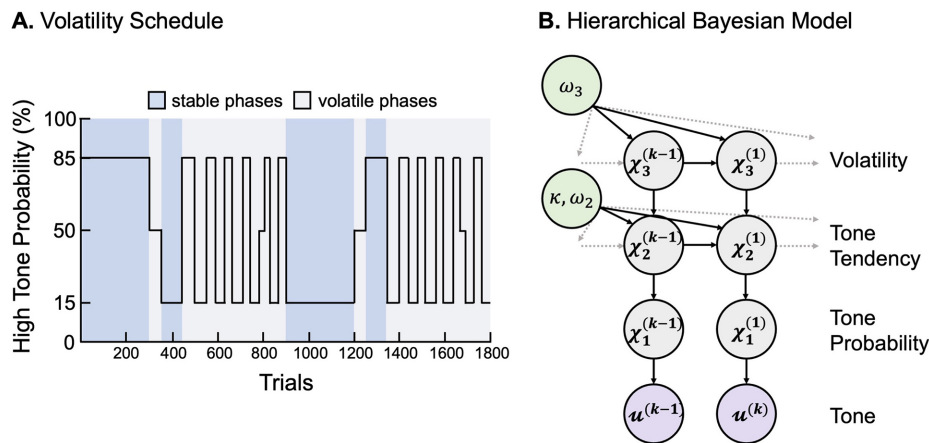


**Fig. 1.** Mismatch volatility schedule and computational model. (A) Volatility schedule of mismatch negativity paradigm. (B) Three-level hierarchical Gaussian filter (HGF) binary perceptual model.

into 600 ms segments around tone onsets (-100 to 500 ms), and baseline correction was performed using a -100 to 0 ms peristimulus window. Eye movement arti-facts were corrected using the signal space projection (SSP) eyeblink correction method in SPM12 (Nolte & Hämäläinen, 2001), which regresses out the leading component from the EEG data based on eye activity (see Supplementary Material for further details). Following eyeblink correction, trials with amplitudes ±75 µV were considered as artefactual and excluded. Channels with over 20% artefactual trials were interpolated for analy-ses. One participant had a single bad channel (P8). The median number of artifact-free trials, taken across all participants, was 1750 (25th percentile: 1623, 75th per-centile: 1783). This total includes all recorded trials, encompassing not only trials defined as standard and deviant tones according to our definition but also all other trials presented during the task. Notably, our para-metric model-based analysis incorporates data from all trials collectively, without differentiating between these conditions.

### 2.5. Difference waveform analysis

In-line with Weber et al., 2022, trial types were defined as follows: deviant trials followed at least 5 repetitions of the other tone (N = 119), and standard trials were defined as the 6th repetition of the same tone (N = 106). In the standard condition, the median number of artifact-free trials was 103 (25th percentile: 97, 75th percentile: 106), and in the deviant condition, it was 116 (25th percentile: 112, 75th percentile: 118). The difference waveform was obtained by subtracting the average standard ERP from the average deviant ERP. Additionally, separate stable and volatile difference waveforms were obtained by sub-tracting the standard ERP from the deviant ERP during

the respective phases. In the stable phase, there were 51 standard tones and 55 deviant tones, whereas in the volatile phase, there were 55 standard tones and 64 deviant tones.

For each subject, difference waveform ERPs were converted into scalp images for all 64 channels using a voxel size of 4.3 mm x 5.4 mm x 2.0 ms. In agreement with previous studies (0, 0, 0, 0), statistical analyses were limited to a 100–400 ms post-stimulus interval, to capture the MMN and *P*300 peaks and to reduce the amount of comparisons. Images were smoothed with a Gaussian kernel (FWHM: 16 x 16 mm) to meet the assumptions of Gaussian random field theory (Kiebel & Friston, 2004; Worsley et al., 1996).

Images from the first-level served as input to our second-level general linear models (GLMs). We tested for the effect of difference waveform expression across sen-sor space and peristimulus time. Significant effects were inferred from thresholded $F$-statistical parametric maps (SPMs) at peak ($p < 0.05$) and cluster ($p < 0.05$) level, which were family-wise error corrected using Gaussian random field theory with a cluster-defining threshold of $p < 0.001$ (Flandin & Friston, 2019).

### 2.6. Computational framework

Single-trial EEG data were modeled using the HGF, a hierarchical Bayesian model of learning under uncertainty (Mathys et al., 2011, 2014), which was previously used to model oddball paradigms (Charlton et al., 2022; Hauke et al., 2023; Weber et al., 2020) (Fig. 1B). The HGF was implemented using the *tapas_ehgf_binary* function of the HGF toolbox (version: 6.0), which is part of the open-source TAPAS software collection (version: 4.0.0) (https://github.com/translationalneuromodeling/tapas /releases/tag/v4.0.0) (Frässle et al., 2021).

Participants were exposed to a sequence of tones in which the high tone were either "deviant" or "standard" depending on their local frequency compared to low tones. We used the HGF to model an individual's implicit learning of the tone sequence. Since the oddball paradigm does not require behavioral responses, and therefore the model cannot be fit to behavior, participants were modeled as surprise-minimizing Bayesian observers (see Supplementary Material for more details). Perceptual parameters were optimized to reduce overall surprise in response to the experienced tone sequence (see Supplementary Tables S1 and S2 for a summary of the parameters).

According to the model, participants' beliefs are recursively updated using the following equation:

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

where $\Delta\mu_i^{(k)}$ denotes the change in the posterior belief at level $i$ on trial $k$. After each tone is experienced, the prediction error from the level below ($\delta_{i-1}^{(k)}$) is computed and then weighted by a ratio of precisions: the precision of the prediction about the level below (sensory precision) ($\hat{\pi}_{i-1}^{(k)}$) and the belief precision at the current level ($\pi_i^{(k)}$). The precision ratio functions as a dynamic learning rate, as it adjusts the magnitude of belief updates based on one's confidence in the sensory input relative to their prior belief. For example, if the environment is stable, the learning rate may be small, resulting in more modest belief updates. By contrast, when the environment is volatile, the learning rate can be increased, enabling greater belief updates and providing a mechanism for flexible adaptation to changing environments.

From the model, we extracted trial-wise estimates of precision-weighted prediction errors (pwPE) at two levels of the hierarchy: low-level sensory pwPEs about the tone tendency $\varepsilon_2^{(k)}$ and high-level volatility pwPEs that update the estimate of environmental volatility $\varepsilon_3^{(k)}$.

### 2.7. Sensor-level single-trial EEG analysis

EEG waveforms were converted to images and smoothed as outlined above. At the single subject level, a GLM with an intercept term and z-standardized computational trajectories was constructed to explain changes in EEG amplitude across trials. Given that precisions may be connected to neurotransmitter-NMDAR (N-methyl-D-aspartate receptor) interactions and PEs to AMPA ($\alpha$-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid) signaling (Friston et al., 2016; Sterzer et al., 2018), two GLMs were included to unpack the effects of pwPEs.

They consisted of the low-level unweighted PE $\delta_1^{(k)}$ and the corresponding precision ratio ($\psi_2$), as well as the high-level unweighted PE $\delta_2^{(k)}$ and corresponding precision ratio ($\psi_3$). The average collinearity between these regressors was $r = 0.088 \pm 0.005$ (mean $\pm$ SD) and $r = 0.119 \pm 0.004$, respectively. Low-level pwPEs $\varepsilon_2^{(k)}$ and high-level pwPEs $\varepsilon_3^{(k)}$ also served as multiple regressors in an additional first-level GLM ($r = 0.535 \pm 0.004$). Consistent with previous studies (Weber et al., 2022), these regressors were not orthogonalized with respect to each other.

The same model-trajectories were used for all participants; however, following EEG preprocessing, these trajectories varied slightly due to rejected trials following artifact rejection. For each computational quantity, we conducted an *F*-test at each point in 2D sensor space and each time point to test the null hypothesis that the correlation between the model-derived learning trajectory and EEG amplitudes was zero, generating beta images that were used for second-level analysis.

At the group level, we specify separate GLMs for each computational quantity, in which beta images from the first-level served as input to the second-level. We identified significant effects using the same threshold criteria as described in Section 2.5.

### 2.8. Source-level EEG analysis

Finally, we applied multiple sparse priors (MSP) source reconstruction to estimate the distribution of cortical sources that give rise to the sensor level single-trial EEG data (Friston et al., 2008). We defined a prior set of sources in line with previous MMN studies, indicating the involvement of specific cortical regions, including bilateral primary auditory cortices (A1), bilateral superior temporal gyri (STG), and bilateral inferior frontal gyri (IFG) (Garrido et al., 2008, Garrido, Kilner, Kiebel, et al., 2009; Giard et al., 1990; Molholm et al., 2005; Opitz et al., 2002) (see Supplementary Table S3 for source coordinates). Source reconstruction was applied to both MMN waveforms (MMN analysis) and single-trial data (computational analysis) to estimate source time courses. We further compared MSP with identically distributed (IID) source reconstruction, finding evidence in favor of MSP, and thus proceeded with the MSP analysis for all subsequent analyses (see Supplementary Material for details). For the computational analysis, we proceeded as in the sensor space analysis and defined three design matrices GLMs (low-level pwPE design: $\delta_1$ and $\psi_2$; high-level pwPE design: $\delta_2$ and $\psi_3$; pwPEs design: $\varepsilon_2$ and $\varepsilon_3$) with an intercept term and z-standardized computational trajectories to explain changes in absolute source amplitude across trials at the first level. First-level betas (reflecting the

correlation between absolute source amplitude and computational quantity) were converted to images and carried to the second-level to test in which of the six sources and when in time each computational variable was expressed. The same statistical thresholds as in the sensor analysis were used.

## 3. RESULTS

### 3.1. Effect of stability on mismatch negativity

Our sensor-level EEG analysis examined the influence of volatility on the mismatch response by comparing the difference waveform (deviant–standard) between stable and volatile periods (Fig. 2). For the stable condition, *F*-tests revealed a significant peak at 141 ms in frontal electrodes (peak, $F_{(1,42)} = 113.94$; $p < 0.001$), aligning with the timing of the MMN effect. In the volatile condition, a significant peak occurred at 145 ms in frontal electrodes (peak, $F_{(1,42)} = 25.70$; $p = 0.034$) (Supplementary Fig. S1).

Comparing stable versus volatile phases revealed significant differences in the waveform amplitude. The difference waveform was significantly larger in stable phases peaking at 152 ms in central electrodes (peak, $t_{(1,42)} = 5.9$; $p = 0.002$; Fig. 3A) and 145 ms in parietal-occipital electrodes (peak, $t_{(1,42)} = 5.1$; $p = 0.019$; Fig. 3B). The difference wave showed greater negativity during stable phases in central electrodes, while parietal clusters exhibited the opposite pattern due to polarity changes. The left-lateralized effect observed in central

electrodes (Fig. 3A) likely results from a relatively weaker MMN component on the left side in the volatile condition, resulting in a more pronounced stable-volatile difference in the waveforms.

### 3.2. Relationship between mismatch stability and functional impairment

Next, we examined the relationship between the sensor-level difference waveform across both phases (stable and volatile) and functioning at baseline. Significant positive correlations were found between the difference waveform and GF: Social, peaking at 332 ms in central-frontal electrodes (cluster, $t_{(1,41)} = 4.5$; $p = 0.003$; Fig. 4A), and GF: Role, peaking at 387 ms in temporal electrodes (peak, $t_{(1,41)} = 4.7$; $p = 0.049$; Fig. 4B).

We investigated the relationship between functional impairment and stability, revealing a significant positive effect between the volatile difference waveform and GF: Social, peaking at 344 ms in central-frontal electrodes (peak, $t_{(1,41)} = 4.8$; $p = 0.036$; Fig. 4C). We also observed a significant positive correlation between the stable difference waveform and GF: Role, peaking at 387 ms in temporal electrodes (cluster, $t_{(1,41)} = 4.3$; $p = 0.049$; Fig. 4D). However, after excluding one subject with poor behavioral performance (based on a hit-rate below 75% in the visual distraction task), the effect with GF: Role fell just short of significance, possibly due to a loss of statistical power (cluster, $t_{(1,40)} = 4.3$; $p = 0.054$).
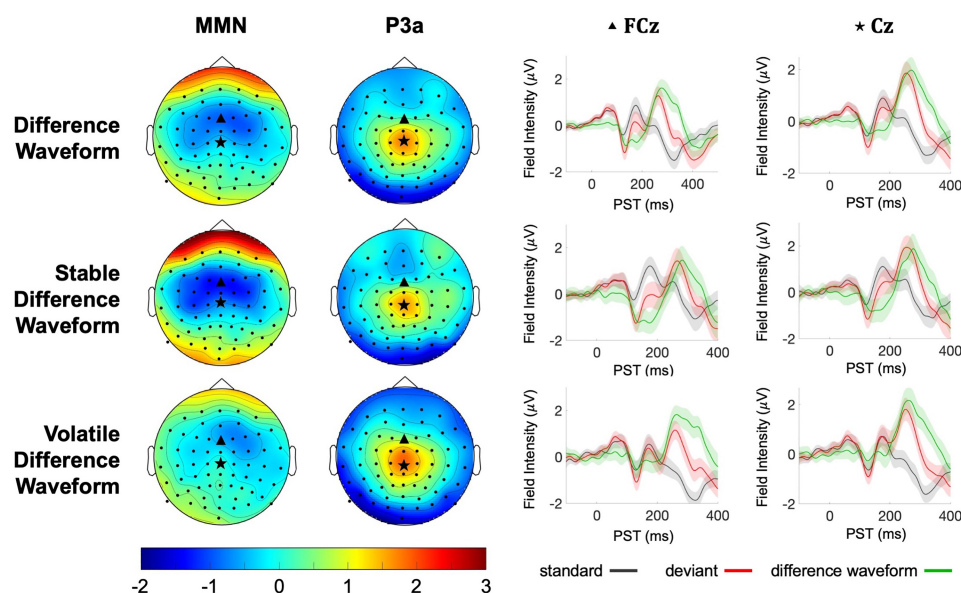


**Fig. 2.** Event-Related Potential (ERP) Scalp Topographies: On the left, scalp voltage topography maps displaying group mean grand-average ERPs. The MMN component was averaged from 150 to 200 ms post-stimulus, and the P3a component was averaged from 250 to 300 ms post-stimulus. All maps are plotted on the same voltage scale (μV). On the right, event-related potential difference waveforms at the FCz and Cz electrodes are shown for the overall difference waveform for the entire paradigm, along with the stable and volatile difference waveforms.
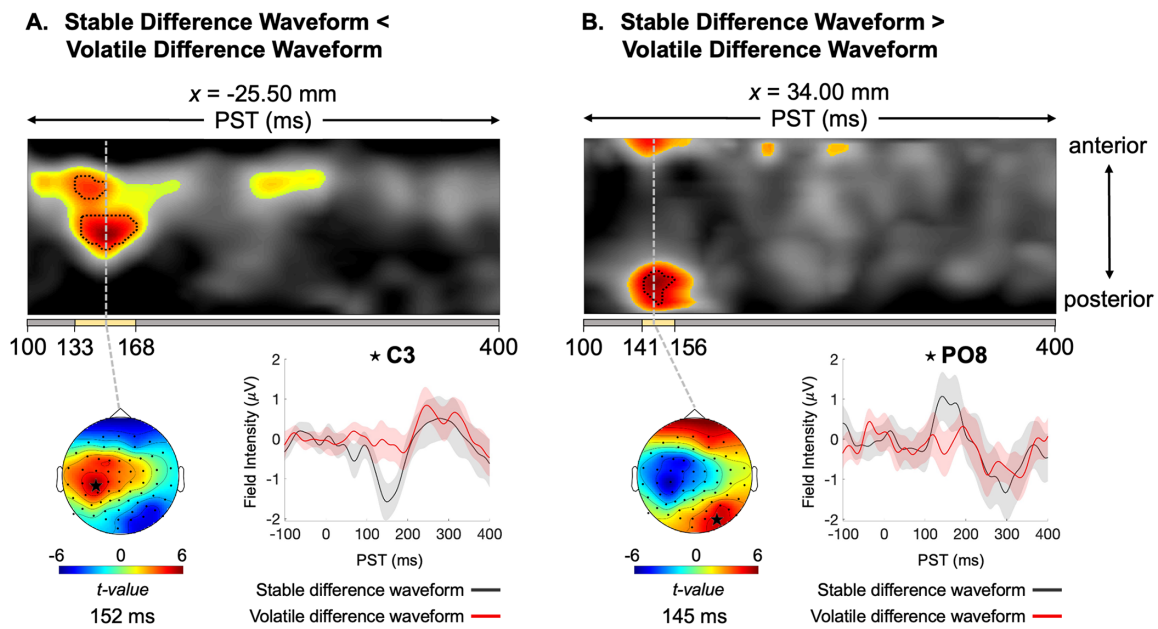
**Fig. 3.** Mismatch responses in volatile phases are reduced compared to stable phases. (A) Maximum intensity projection *t*-map illustrating the contrast between stable MMN < volatile MMN ERPs across anterior to posterior scalp locations (top). Significant peak-level effects ($p < 0.05$, whole-volume FWE-corrected) are outlined by black contours, while the coloured area indicates *t*-values exceeding the cluster-defining threshold of $p < 0.001$, uncorrected. The yellow bar at the bottom of the *t*-map indicates the time range of significant peak effects, from earliest to latest significant time points. On the left, the scalp map displays the peak effect of the given cluster using an *F*-map at the indicated PST, displayed on a 2D sensor layout. On the right, difference waveforms (deviant–standard) are shown for each phase, with the chosen sensor location on the scalp map marked with a star. (B) Expression of clusters in time x sensor space where ERPs to tones in stable phases were more positive than ERPs to tones in volatile phase. MMN: mismatch negativity; PST: peri-stimulus time; FWE: family wise error.

These effects remained significant when including either working memory performance (assessed with the digit span backward task (Wechsler, 1981)), age, or cannabis consumption as covariates, except for the correlation between GF: Role and stable difference waveform (Fig. 4D). This correlation was only marginally significant when either working memory, age, or cannabis was included as a covariate ($p_{cluster} = 0.063$, $p_{cluster} = 0.054$ and $p_{cluster} = 0.087$ respectively). Given the borderline significance of this result prior to covariate adjustment ($p_{cluster} = 0.049$), this result should be interpreted with caution and further validation with a larger sample size is needed. Notably, none of these covariates had a significant impact on the difference waveform.

### 3.3. Cortical source of mismatch stability and functional impairment relationship

Next, we identified the cortical sources underlying the difference waveform and its correlations with functioning, as shown in Figure 4. Source activity of the difference waveform is detailed in the Supplementary Material. We found a significant negative correlations between GF: Role and source activity in left STG peaking at 160 ms for

the stable difference waveform (peak, $t_{(1,41)} = 4.8$; $p < 0.001$; Fig. 5A). Notably, this effect in source space coincides with the cluster identified in our sensor space analysis, which exhibited marginal significance in sensor space, peaking at 145 ms in the left temporal region (peak, $t_{(1,41)} = 4.1$; $p = 0.056$; Fig. 4D). The presence of a significant effect in source space might suggest that our sensor space analysis lacked sufficient statistical power, as the sensor analysis was conducted across the entire sensor space, whereas the source analysis was constrained to six sources that were previously associated with MMN generation.

### 3.4. Expression of hierarchical prediction errors and precision ratios

At the sensor level, we observed significant correlations between the low-level (sensory) and high-level (volatility) unweighted PEs and precision ratio trajectories with EEG amplitude (Fig. 6). Specifically, low-level PEs showed an early peak at 137 ms in frontal-central electrodes (peak, $F_{(1,42)} = 140.9$; $p < 0.001$; Fig. 6A), while high-level PEs peaked at 141 ms in frontal-central electrodes (peak, $F_{(1,42)} = 136.2$; $p < 0.001$; Fig. 6B). Additionally, the low-
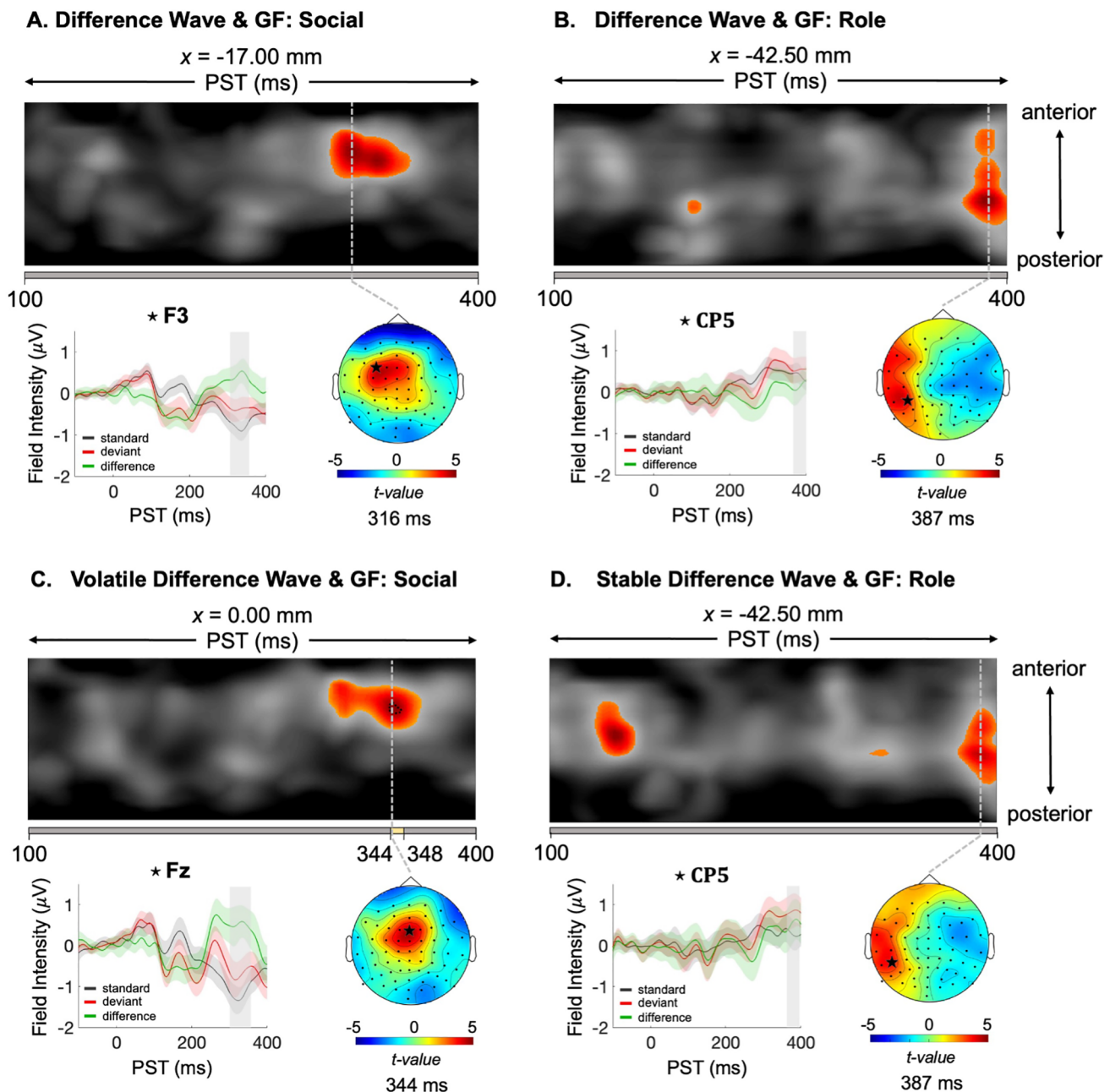
**Fig. 4.** Positive correlation of the difference waveform with psychosocial function. (A–D) Maximum intensity projection *t*-map illustrating the correlation between global function (GF) and difference waveform ERPs across and anterior to posterior scalp locations (top). Significant peak-level effects ($p < 0.05$, whole-volume FWE-corrected) are outlined by black contours, while the colored area indicates *t*-values exceeding the cluster-defining threshold of $p < 0.001$, uncorrected. The yellow bar at the bottom of the *t*-map indicates the time range of significant peak effects, from earliest to latest significant time points. In the absence of a yellow bar, only cluster-level effects were significant. Note the statistical analysis was limited to 100–400 ms post-stimulus. On the left, the grand-average ERPs for standard, deviant, and their difference waveform (deviant – standard) at the peak effect location are shown. The shaded gray area indicates the duration of the significant effect for the specified cluster. On the right, the scalp map highlights the selected sensor location, marked by a star. The scalp map displays the peak effect of the given cluster using an *F*-map at the indicated PST, displayed on a 2D sensor layout. PST: peri-stimulus time; FWE: family wise error.

## A.  Stable MMN and GF Role: Negative Correlation
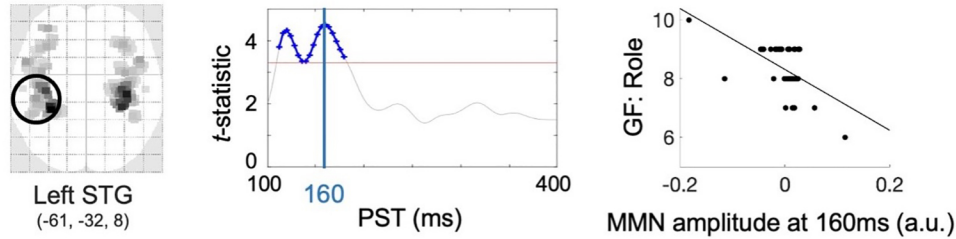


## B.  Low-level PE ($\delta_1$) and GF Role: Negative Correlation
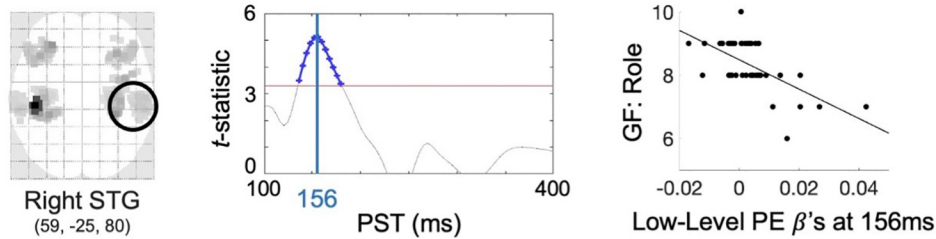


## C.  High-level PE ($\delta_2$) and GF Role: Negative Correlation
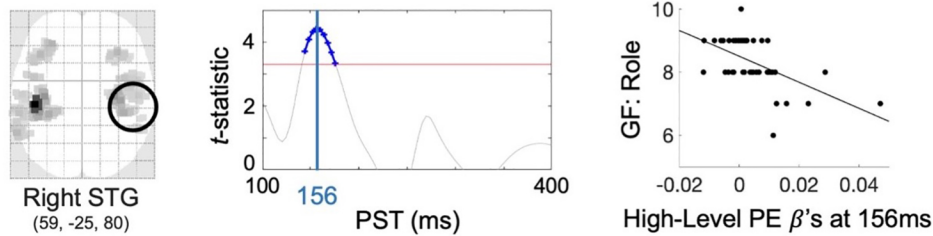


**Fig. 5.**   Cortical sources associated with functional impairment. (A) A negative correlation between the stable mismatch response and GF: Role scores were observed, peaking at 160 ms (301–309 ms) in the left superior temporal gyrus (STG). The source activation estimated from the difference waveform (MMN) at the peak time point (160 ms) is depicted using an SPM-glass brain (left). Significant $t$-contrasts of the stable mismatch response and global function (GF) scores are presented for the given source (left STG) over peri-stimulus time (PST) (middle). The significance threshold using peak-family wise error (FWE) correction is indicated by the red horizontal line, and all time points above this threshold are colored in blue. The peak time point is marked by the blue vertical line. The scatter plot (right) illustrates the relationship between GF scores and the stable MMN amplitude at the peak time point (160 ms) and source (left STG). (B) A negative correlation between the low-level ($\delta_1$) and (C) high-level ($\delta_2$) unweighted prediction errors (PEs) and global function (GF): Role scores was observed, peaking at 156 ms in the right STG. The source activation was estimated from the grand-averaged difference waveform (10% highest–10% lowest $\delta_1$ or $\delta_2$ trials, respectively) at the peak time point (156 ms). MMN: mismatch negativity; PST: peri-stimulus time; GF: global function; A1: primary auditory cortex; STG: superior temporal gyrus.
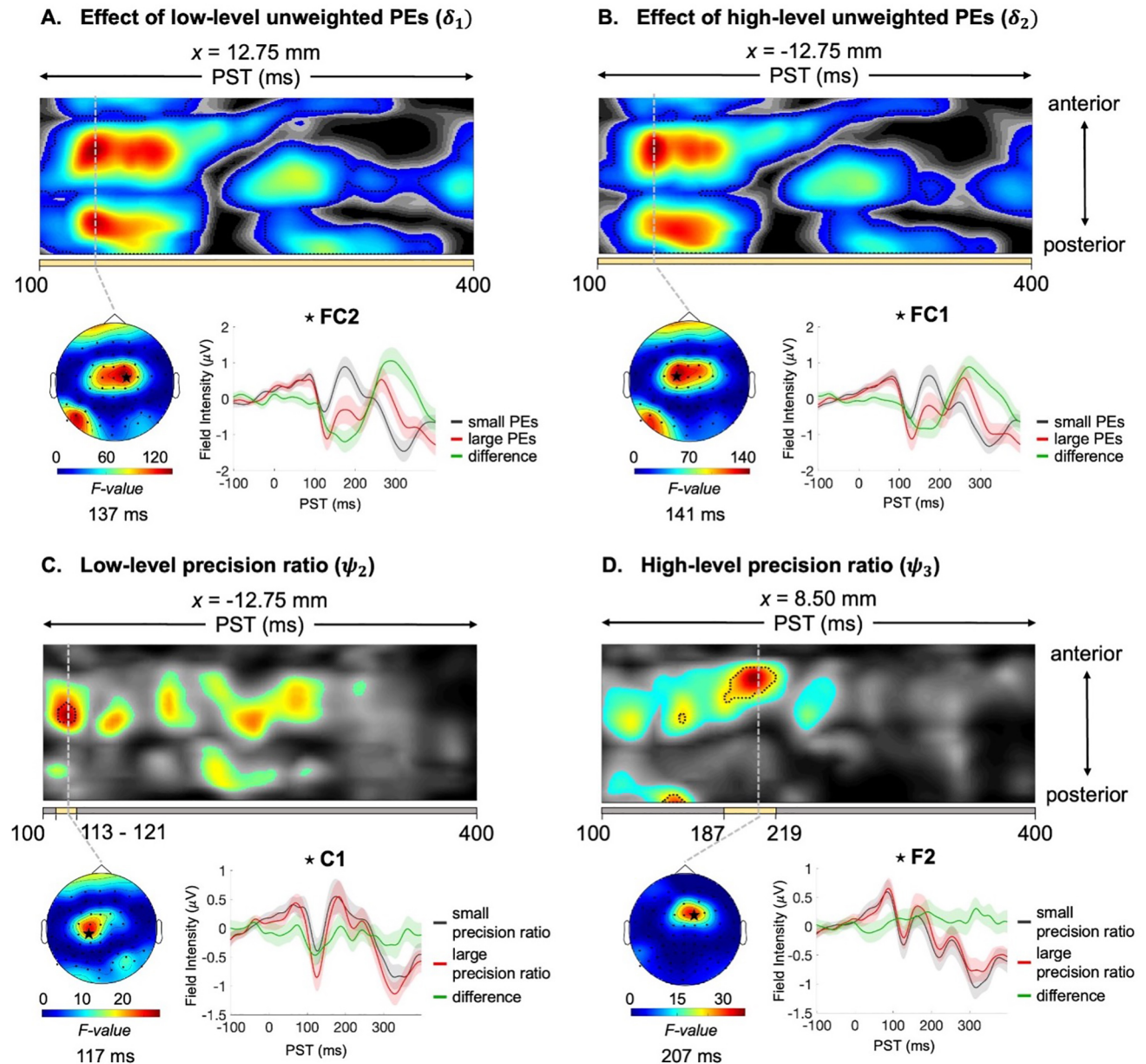
level precision ratio peaked at 117 ms in central electrodes (peak, $F_{(1,42)} = 28.9$; $p = 0.015$; Fig. 6C) whereas the high-level precision ratio peaked at 207 ms in frontal electrodes (peak, $F_{(1,42)} = 36.3$; $p = 0.002$; Fig. 6D). Correlations were also found between trial-wise EEG activity and both low-level and high-level pwPEs (Supplementary Fig. S2). No significant correlations were found between the model trajectories and functioning in sensor space.

### 3.5.   Cortical sources of hierarchical prediction errors and functional impairment relationships

Next, we analyzed the cortical generators of these sensor-level unweighted PE and precision effects. Our findings revealed low-level and high-level unweighted PE expression in the left STG (peak, $F_{(1,42)} = 17.4$; $p = 0.003$ and peak, $F_{(1,42)} = 21.2$; $p = 0.001$, respectively), as well as high-level unweighted PE expression in the left (peak, $F_{(1,42)} = 18.3$; $p = 0.003$) and right A1 (peak, $F_{(1,42)} = 22.4$; $p = 0.001$). The left A1 was additionally associated with high-level precision ratio (peak, $F_{(1,42)} = 12.8$; $p = 0.022$).

Although no correlations with function were found at the sensor level, significant correlations emerged at the source level. Similar to our mismatch analysis, this difference could be attributed to the increased power in the source analysis. Notably, significant negative correlations emerged between GF: Role and the expression of both low-level unweighted PE (peak, $t_{(1,41)} = 5.2$; $p < 0.001$;

**A.  Effect of low-level unweighted PEs ($\delta_1$)**

$x = 12.75$ mm



**B.  Effect of high-level unweighted PEs ($\delta_2$)**

$x = -12.75$ mm



**C.  Low-level precision ratio ($\psi_2$)**

$x = -12.75$ mm



**D.  High-level precision ratio ($\psi_3$)**

$x = 8.50$ mm



**Fig. 6.**   Unweighted prediction error (PE) and precision ratio expression. Maximum intensity projection $F$-map illustrating the effect of (A) low-level sensory PEs ($\varepsilon_2$), (B) high-level volatility PEs ($\varepsilon_3$), (C) low-level sensory precision ratio ($\psi_2$), and (D) high-level volatility precision ratio ($\psi_3$) on EEG amplitudes across anterior to posterior scalp locations (top). Significant peak-level effects ($p < 0.05$, whole-volume FWE-corrected) are outlined by black contours, while the colored area indicates $F$-values exceeding the cluster-defining threshold of $p < 0.001$, uncorrected. The yellow bar at the bottom of the $F$-map indicates the time range of significant peak effects, from earliest to latest significant time points. On the left, the scalp map displays the peak effect of the given cluster using an $F$-map at the indicated PST, displayed on a 2D sensor layout. On the right, ERPs were averaged across the electrode at the peak of the significant clusters using the 10% largest and 10% smallest PE (or precision ratio) values. The selected electrode position indicated by a star on the scalp map. MMN: mismatch negativity; PST: peri-stimulus time; FWE: family wise error.

Fig. 5B) and high-level unweighted PE (peak, $t_{(1,41)} = 4.4$; $p = 0.001$; Fig. 5C) in the right STG peaking at 156 ms. These source effects remained significant when including either working memory performance, age, or cannabis consumption as covariates (see Supplementary Material for detailed discussion on the effect of covariates in the source-space analysis).

## 4.  DISCUSSION

In this study, we investigated the EEG signatures of hierarchical predictive errors during ERP generation using a modified auditory oddball paradigm in healthy controls. We found that stability significantly impacted the difference waveform, with a reduced expression of the auditory

MMN in volatile compared to stable phases. Importantly, despite using a novel auditory oddball task, these findings replicate established effects in the literature (Dzafic et al., 2020; Kotchoubey, 2014; Todd et al., 2011, 2014; Weber et al., 2022), confirming the validity and effectiveness of our experimental approach.

Importantly, we found that this difference waveform was differentially related to psychosocial functioning, with the stable condition correlating with Global Functioning: Social scores and the volatile condition with Global Functioning: Role scores. Additionally, we observed that both PEs and precisions were expressed during the oddball task, with PEs correlating with ERPs over a larger time window and precision ratios at distinct points, and these PE expressions were linked to functioning in the superior temporal gyrus. These findings highlight the nuanced role of predictive coding in social and role functioning, offering insights into the neurobiological underpinnings of psychosocial behaviors.

### 4.1. Difference waveform and psychosocial functioning

We observed a positive correlation between the difference waveform across the entire paradigm and GF: Social scores, peaking at 332 ms post-stimulus (Fig. 4A). This correlation was significant during the volatile phase of the paradigm, peaking between 344 and 348 ms post-stimulus (Fig. 4C), but was not significant during the stable phase. The timing and topography of this correlation corresponded with the P3a component, suggesting that smaller P3a amplitudes may be associated with reduced psychosocial functioning in HCs (Light et al., 2007), a finding replicated across psychosis spectrum disorders (Carrión et al., 2015; Hermens et al., 2010; Murphy et al., 2020). Given the age-related decline of P3a amplitudes (Kiang et al., 2009) and its reduction across clinical populations (Chen et al., 2015; Kaur et al., 2011; Kim et al., 2020; Light et al., 2015; Solís-Vivanco et al., 2015), a blunted P3a response may serve as an indicator of broader cognitive dysfunction.

This is particularly relevant given the inherent volatility of the social world, where accurate interpretation of social cues—integrating facial expressions, body language, tone of voice and conversational context—is crucial for inferring others' intentions. Moreover, these cues must be weighed against their relative reliability (Griffin & Fletcher, 2017) and a failure to accurately integrate and interpret social information may lead to the generation of aberrant beliefs about others' intentions or actions (Diaconescu et al., 2019).

Moreover, we identified an additional correlation between the difference waveform during the stable phase and GF: Role scores (Fig. 4B, D). Although this effect was of marginal significance, it suggests a potential link between neural responses and role functioning even in stable conditions. Caution is warranted in interpreting this finding, and replication in a larger sample is required to substantiate this effect.

Our source-level analysis further revealed that a more positive stable difference waveform amplitude within the left STG, peaking at 160 ms post-stimulus, correlated with poorer role functioning. This effect coincides with the timing of the MMN component. While previous studies have localized the MMN to the left STG and observed reductions in MMN amplitude among patients with schizophrenia (Erickson et al., 2016), our study is the first to link MMN amplitude in the left STG to role functioning within a healthy population. Interestingly, the left STG has consistently shown grey matter loss in patients with schizophrenia (Vita et al., 2012), and reductions in MMN amplitude have been associated with poorer psychosocial functioning in this population (Hamilton et al., 2018; Thomas et al., 2017). Our findings suggest that even among healthy individuals, variability in MMN amplitude in the left STG may be linked to individual differences in role functioning, offering new insights into the neurobiological mechanisms underlying psychosocial behaviors.

### 4.2. Multiple, hierarchically-related prediction errors underlie the MMN

In line with Weber et al. (2022) and other single-trial MMN analyses (Hauke et al., 2023; Stefanics et al., 2018; Weber et al., 2020), we found significant associations between hierarchical PE expression and trial-wise EEG activity (Fig. 6). Notably, correlations between unweighted PEs and EEG amplitudes occurred mainly in three distinct time windows or cluster groups, potentially corresponding to the MMN, P3a, and reorienting negativity (RON) component. The RON, characterized by a negative waveform, signals attentional reorientation, expressed after the P3a component in frontal electrodes. These findings suggest a potential link between the MMN/P3a/RON complex and hierarchical PE expression.

Previous studies have investigated hierarchical PE expression in the context of MMN generation using oddball-like paradigms, such as the local-global paradigm. However, the nature of PEs identified through the HGF is distinct from those identified in the local-global paradigm. In the HGF, hierarchical levels are coupled via their variances, implying a temporal hierarchy where higher levels reflect the slower rates of change relative to the level below (Hauke, 2022). In contrast, the local-global paradigm assumes a hierarchy based on "what" is changing—that is, local or global deviance. Hence, the

use of hierarchical Bayesian models is useful for understanding the dynamics of change, rather than just the occurrence of change.

Using source reconstruction, we found low-level and high-level unweighted PE expression in the left STG, coinciding with the timing of the MMN, and high-level precision ratio expression in the right STG, partially aligning with the P3a component. Contrary to our initial hypothesis that low-level PEs would be generated in early sensory regions and high-level PEs in higher-level regions, our findings suggest a more complex distribution of predictive coding processes across brain regions. Additionally, a negative correlation between GF: Role and unweighted low- and high-level PEs in the right STG was found at 156 ms post-stimulus, highlighting the STG's involvement in PE computation, with increased PE expression linked to diminished role functioning. Notably, this effect aligns with the correlation between the positive stable difference waveform amplitude and role functioning within the left STG, peaking at 160 ms post-stimulus. Taken together, these ERP-based and model-based findings suggest a role for the STG in MMN–PE expression and its association with role functioning.

Furthermore, higher volatility precision ratios (i.e., learning rate) correlated with decreased social function in the primary auditory cortex at 254 ms, suggesting that individuals with lower social function may be more sensitive to environmental volatility. Previous work has shown that increased expectations of volatility are associated with higher levels of paranoia (Reed et al., 2020) and emerging psychosis (Cole et al., 2020; Hauke et al., 2024). This finding aligns with our earlier observations where the volatile difference waveform correlated with GF: Social (Fig. 4C), indicating a potential role for volatility precision ratio in later P3a expression. Higher cognitive levels (e.g., frontal cortex) may represent expected predictions (e.g., anticipated tone), while lower levels (e.g., primary auditory cortex) compute weighted PEs about the tones and environmental volatility.

Notably, despite using a non-social task with healthy controls, we still observed that changes in volatile MMN expression and volatility learning were associated with reduced social functioning, specifically. More investigation is needed to understand how social context influences volatility perception in both healthy individuals and across psychosis spectrum disorders (Gibbs-Dean et al., 2023).

### 4.3.  Limitations

This study has several limitations. First, as the MMN paradigm is a passive task, the HGF model could not be fit to participants' behavior. Instead, belief trajectories were simulated based on the assumption of a Bayes-optimal learner, consistent with previous studies (Hauke et al., 2023; Weber et al., 2020, 2022). Future research should consider employing an active oddball task (Hamilton et al., 2019), which would allow for direct measurement of participants' behavior. This approach would allow us to perform model selection based on participants' behavior and to test the current computational approach against other theoretical models of MMN generation (Lieder, Stephan, et al., 2013). Second, since the data presented here were collected as part of an early psychosis study (Hauke et al., 2024), the functioning assessment employed was not tailored to or developed for the assessment of broader populations (Cornblatt et al., 2007). However, our results suggest that there are relevant relationships between early auditory information processing and psychosocial functioning. Thus, future studies should include more comprehensive assessments of functioning to better characterize the effects (e.g., WHDAS 2.0) (Gold, 2014).

### 4.4.  Future directions

Future research should leverage computational models with greater physiological detail, like connectome-based neurophysiological modeling (Momi et al., 2023) and dynamic causal modeling (David et al., 2006; Friston et al., 2003, 2019), to delve deeper into the neurobiology of hierarchical PEs and MMN expression. These models can shed light on local neural dynamics, receptor densities, and the interplay between precisions, PEs, and AMPAR/NMDAR function. Finally, to enhance the external validity of our findings related to early information processing in HCs, we recommend that future studies incorporate larger sample sizes and measures like WHODAS 2.0 to bolster the generalizability of the results.

### 5.  CONCLUSION

In conclusion, our exploratory study investigated the neural correlates of hierarchical PEs during MMN generation using a novel auditory oddball paradigm. We found a significant effect of stability on the mismatch difference waveform, with larger responses in stable phases and associations with psychosocial functioning. Moreover, we find evidence for the role of hierarchically-related PEs in the generation of the MMN and their association with functioning in the superior temporal gyrus. These results underscore the importance of predictive coding in understanding in early auditory information processing and psychosocial functioning.

## DATA AND CODE AVAILABILITY

The analysis code for this study is publicly available at https://github.com/colleenc11/COMPI_MMN. The data is publicly available at https://osf.io/exsj3/.

## AUTHOR CONTRIBUTIONS

C.E.C. performed the data analysis and drafted the manuscript. D.J.H. conceptualized the study, recruited participants, collected data, contributed to the development of the analysis pipelines, and revised the manuscript. M.W. recruited participants, collected data, and revised the manuscript. C.A. provided rater training, supervision and revised the manuscript. R.B. recruited participants, collected data, and revised the manuscript. S.B. and V.R. provided infrastructure, supervision and revised the manuscript. A.O.D. conceptualized and supervised the study, and revised the manuscript.

## FUNDING

## DECLARATION OF COMPETING INTEREST

The authors have no competing interests to declare.

## SUPPLEMENTARY MATERIALS

Supplementary material for this article is available with the online version here: https://doi.org/10.1162/imag_a_00461.

## REFERENCES

Asadzadeh, S., Yousefi Rezaii, T., Beheshti, S., Delpak, A., & Meshgini, S. (2020). A systematic review of EEG source localization techniques and their applications on diagnosis of brain abnormalities. *Journal of Neuroscience Methods*, *339*, 108740. https://doi.org/10.1016/j.jneumeth.2020.108740

Camalier, C. R., Scarim, K., Mishkin, M., & Averbeck, B. B. (2019). A comparison of auditory oddball responses in dorsolateral prefrontal cortex, basolateral amygdala, and auditory cortex of macaque. *Journal of Cognitive Neuroscience*, *31*(7), 1054–1064. https://doi.org/10.1162/jocn_a_01387

Carrión, R. E., Cornblatt, B. A., McLaughlin, D., Chang, J., Auther, A. M., Olsen, R. H., & Javitt, D. C. (2015). Contributions of early cortical processing and reading ability to functional status in individuals at clinical high risk for psychosis. *Schizophrenia Research*, *164*(1–3), 1–7. https://doi.org/10.1016/j.schres.2015.01.030

Charlton, C. E., Lepock, J. R., Hauke, D. J., Mizrahi, R., Kiang, M., & Diaconescu, A. O. (2022). Atypical prediction error learning is associated with prodromal symptoms in individuals at clinical high risk for psychosis. *Schizophrenia*, *8*(1), 105. https://doi.org/10.1038/s41537-022-00302-3

Chen, J., Zhang, Y., Wei, D., Wu, X., Fu, Q., Xu, F., Wang, H., Ye, M., Ma, W., Yang, L., & Zhang, Z. (2015). Neurophysiological handover from MMN to P3a in first-episode and recurrent major depression. *Journal of Affective Disorders*, *174*, 173–179. https://doi.org/10.1016/j.jad.2014.11.049

Cole, D. M., Diaconescu, A. O., Pfeiffer, U. J., Brodersen, K. H., Mathys, C. D., Julkowski, D., Ruhrmann, S., Schilbach, L., Tittgemeyer, M., Vogeley, K., & Stephan, K. E. (2020). Atypical processing of uncertainty in individuals at risk for psychosis. *NeuroImage: Clinical*, *26*, 102239. https://doi.org/10.1016/j.nicl.2020.102239

Cornblatt, B. A., Auther, A. M., Niendam, T., Smith, C. W., Zinberg, J., Bearden, C. E., & Cannon, T. D. (2007). Preliminary findings for two new measures of social and role functioning in the prodromal phase of schizophrenia. *Schizophrenia Bulletin*, *33*(3), 688–702. https://doi.org/10.1093/schbul/sbm029

David, O., Kiebel, S. J., Harrison, L. M., Mattout, J., Kilner, J. M., & Friston, K. J. (2006). Dynamic causal modeling of evoked responses in EEG and MEG. *NeuroImage*, *30*(4), 1255–1272. https://doi.org/10.1016/j.neuroimage.2005.10.045

Diaconescu, A. O., Hauke, D. J., & Borgwardt, S. (2019). Models of persecutory delusions: A mechanistic insight into the early stages of psychosis. *Molecular Psychiatry*, *24*(9), 1258–1267. https://doi.org/10.1038/s41380-019-0427-z

Dzafic, I., Randeniya, R., Harris, C. D., Bammel, M., & Garrido, M. I. (2020). Statistical learning and inference is impaired in the nonclinical continuum of psychosis. *The Journal of Neuroscience*, *40*(35), 6759–6769. https://doi.org/10.1523/JNEUROSCI.0315-20.2020

Erickson, M. A., Ruffle, A., & Gold, J. M. (2016). A meta-analysis of mismatch negativity in schizophrenia: From clinical risk to disease specificity and progression. *Biological Psychiatry*, *79*(12), 980–987. https://doi.org/10.1016/j.biopsych.2015.08.025

Flandin, G., & Friston, K. J. (2019). Analysis of family-wise error rates in statistical parametric mapping using random field theory. *Human Brain Mapping*, *40*(7), 2052–2054. https://doi.org/10.1002/hbm.23839

Frässle, S., Aponte, E. A., Bollmann, S., Brodersen, K. H., Do, C. T., Harrison, O. K., Harrison, S. J., Heinzle, J., Iglesias, S., Kasper, L., Lomakina, E. I., Mathys, C., Müller-Schrader, M., Pereira, I., Petzschner, F. H., Raman, S., Schöbi, D., Toussaint, B., Weber, L. A., Yao, Y., … Stephan, K. E. (2021). TAPAS: An open-source software package for translational neuromodeling and computational psychiatry. *Frontiers in Psychiatry*, *12*, 680811. https://doi.org/10.3389/fpsyt.2021.680811

Friston, K. (2003). Learning and inference in the brain. *Neural Networks*, *16*(9), 1325–1352. https://doi.org/10.1016/j.neunet.2003.06.005

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K., Brown, H. R., Siemerkus, J., & Stephan, K. E. (2016). The dysconnection hypothesis. *Schizophrenia Research*, *176*(2–3), 83–94. https://doi.org/10.1016/j.schres.2016.07.014

Friston, K., Harrison, L., Daunizeau, J., Kiebel, S., Phillips, C., Trujillo-Barreto, N., Henson, R., Flandin, G., & Mattout, J. (2008). Multiple sparse priors for the M/EEG inverse problem. *NeuroImage*, *39*(3), 1104–1120. https://doi.org/10.1016/j.neuroimage.2007.09.048

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, *19*(4), 1273–1302. https://doi.org/10.1016/s1053-8119(03)00202-7

Friston, K. J., Preller, K. H., Mathys, C., Cagnan, H., Heinzle, J., Razi, A., & Zeidman, P. (2019). Dynamic causal modelling revisited. *NeuroImage*, *199*, 730–744. https://doi.org/10.1016/j.neuroimage.2017.02.045

Fong, C. Y., Law, W. H. C., Uka, T., & Koike, S. (2020). Auditory mismatch negativity under predictive coding framework and its role in psychotic disorders. *Frontiers in Psychiatry*, *11,* 557932. https://doi.org/10.3389/fpsyt.2020.557932

Garrido, M. I., Friston, K. J., Kiebel, S. J., Stephan, K. E., Baldeweg, T., & Kilner, J. M. (2008). The functional anatomy of the MMN: A DCM study of the roving paradigm. *NeuroImage*, *42*(2), 936–944. https://doi.org/10.1016/j.neuroimage.2008.05.018

Garrido, M. I., Kilner, J. M., Kiebel, S. J., & Friston, K. J. (2009). Dynamic causal modeling of the response to frequency deviants. *Journal of Neurophysiology*, *101*(5), 2620–2631. https://doi.org/10.1152/jn.90291.2008

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology*, *120*(3), 453–463. https://doi.org/10.1016/j.clinph.2008.11.029

Giard, M. H., Perrin, F., Pernier, J., & Bouchet, P. (1990). Brain generators implicated in the processing of auditory stimulus deviance: A topographic event-related potential study. *Psychophysiology*, *27*(6), 627–640. https://doi.org/10.1111/j.1469-8986.1990.tb03184.x

Gibbs-Dean, T., Katthagen, T., Tsenkova, I., Ali, R., Liang, X., Spencer, T., & Diederen, K. (2023). Belief updating in psychosis, depression and anxiety disorders: A systematic review across computational modelling approaches. *Neuroscience and Biobehavioral Reviews*, *147*, 105087. https://doi.org/10.1016/j.neubiorev.2023.105087

Gold, L. H. (2014). DSM-5 and the assessment of functioning: The World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0). *The Journal of the American Academy of Psychiatry and the Law*, *42*(2), 173–181. https://doi.org/10.1016/j.jphys.2020.11.004

Griffin, J. D., & Fletcher, P. C. (2017). Predictive processing, source monitoring, and psychosis. *Annual Review of Clinical Psychology*, *13*, 265–289. https://doi.org/10.1146/annurev-clinpsy-032816-045145

Hamilton, H. K., Perez, V. B., Ford, J. M., Roach, B. J., Jaeger, J., & Mathalon, D. H. (2018). Mismatch negativity but not P300 is associated with functional disability in schizophrenia. *Schizophrenia Bulletin*, *44*(3), 492–504. https://doi.org/10.1093/schbul/sbx104

Hamilton, H. K., Roach, B. J., Bachman, P. M., Belger, A., Carrion, R. E., Duncan, E., Johannesen, J. K., Light, G. A., Niznikiewicz, M. A., Addington, J., Bearden, C. E., Cadenhead, K. S., Cornblatt, B. A., McGlashan, T. H., Perkins, D. O., Seidman, L. J., Tsuang, M. T., Walker, E. F., Woods, S. W., Cannon, T. D., … Mathalon, D. H. (2019). Association between P300 responses to auditory oddball stimuli and clinical outcomes in the psychosis risk syndrome. *JAMA Psychiatry*, *76*(11), 1187–1197. https://doi.org/10.1001/jamapsychiatry.2019.2135

Hauke, D. J. (2022). *Hierarchical Bayesian inference in psychosis* [Doctoral dissertation, University of Basel]. edoc. https://edoc.unibas.ch/89780/

Hauke, D. J., Charlton, C. E., Schmidt, A., Griffiths, J. D., Woods, S. W., Ford, J. M., Srihari, V. H., Roth, V., Diaconescu, A. O., & Mathalon, D. H. (2023). Aberrant hierarchical prediction errors are associated with transition to psychosis: A computational single-trial analysis of the mismatch negativity. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *8*(12), 1176–1185. https://doi.org/10.1016/j.bpsc.2023.07.011

Hauke, D. J., Wobmann, M., Andreou, C., Mackintosh, A. J., de Bock, R., Karvelis, P., Adams, R. A., Sterzer, P., Borgwardt, S., Roth, V., & Diaconescu, A. O. (2024). Altered perception of environmental volatility during social learning in emerging psychosis. *Computational Psychiatry*, *8*(1), 1–22. https://doi.org/10.5334/cpsy.95

Hermens, D. F., Ward, P. B., Hodge, M. A., Kaur, M., Naismith, S. L., & Hickie, I. B. (2010). Impaired MMN/P3a complex in first-episode psychosis: Cognitive and psychosocial associations. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, *34*(6), 822–829. https://doi.org/10.1016/j.pnpbp.2010.03.019

Kanai, R., Komura, Y., Shipp, S., & Friston, K. (2015). Cerebral hierarchies: Predictive processing, precision and the pulvinar. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *370*(1668), 20140169. https://doi.org/10.1098/rstb.2014.0169

Kaur, M., Battisti, R. A., Ward, P. B., Ahmed, A., Hickie, I. B., & Hermens, D. F. (2011). MMN/P3a deficits in first episode psychosis: Comparing schizophrenia-spectrum and affective-spectrum subgroups. *Schizophrenia Research*, *130*(1–3), 203–209. https://doi.org/10.1016/j.schres.2011.03.025

Kiang, M., Braff, D. L., Sprock, J., & Light, G. A. (2009). The relationship between preattentive sensory processing deficits and age in schizophrenia patients. *Clinical Neurophysiology*, *120*(11), 1949–1957. https://doi.org/10.1016/j.clinph.2009.08.019

Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS Computational Biology*, *4*(11), e1000209. https://doi.org/10.1371/journal.pcbi.1000209

Kiebel, S. J., & Friston, K. J. (2004). Statistical parametric mapping for event-related potentials: I. Generic considerations. *NeuroImage*, *22*(2), 492–502. https://doi.org/10.1016/j.neuroimage.2004.02.012

Kim, S., Baek, J. H., Shim, S. H., Kwon, Y. J., Lee, H. Y., Yoo, J. H., & Kim, J. S. (2020). Mismatch negativity indices and functional outcomes in unipolar and bipolar depression. *Scientific Reports*, *10*(1), 12831. https://doi.org/10.1038/s41598-020-69776-4

Kotchoubey, B. (2014). First love does not die: A sustaining primacy effect on ERP components in an oddball paradigm. *Brain Research*, *1556*, 38–45. https://doi.org/10.1016/j.brainres.2014.02.012

Lieder, F., Daunizeau, J., Garrido, M. I., Friston, K. J., & Stephan, K. E. (2013). Modelling trial-by-trial changes in the mismatch negativity. *PLoS Computational Biology*, *9*(2), e1002911. https://doi.org/10.1371/journal.pcbi.1002911

Lieder, F., Stephan, K. E., Daunizeau, J., Garrido, M. I., & Friston, K. J. (2013). A neurocomputational model of the mismatch negativity. *PLoS Computational Biology*, *9*(11), e1003288. https://doi.org/10.1371/journal.pcbi.1003288

Light, G. A., Swerdlow, N. R., & Braff, D. L. (2007). Preattentive sensory processing as indexed by the MMN and P3a brain responses is associated with cognitive and psychosocial functioning in healthy adults. *Journal of Cognitive Neuroscience*, *19*(10), 1624–1632. https://doi.org/10.1162/jocn.2007.19.10.1624

Light, G. A., Swerdlow, N. R., Thomas, M. L., Calkins, M. E., Green, M. F., Greenwood, T. A., Gur, R. E., Gur, R. C., Lazzeroni, L. C., Nuechterlein, K. H., Pela, M., Radant, A. D., Seidman, L. J., Sharp, R. F., Siever, L. J., Silverman, J. M., Sprock, J., Stone, W. S., Sugar, C. A., Tsuang, D. W., … Turetsky, B. I. (2015). Validation of mismatch negativity and P3a for use in multi-site studies of schizophrenia: Characterization of demographic, clinical, cognitive, and functional correlates in COGS-2. *Schizophrenia Research*, *163*(1–3), 63–72. https://doi.org/10.1016/j.schres.2014.09.042

Mathys, C., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A Bayesian foundation for individual learning under uncertainty. *Frontiers in Human Neuroscience*, *5*, 39. https://doi.org/10.3389/fnhum.2011.00039

Mathys, C. D., Lomakina, E. I., Daunizeau, J., Iglesias, S., Brodersen, K. H., Friston, K. J., & Stephan, K. E. (2014). Uncertainty in perception and the Hierarchical Gaussian Filter. *Frontiers in Human Neuroscience*, *8*, 825. https://doi.org/10.3389/fnhum.2014.00825

Molholm, S., Martinez, A., Ritter, W., Javitt, D. C., & Foxe, J. J. (2005). The neural circuitry of pre-attentive auditory change-detection: An fMRI study of pitch and duration mismatch negativity generators. *Cerebral Cortex*, *15*(5), 545–551. https://doi.org/10.1093/cercor/bhh155

Momi, D., Wang, Z., & Griffiths, J. D. (2023). TMS-evoked responses are driven by recurrent large-scale network dynamics. *eLife*, *12*, e83232. https://doi.org/10.7554/eLife.83232

Murphy, T. K., Haigh, S. M., Coffman, B. A., & Salisbury, D. F. (2020). Mismatch negativity and impaired social functioning in long-term and in first episode schizophrenia spectrum psychosis. *Frontiers in Psychiatry*, *11*, 544. https://doi.org/10.3389/fpsyt.2020.00544

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*(12), 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026

Nolte, G., & Hämäläinen, M. S. (2001). Partial signal space projection for artefact removal in MEG measurements: A theoretical analysis. *Physics in Medicine and Biology*, *46*(11), 2873–2887. https://doi.org/10.1088/0031-9155/46/11/308

Opitz, B., Rinne, T., Mecklinger, A., von Cramon, D. Y., & Schröger, E. (2002). Differential contribution of frontal and temporal cortices to auditory change detection: fMRI and ERP results. *NeuroImage*, *15*(1), 167–174. https://doi.org/10.1006/nimg.2001.0970

Parras, G. G., Nieto-Diego, J., Carbajal, G. V., Valdés-Baizabal, C., Escera, C., & Malmierca, M. S. (2017). Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. *Nature Communications*, *8*(1), 2148. https://doi.org/10.1038/s41467-017-02038-6

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Reed, E. J., Uddenberg, S., Suthaharan, P., Mathys, C. D., Taylor, J. R., Groman, S. M., & Corlett, P. R. (2020). Paranoia as a deficit in non-social belief updating. *eLife*, *9*, e56345. https://doi.org/10.7554/eLife.56345

Ro, E., & Clark, L. A. (2009). Psychosocial functioning in the context of diagnosis: Assessment and theoretical issues. *Psychological Assessment*, *21*(3), 313–324. https://doi.org/10.1037/a0016611

Schöbi, D., Homberg, F., Frässle, S., Endepols, H., Moran, R. J., Friston, K. J., Tittgemeyer, M., Heinzle, J., & Stephan, K. E. (2021). Model-based prediction of muscarinic receptor function from auditory mismatch negativity responses. *NeuroImage*, *237*, 118096. https://doi.org/10.1016/j.neuroimage.2021.118096

Solís-Vivanco, R., Rodríguez-Violante, M., Rodríguez-Agudelo, Y., Schilmann, A., Rodríguez-Ortiz, U., & Ricardo-Garcell, J. (2015). The P3a wave: A reliable neurophysiological measure of Parkinson's disease duration and severity. *Clinical Neurophysiology*, *126*(11), 2142–2149. https://doi.org/10.1016/j.clinph.2014.12.024

Stefanics, G., Heinzle, J., Horváth, A. A., & Stephan, K. E. (2018). Visual mismatch and predictive coding: A computational single-trial ERP study. *The Journal of Neuroscience*, *38*(16), 4020–4030. https://doi.org/10.1523/JNEUROSCI.3365-17.2018

Sterzer, P., Adams, R. A., Fletcher, P., Frith, C., Lawrie, S. M., Muckli, L., Petrovic, P., Uhlhaas, P., Voss, M., & Corlett, P. R. (2018). The predictive coding account of psychosis. *Biological Psychiatry*, *84*(9), 634–643. https://doi.org/10.1016/j.biopsych.2018.05.015

Thomas, M. L., Green, M. F., Hellemann, G., Sugar, C. A., Tarasenko, M., Calkins, M. E., Greenwood, T. A., Gur, R. E., Gur, R. C., Lazzeroni, L. C., Nuechterlein, K. H., Radant, A. D., Seidman, L. J., Shiluk, A. L., Siever, L. J., Silverman, J. M., Sprock, J., Stone, W. S., Swerdlow, N. R., … Light, G. A. (2017). Modeling deficits from early auditory information processing to psychosocial functioning in schizophrenia. *JAMA Psychiatry*, *74*(1), 37–46. https://doi.org/10.1001/jamapsychiatry.2016.2980

Todd, J., Heathcote, A., Whitson, L. R., Mullens, D., Provost, A., & Winkler, I. (2014). Mismatch negativity (MMN) to pitch change is susceptible to order-dependent bias. *Frontiers in Neuroscience*, *8*, 180. https://doi.org/10.3389/fnins.2014.00180

Todd, J., Provost, A., & Cooper, G. (2011). Lasting first impressions: A conservative bias in automatic filters of the acoustic environment. *Neuropsychologia*, *49*(12), 3399–3405. https://doi.org/10.1016/j.neuropsychologia.2011.08.016

Vita, A., De Peri, L., Deste, G., & Sacchetti, E. (2012). Progressive loss of cortical gray matter in schizophrenia: A meta-analysis and meta-regression of longitudinal MRI studies. *Translational Psychiatry*, *2*(11), e190. https://doi.org/10.1038/tp.2012.116

Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., & Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(51), 20754–20759. https://doi.org/10.1073/pnas.1117807108

Weber, L. A., Diaconescu, A. O., Mathys, C., Schmidt, A., Kometer, M., Vollenweider, F., & Stephan, K. E. (2020). Ketamine affects prediction errors about statistical regularities: A computational single-trial analysis of the mismatch negativity. *The Journal of Neuroscience*, *40*(29), 5658–5668. https://doi.org/10.1523/JNEUROSCI.3069-19.2020

Weber, L. A., Tomiello, S., Schöbi, D., Wellstein, K. V., Mueller, D., Iglesias, S., & Stephan, K. E. (2022). Auditory mismatch responses are differentially sensitive to changes in muscarinic acetylcholine versus dopamine receptor function. *eLife*, *11*, e74835. https://doi.org/10.7554/eLife.74835

Wechsler, D. (1981). *Wechsler adult intelligence scale-revised (WAIS-R)*. Psychological Corporation. https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Wechsler+adult+intelligence+scale-revised+%28WAIS-R%29&btnG=

Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., & Evans, A. C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping*, *4*(1), 58–73. https://doi.org/10.1002/(SICI)1097-0193(1996)4:1<58::AID-HBM4>3.0.CO;2-O