

Model-Based Approaches to Investigating Mismatch Responses in Schizophrenia

Dirk C. Gütlin^{1,*} , Hannah H. McDermott^{1,*}, Miro Grunde¹,
and Ryszard Aukštulewicz¹ 

Clinical EEG and Neuroscience
2025, Vol. 56(1) 8–21
© EEG and Clinical Neuroscience
Society (ECNS) 2024



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/15500594241253910
journals.sagepub.com/home/eeg



Abstract

Alterations of mismatch responses (ie, neural activity evoked by unexpected stimuli) are often considered a potential biomarker of schizophrenia. Going beyond establishing the type of observed alterations found in diagnosed patients and related cohorts, computational methods can yield valuable insights into the underlying disruptions of neural mechanisms and cognitive function. Here, we adopt a typology of model-based approaches from computational cognitive neuroscience, providing an overview of the study of mismatch responses and their alterations in schizophrenia from four complementary perspectives: (a) connectivity models, (b) decoding models, (c) neural network models, and (d) cognitive models. Connectivity models aim at inferring the effective connectivity patterns between brain regions that may underlie mismatch responses measured at the sensor level. Decoding models use multivariate spatiotemporal mismatch response patterns to infer the type of sensory violations or to classify participants based on their diagnosis. Neural network models such as deep convolutional neural networks can be used for improved classification performance as well as for a systematic study of various aspects of empirical data. Finally, cognitive models quantify mismatch responses in terms of signaling and updating perceptual predictions over time. In addition to describing the available methodology and reviewing the results of recent computational psychiatry studies, we offer suggestions for future work applying model-based techniques to advance the study of mismatch responses in schizophrenia.

Keywords

mismatch negativity, schizophrenia, computational psychiatry, connectivity, decoding

Received September 30, 2023; revised February 9, 2024; accepted April 23, 2024.

Introduction

Efficient perceptual processing depends on detecting unexpected changes in the sensory environment. For example, when a surprising stimulus follows a sequence of expected ones, macroscopic neural measurements like the electroencephalogram (EEG) yield discernible responses signaling the mismatch between previous sensory regularities and present stimuli. One well-studied neural correlate of this deviance detection in the auditory system is the mismatch negativity (MMN). It is quantified as the difference in response waveforms between deviants versus standards. MMN is observed in classic oddball paradigms¹ where standards are repeated many times,² and in complex stimulus sequences where stimulus expectation is based on context rather than repetition.³ In cognitive and computational terms, it has been linked to prediction error (PE) signaling, resulting from a comparison between the brain's predictions and incoming stimuli.^{2,4} Besides MMN, which occurs around 100 to 200 ms after stimulus onset and is modulated by attention only in some conditions,^{5,6} another widely studied mismatch response (MMR) is the P3 component⁷ of the event-related potential (ERP), occurring around 250 to 500 ms after stimulus onset and more influenced by

attention.⁸ P3 is traditionally linked to attentional switches in response to salient stimulus features but more recent studies have suggested that MMN and P3 are two stages of a predictive hierarchy⁹ and cross-modal evidence of mismatch processing has been found in the P3 time range.¹⁰

MMRs are considered potential schizophrenia biomarkers¹¹ mainly due to established MMN reduction in patients,¹² in addition to more nuanced effects on P3 amplitude.¹³ Importantly, MMN reduction is thought to reflect underlying neurobiological impairments, such as glutamatergic N-methyl-D-aspartate receptor (NMDAR) hypofunction.^{14,15} As shown over 20 years ago, administering ketamine (NMDAR antagonist) to healthy volunteers led to reduced MMN, correlating with psychotic-like

¹Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany

*Equal contribution.

Corresponding Author:

Ryszard Aukštulewicz, Department of Education and Psychology, Freie Universität Berlin, Habelschwerdter Allee 45, Berlin 14195, Germany.
Email: ryszard.aukštulewicz@fu-berlin.de

Full-color figures are available online at <http://journals.sagepub.com/home/eeg>.

symptoms.¹⁶ Other neuromodulatory impairments^{17,18} are less clear, with MMN sensitivity reported for cholinergic¹⁹ but not dopaminergic (DA)^{20,21} psychopharmacological manipulations. MMN attenuation has also been shown in prodromal phases and at-risk populations^{22,23} (but see Refs.^{24,25}), indicating its potential as an early biomarker.

Interest in schizophrenia biomarkers has fueled the growth of computational psychiatry,^{26–28} which employs computational models to uncover the mechanisms of psychiatric disorders. These biomarkers can relate to diagnosis (indicating disease presence), prognosis (assessing the likely course of the disease), and risk (estimating the chances of being diagnosed), among other clinical aspects.¹¹ Beyond biomarker identification, computational psychiatry aims to elucidate schizophrenia pathophysiology and inter-individual differences for personalized treatments. Here, we adopt the typology of computational approaches to neuroscientific data outlined in a foundational review of computational cognitive neuroscience,²⁹ including two main types: “from data to theory” and “from theory to data” (Figure 1).

Approaches going from data to theory are largely data-driven, make few assumptions about the underlying cognitive processes (such as change detection), and often (but not always) rely on standard statistical techniques such as regression and classification. These approaches include (a) models of connectivity and dynamics, which in the context of MMRs

aim at elucidating effective connectivity patterns between distinct (sub)-cortical regions underlying the measured signals; and (b) decoding and representational models, which use multivariate pattern analysis (MVPA) and classification techniques to uncover the information present in population responses such as MMRs. Models of connectivity and dynamics are primarily distinguished from decoding and representational models not in terms of mathematical or statistical techniques, but rather in terms of the research questions they are used to elucidate. Connectivity models can provide more mechanistic insights into putative neurophysiological impairments, related to impaired connectivity between brain regions or neuromodulation within distinct neural populations, which may underlie altered MMRs. Conversely, decoding models can be adapted to complex stimulus sequences and experimental paradigms, going beyond simple differences in MMRs between clinical populations and inferring more fine-grained representational contents of MMR responses.

In contrast, approaches going from theory to data typically encompass explicit assumptions about the underlying cognitive processes, and often rely on more complex computational modeling techniques. These approaches include (c) neural network models, comprising neurobiologically inspired but relatively abstract connectionist models mapping inputs to outputs, which can take a form of mapping stimulus inputs to neural

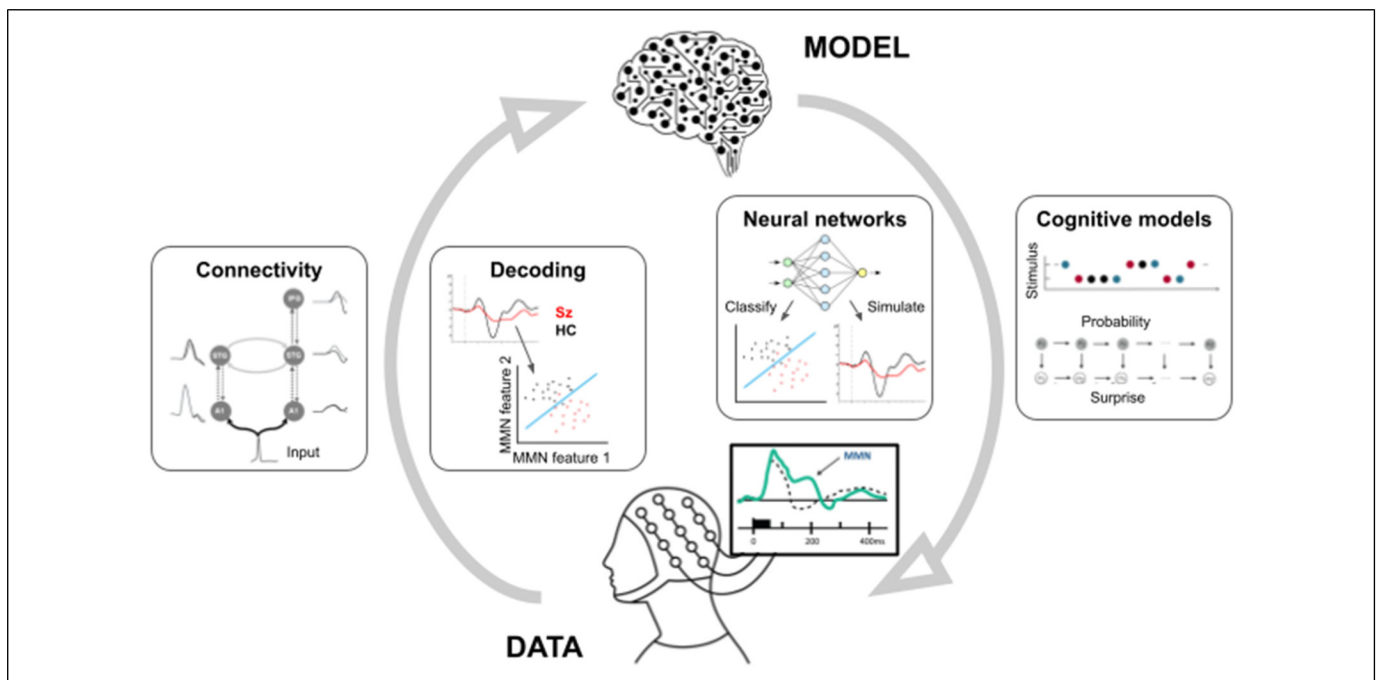


Figure 1. Overview of modeling techniques. In applying computational models to the study of MMRs in schizophrenia, two broadly complementary approaches can be used. Going from empirical data to theoretical/computational models, computational approaches include connectivity modeling (inferring effective connectivity patterns mediating mismatch signaling in patients and controls) and decoding techniques (eg, allowing the classification of study participants into patients and controls based on multivariate MMR features). Going from models to empirical data, computational approaches include neural networks (which, similar to decoding, can allow classifying study participants, but also encompass eg MMR simulation studies) and cognitive models (typically quantifying stimulus sequences in computational terms based on eg probability and surprise).

(eg, MMR) outputs, or neural inputs to clinically relevant (eg, diagnostic or prognostic) outputs; and (d) cognitive models, which aim at understanding the theoretical and algorithmic properties governing trial-by-trial MMR dynamics. While both aim at reproducing cognitive phenomena related to change detection, these models differ in terms of levels of abstraction and complexity: neural network models are composed of a large number of relatively simple building blocks (artificial “neurons”) performing basic operations, while cognitive models have fewer components which perform more abstract operations (eg, quantifying information-theoretical quantities related to surprise). They also differ in terms of research questions they are typically used to answer. While neural network models can be used similarly to decoding models, they can also be used to simulate altered MMRs following different manipulations of network parameters. Conversely, cognitive models can be used to infer eg learning impairments which may subservise MMR alterations in clinical populations.

We will now turn to presenting a survey of each of these four types of model-based approaches—both in the broader context of investigating MMRs, and in the more specific context of the role that MMRs may play in schizophrenia.

Connectivity Models

As mentioned earlier, MMN reduction in schizophrenia has been putatively linked to underlying NMDAR hypofunction.¹⁴ NMDAR hypofunction is also central to the influential “dysconnection” hypothesis of schizophrenia,^{30,31} suggesting abnormal interactions between NMDAR function and other neuromodulatory systems. These interactions lead to aberrant connectivity patterns, including extrinsic (long-range ascending and descending projections) and intrinsic (local synaptic gain control) connectivity. Importantly for the interpretation of MMN in terms of PE responses,² different types of connections have been linked to predictive processing. Descending connections are thought to mediate predictions, ascending connections subservise PEs, while gain control scales their precision. Identifying connectivity patterns which may be impaired in schizophrenia not only has neurobiological relevance, but can also relate to cognitive symptoms.

Among a myriad of connectivity estimation techniques applicable to EEG^{32,33} (for fMRI applications, see elsewhere^{34,35}), we will focus primarily on model-based effective connectivity analysis (as opposed to model-free and/or functional connectivity such as phase locking value, in which connections are symmetric³⁶) based on time-domain data such as ERPs (as opposed to frequency-domain). In particular, dynamic causal modeling (DCM) is a hypothesis-driven, model-based effective connectivity approach³⁷ aspiring to biological realism by modeling interactions between excitatory and inhibitory cells within and between brain regions. Models are typically fitted to individual participant’s data, yielding connectivity estimates which maximize model accuracy (goodness of

fit) while minimizing complexity (preventing overfitting). Connectivity parameters are compared between experimental manipulations (eg, deviants vs standards) or participant groups (eg, patients vs controls) to identify connections sensitive to these factors.^{5,19} DCM has often been applied to EEG or MEG (magnetoencephalographic) data acquired in oddball sequences to uncover connections modulated by stimulus deviance.^{38,39} This approach established a standard model of auditory oddball processing, including bilateral primary auditory cortices (A1), superior temporal gyri (STG), and inferior frontal gyri (IFG), which has since been applied in many MMN studies,^{40–43} including psychopharmacological manipulations^{19,44} and clinical groups.^{45,46} Below, we review such applications in the context of schizophrenia (Figure 2). For more comprehensive connectivity-related results, please refer to recent reviews on early-stage schizophrenia, at-risk mental states,⁴⁷ and the ketamine model of NMDAR hypofunction.⁴⁸

Connectivity models applied to schizophrenia research cover various paradigms, such as classic and multiple oddball,^{49–51} roving oddball,^{52–56} and more complex sequences.^{57–59} In a classic duration oddball study,⁶⁰ patients showed increased temporal-to-prefrontal ascending connectivity and reduced prefrontal cross-hemispheric connectivity, quantified using partial directed coherence. The latter finding was linked to greater negative symptom severity. However, DCM studies on duration and frequency deviants^{61–63} revealed different results, indicating impaired intrinsic (self-)connectivity with putative links to abnormal neuromodulation.¹⁹ This is likely due to unique properties of DCM, which can distinguish between extrinsic and intrinsic connectivity. The locus of intrinsic connectivity modulation varied across studies. An earlier report⁶² suggested that patients and (to a lesser degree) unaffected relatives showed increased prefrontal self-inhibition (interpreted as reflecting NMDAR abnormalities). This finding was recently replicated,⁶³ but additionally positive auditory symptoms were linked to A1 disinhibition. However, another recent study⁶¹ linked prefrontal connectivity to chronic psychosis, whereas first-episode patients showed localized connectivity changes in the left A1. Finally, a multiple-deviant study has compared children at familial risk of schizophrenia or bipolar disorder to age-matched controls.⁵¹ Both at-risk cohorts showed impaired connectivity in A1 and stronger forward connectivity to the prefrontal cortex. Notably, children at risk of schizophrenia exhibited different connectivity patterns than those at risk of bipolar disorder, including impaired intrinsic STG and prefrontal connectivity, as well as weaker extrinsic connectivity to the STG. Based on these studies, MMN alterations in schizophrenia and at-risk populations may stem from disrupted intrinsic (and, to a lesser extent, extrinsic) connectivity in different cortical regions, influenced by disease progression stage and symptom severity.

DCM studies on the roving oddball paradigm showed diverse findings depending on the studied cohort. One MEG study in adolescents with schizophrenia⁵² revealed altered cross-hemispheric connectivity, emphasizing stronger right-to-left

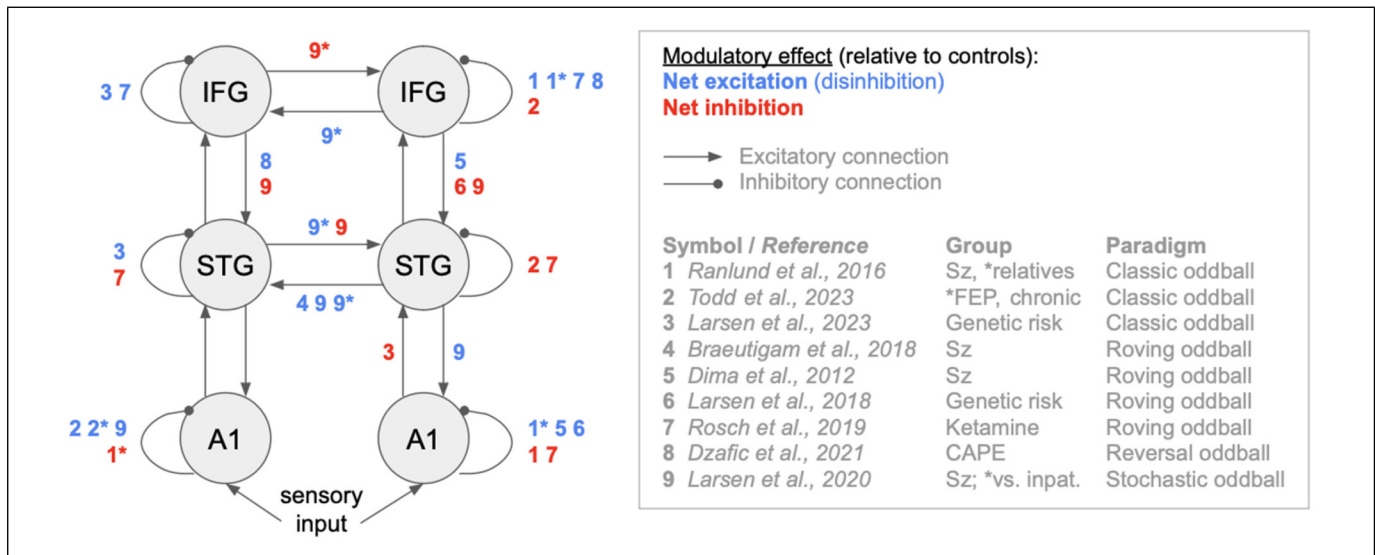


Figure 2. Qualitative overview of effective connectivity results. The graph shows the reported modulatory effects of Sz-relevant groups versus controls, mapped onto the widely used dynamic causal model (DCM) of auditory mismatch responses. The most consistent effects include IFG and A1 disinhibition, right STG inhibition, as well as increased right-to-left STG connectivity in patients or related groups. While the reviewed studies show qualitatively heterogeneous results (largely dependent on the investigated cohort and paradigm), please note that this overview should not be interpreted as a direct comparison of posterior parameter estimates between studies, as different studies may select different winning models. Sz: schizophrenia diagnosis; FEP: first-episode psychosis; CAPE: community assessment of psychotic experiences (quantifying psychotic-like experiences); inpat.: inpatients without psychosis. In case of multiple groups investigated, asterisk denotes the modulatory effect on connectivity associated with membership of a specific group.

connections between bilateral STG; however, intrinsic connectivity was not analyzed. Another MEG study⁵³ with a small sample of 14 schizophrenia patients found reduced self-connectivity in the (right) A1 and increased descending frontotemporal connectivity. The same subset of connections was found in a study involving young nonpsychotic 22q11.2 deletion carriers⁵⁴ at genetic risk for schizophrenia. Here, however, both connectivity parameters were nominally reduced in at-risk individuals, albeit the effects did not survive correction for multiple comparisons. Finally, a ketamine study⁵⁶ linked NMDAR blockade to reduced intrinsic inhibition in prefrontal regions. Overall, these studies predominantly identified changes to intrinsic and descending connectivity, although their specific pattern varied across groups (diagnosed schizophrenia, genetic risk, or pharmacologically induced NMDAR hypofunction).

In complex MMR-evoking sequences where stimulus probabilities change over time, studies have shown that neural responses in healthy volunteers can be influenced by whether sounds are initially perceived as standards or deviants, and how rapidly this assignment changes.⁶⁴ In research involving patients with schizophrenia, nonpsychotic inpatients, and healthy controls, it was found that schizophrenia diagnosis was associated with reduced MMN in stable sequences and P3 in volatile sequences.⁵⁸ DCM linked these reductions to decreased intrinsic connectivity in the left A1 and right IFG. Additionally, symptom severity correlated with changes in frontotemporal connectivity. However, in a study involving undiagnosed individuals with subclinical psychotic-like

experiences, DCM of EEG data did not predict prodromal scores, unlike raw data features based on ERPs.⁵⁹ Finally, another study compared diagnosed schizophrenia patients with nonpsychotic inpatients and healthy controls, using a stochastic oddball paradigm.⁵⁷ Here, both inpatient groups showed decreased intrinsic connectivity in the left A1 and descending frontotemporal connectivity. However, schizophrenia patients differed from nonpsychotic inpatients primarily in cross-hemispheric connections across the cortical hierarchy, which also correlated with psychotic-like symptom severity. It should be noted that these connections were distinct from the connections found in a study from the same group using reversal sequences.⁵⁸ Given such heterogeneous findings across participant cohorts and stimulus sequences, the pattern of connectivity alterations in patients with schizophrenia and related groups remains elusive, and the field would benefit from independent replications of the reported findings.

Decoding Models

In contrast to connectivity models which characterize the interactions between regions, decoding models can reveal the information present in a region's population activity. As a type of MVPA, decoding models allow researchers to exploit the fine-grained multivariate information present in most neuroimaging and electrophysiological data. Decoding and other types of MVPA have helped uncover the content of regional representations of the brain, which adds a functional interpretation of brain

activity.²⁹ Decoding and encoding models can be conceptualized as the inverse of one another. In short, decoding models take data and use it to build a model of representations, whereas encoders begin with the model and use it to predict the data. Decoding models such as support vector machines (SVM), especially in combination with representational similarity analysis (RSA), have proved invaluable in other areas of research, such as perception and learning.⁶⁵ RSA stipulates that stimuli with more distinct neural representations are easier to decode, thus, representational similarity can be indexed by the degree of decodability. The decodability of all possible pairwise combinations of stimuli are compared in a representational dissimilarity matrix (RDM).⁶⁶ In contrast, SVM attempts to identify a reproducible hyperplane that maximizes the distance between two categories.⁶⁷ Despite its success in other areas, MVPA is severely underutilized in MMN and schizophrenia research, both separately and combined. One study assessed MMN in healthy participants using RSA analysis on a roving oddball paradigm.⁶⁸ The researchers found that acoustic features were decodable from the topography of MMRs, although at later latencies than typical MMN. This approach could be highly useful, and easily translated, to a population with schizophrenia. By decoding stimulus features from the MMN and comparing the resulting decoding accuracies between patients with schizophrenia and healthy controls, it may be possible to tell if MMRs are sensitive to different stimulus features in schizophrenia than in healthy controls.

Alternatively, instead of decoding stimulus features from MMRs, MVPA can be applied to MMRs to decode schizophrenia diagnosis (Figure 3A). For instance, a previous study applied SVM to fMRI-derived MMR, comparing patients with schizophrenia to healthy controls in an auditory mismatch task.⁴⁹ SVM was applied to two types of data features; MMR-related brain activation patterns across multiple fMRI voxels, and functional connectivity measured by a correlation analysis across the whole brain. The activation-based features showed 83% participant classification accuracy already within four regions of interest (ROI), while the functional connectivity dataset performed similarly or worse for up to 10 ROIs. However, inclusion of up to 24 ROIs in the functional connectivity reached a maximal accuracy of 90%. In short, the study found that the inclusion of functional connectivity measures across the distributed networks yielded higher classification accuracy. Thus, MVPA methods capitalized on altered functional connectivity to infer the presence of schizophrenia diagnosis, which is consistent with previous research and outlines MMR's potential as a promising diagnostic biomarker of core impairments in schizophrenia.³⁰ Another study⁶⁹ also used a decoding model to assess MMN as a diagnostic biomarker of schizophrenia. Using a similar experimental paradigm, this study measured EEG activity in patients with schizophrenia and healthy controls who listened to different types of oddball sequences. Oddball stimuli were based on three physical aspects: stimulus duration, aural gap, and interstimulus interval. Two separate MVPA models were tested; SVM and

a Gaussian process classifier (GPC). These models elicited accuracy of up to 80%, with the best performing models being generated via the GPC in response to a gap stimulus paradigm. Global functioning scores predicted by the model were shown to have a 73% correlation with true scores, providing additional evidence for MMN responses as a diagnostic biomarker of schizophrenia and symptom severity.

Neural Network Models

As an alternative to SVM and other classical machine learning decoding models, deep neural network (DNN) models are gaining popularity for schizophrenia-related decoding tasks, especially for automatically classifying participants based on their diagnosis. As mentioned above, MMR-related features can be used as a practical basis for classification, with their employment usually increasing the distinguishability between patients and healthy individuals (relative to other data features).^{70,71} This holds for detection based on classical machine learning models^{69,72-78} (see decoding models) as well as DNN models.^{79,80} While DNN models might still rely on MMR-related variables of the underlying neuroimaging data, they do not necessarily rely on the heuristic preprocessing of MMR-related features, since they are usually capable of extracting classification-relevant information directly from the raw data.⁷⁹⁻⁸⁵ Accordingly, DNN models have the advantage of classifying based on intricate and complex markers of schizophrenia which generally yield a higher accuracy of detecting the underlying trait than classical machine learning models,⁷⁰ often scoring as high as 90% to 98%.^{71,80,86-91} However, this results in a decrease in interpretability (a consequence of high parameter count in most DNNs and the resulting "black box" effect). While this accuracy-interpretability tradeoff might be acceptable for classification, interpretability becomes challenging when using DNN as scientific models of the brain.

Using DNN as scientific models is a relatively new but quickly expanding field in neuroscience, which has led to the emergence of the neuroconnectionist research program.⁹² Its main mode of operation is to implement different neural network models (altered in a way to account for a given neuroscientific hypothesis), which are used to encode, decode, or replicate various aspects of empirical neurophysiological responses. This approach can include comparing behavioral, activational, or representational data features, but can also include additional methods like *in silico* lesion studies, which investigate the influence of altering DNN parameters on simulated responses. An overview of the rationale and analysis techniques can be found elsewhere.⁹² The neuroconnectionist research program has mostly gained traction in vision neuroscience.⁹³⁻⁹⁷ However, the same principles can be applied for the investigation of MMR⁹⁸⁻¹⁰¹ and schizophrenia.¹⁰² Cortez-Briones et al⁷¹ give an extensive overview of previous studies using DNN models in schizophrenia research. While to date, there are few studies applying this approach to MMRs, the flexibility of DNN models makes them a very promising method for future

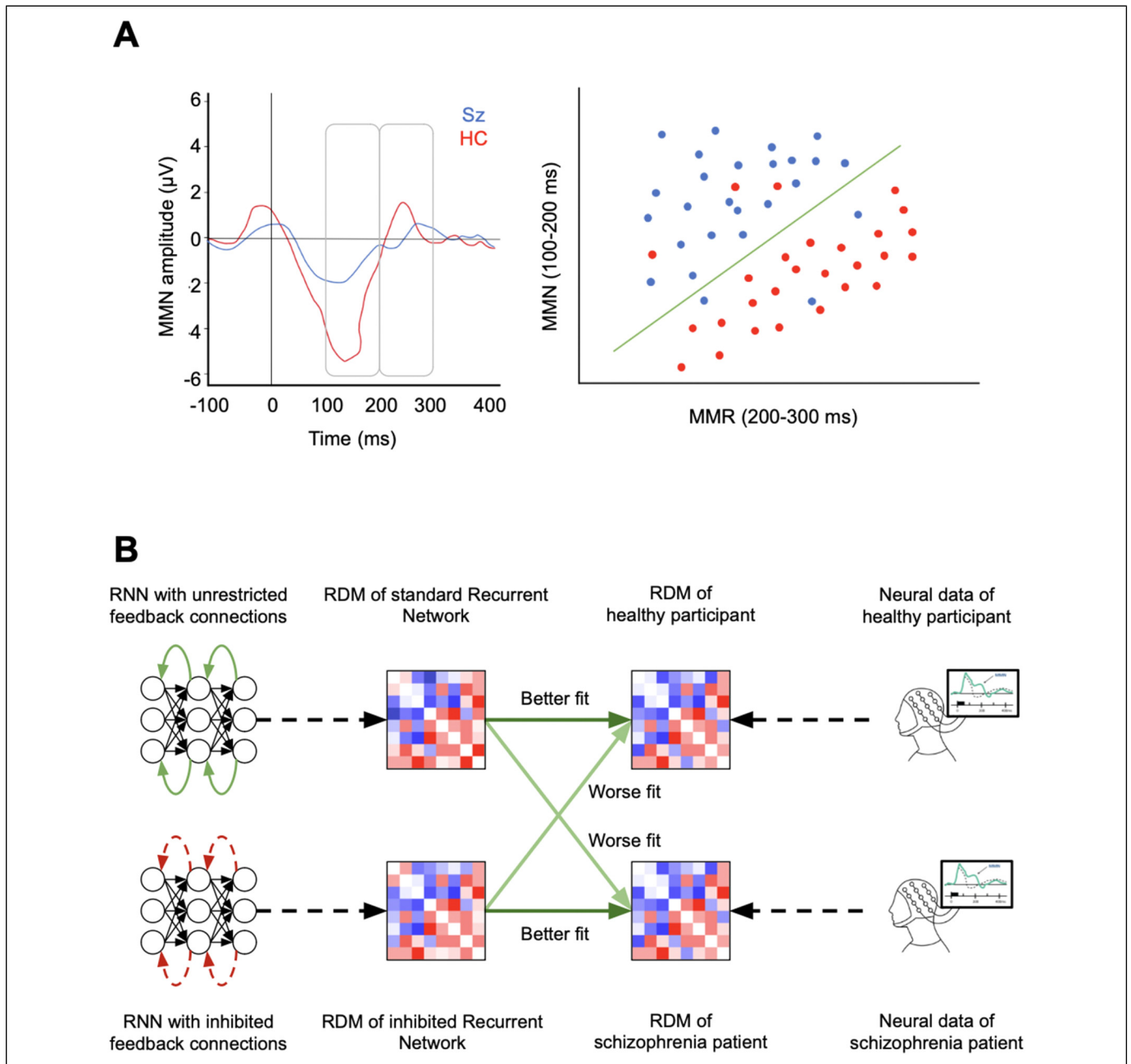


Figure 3. Possible applications of decoding and neural network models. (A) Example of SVM application to decode schizophrenia diagnosis. By applying SVM to a multivariate set of MMR features (eg, EEG amplitudes in the MMR window vs a later time window), it is possible to classify participants based on diagnosis. The SVM creates a hyperplane which separates the data into classes with up to 90-98% accuracy. (B) Possible application of DNN models in MMR/schizophrenia research. Two recurrent neural network (RNN) models are created, with their mechanisms altered in a fashion that represents a given hypothesis in schizophrenia research (In this example: schizophrenia patients exhibit impaired top-down feedback). Then, neuroconnectionist methods (here: RSA) are used to compare the dynamics or representations between the model variations and neurophysiological data from healthy and schizophrenia-diagnosed participants. If the representational dissimilarity matrix (RDM) of the hypothesis-altered model better fits the RDM of schizophrenia patients than the RDM of the standard model (and vice versa for healthy participants), the altered underlying mechanism can be taken as a better model for the corresponding neural mechanisms of schizophrenia.

research. A potential application of DNN models in MMR/schizophrenia research is illustrated in Figure 3B.

Beyond DNNs, other neural network models include attractor-based models, like Hopfield Networks¹⁰³ or spiking

neural networks (SNN).¹⁰⁴ These models aim to simulate certain physical attributes of the brain, and have been employed for neural simulations for several decades. Attractor-based networks have been used to model MMR in general,^{105,106} as well

as in connection with schizophrenia research (for a review, see Ref.¹⁰⁷). As an example, SNNs have been used to simulate neural noise levels and link the alteration of NMDA, gamma-aminobutyric acid (GABA) and DA receptors to increased signal-to-noise ratio and impaired MMR, pointing out similarities to neurophysiological data of schizophrenia patients.¹⁰⁸ The main advantage of using attractor-based models is their depth of biological realism, which is often used to model cell-level dynamics such as excitability and neurotransmitter responses. While until recently, the strength of attractor-based neural networks was usually restricted to purely simulatory approaches, recent advances have enabled the training of these models for decision and perception tasks. This raises the possibility that these types of neural networks, similarly to DNNs, can be used as comprehensive models of cognitive functions and their disruption in neuropsychiatric disorders.^{109,110}

Cognitive Models

In recent years, the study of MMRs has shifted from average-based analyses to single-trial modeling, capturing more intricate brain response dynamics. MMRs such as MMN and P3 have been proposed to reflect a neuronal expression of error or mismatch between current sensory inputs and those predicted under the brain's generative model.^{4,111,112} Thus, MMRs may provide information about the brain's inference process on the environment's statistical structure.³ Given this sensitivity, MMRs should depend on gradual changes in input statistics, necessitating the analysis of trial-by-trial dynamics. Consequently, a growing area of research has employed computational models, particularly Bayesian observer models, to examine single-trial EEG dynamics of MMN in audition,^{113–117} vision^{118,119} and somatosensation,^{10,120,121} and P3 across different senses.^{116,120–128}

Bayesian observer models can incorporate sequential information to infer the probability of new observations (Figure 4A). Given the evidence pointing towards a probabilistic inference-based model of psychosis,^{28,129,130} dissecting various aspects of sequential inference and learning could provide insights into differences between healthy and clinical populations. At the level of probabilistic inference, models might differentiate between quantifying (a) stimulus probability,^{115,122,128} typical for classic oddball paradigms where deviants are defined by low probability^{4,131}; (b) repetition/alternation probability,^{132–134} given the sensitivity of MMN to (un)expected stimulus repetitions,^{2,135,136} which may be disrupted in schizophrenia^{137,138}; and (c) transition probability,^{10,116,121,127,139} typical for Markovian stimulus sequences¹⁴⁰ and considered essential for probabilistic sequence processing.¹³⁹ These probabilistic quantities, estimated by the models, are used with different read-out functions to connect model dynamics to brain data, highlighting different aspects of the inference process reflected in MMRs.¹¹⁵ The read-out functions include (a) novelty detection,^{123,124,127,128} indicating MMR sensitivity to the degree of surprise following new observations^{115,141–143} and quantified using information-theoretical surprise measures¹⁴⁴; (b) belief commitment,^{145–149} whereby probabilistic inference is weighted by prediction confidence, quantified as confidence-corrected surprise^{10,121,150,151}; and (c) model update,^{10,115,120,121,123,128,152} indicating the degree to which new observations trigger adjustments of the generative model,^{115,153,154} quantified as Bayesian surprise.^{155,156} Recent work yields evidence that earlier MMRs (eg, MMN) may be more related to belief commitment, while later MMRs (eg, P3) rather reflect model updates.^{10,121}

A widely used observer model in computational psychiatry^{117,159–165} is the hierarchical Gaussian filter (HGF),^{166,167} which models neural or behavioral responses as PEs scaled by prediction precision (inverse variance of belief distribution).

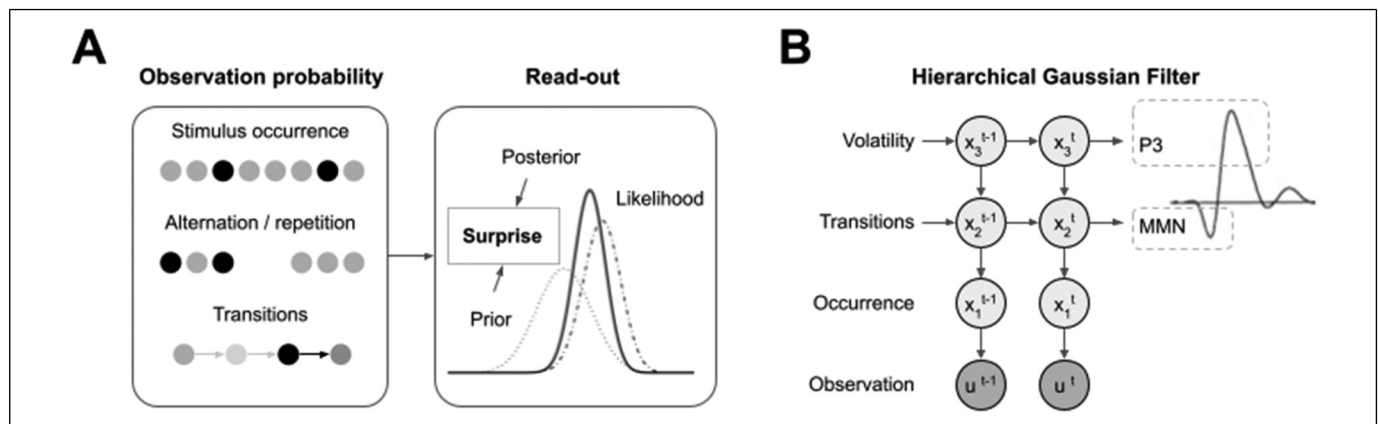


Figure 4. Cognitive models. (A) Observation probabilities include probabilistic quantities related to stimulus occurrence, alternation/repetition, and transitions between stimuli. These probabilistic quantities are subject to different read-out functions based on surprise. (B) Modeling MMRs using the HGF indicated that the MMN and P3 can be mapped onto different hierarchical levels of predictions and PEs. The directed graph shows a typical HGF architecture, tracking probability estimates over time. The highest level relates to volatility estimates and has been linked to the P3, while the lower level relates to transition probability estimates and has been linked to the MMN.^{117,157} Both levels have been shown to be altered in schizophrenia.^{158,159}

These prediction-weighted PEs are used to update predictions of the next observation, with multiple hierarchically organized levels where each level sends predictions to the lower level, and resulting PEs are sent back up the hierarchy. Lower-level PEs typically track stimulus transition probabilities, while higher-level PEs track changes in these probabilities over time. Thus, the HGF excels in modeling probabilistic inference and learning of dynamic stimulus sequences with changing statistical features.^{168,169} The HGF is also versatile, as it can capture behavioral and neural responses across a wide range of tasks,^{117,118,157,159,169–171} and aligns with normative theories of brain function based on probabilistic inference such as hierarchical predictive coding.^{172,173}

The HGF has been used in studies investigating neurocomputational mechanisms mediating schizophrenia and related conditions. In a ketamine study of the roving oddball paradigm,¹¹⁷ stimulus sequences were modeled to generate lower versus higher-level PEs. Similar to findings from surprise-based studies mentioned earlier, this research indicated that MMN reflects lower-level PEs about stimulus transitions, while P3 reflects higher-level PEs regarding their volatility (ie, how fast these transitions change), used to update estimates of environmental statistics (Figure 4B). Importantly, the study found that ketamine reduced high-level PE signaling, suggesting that NMDAR antagonism impairs probabilistic inference related to abstract statistical regularities (but see a recent study showing low-level effects of ketamine¹⁷⁴). Another study, using cholinergic antagonist biperiden and DA antagonist amisulpride (and respective agonists),¹⁵⁷ replicated the finding that lower- versus higher-level PEs can be mapped onto MMN versus P3. Furthermore, the study found that biperiden (compared to placebo) decreased the correlation of EEG amplitudes with low-level PEs, but increased the correlation with high-level (volatility) PEs, suggesting a different effect of DA antagonism than the NMDAR modulation described earlier.¹¹⁷ No other drug effects were identified in this study.

In recent studies, the HGF has been directly applied to investigate schizophrenia.¹⁷⁵ In one study of MMN responses,¹⁵⁸ early-stage patients were compared with individuals at clinical risk and with healthy controls. Both low (sensory) and high-level (volatility) PEs were altered in patients and those at risk, compared to controls. Furthermore, low-level PEs could predict the conversion to psychosis in at-risk individuals. Another study, using fMRI data to model PE signals, involved participants with diagnosed schizophrenia with varying levels of delusions and healthy participants with varying delusional-like ideation.¹⁵⁹ Participants engaged in a reversal learning task with volatile (ie, frequently changing) stimulus contingencies. Delusions (regardless of diagnosis) led to increased precision-weighted PE-related neural activation in frontostriatal regions. In contrast, schizophrenia diagnosis (independent of delusion strength) resulted in overestimated environmental volatility and weaker neural correlates of volatility in the anterior insula, medial frontal, and angular gyrus. This suggests that schizophrenia may be associated with false beliefs regarding

environmental volatility and their impaired neural encoding. However, since patients in the latter study were relatively older than in the former study (mean >30 vs <25 years),¹⁵⁸ model-based correlates of schizophrenia may change during disease progression, moving from lower to higher levels of the processing hierarchy—as previously hypothesized.¹²⁹

Limitations of Statistical and Machine Learning Models in Computational Psychiatry

When using statistical and machine learning tools to infer brain functions it is important to address that these models are usually highly driven by the underlying data and can be subject to overfitting if not validated properly. This is especially true for highly sensitive models, such as nonlinear SVM or DNN. When interpreting or devising studies employing such models, it is therefore crucial to keep in mind that their generalizability only goes so far as the underlying sample of subjects and validation procedure allow for. This is especially relevant in clinical group prediction where the implied claim of a statistical or machine learning approach is to provide criteria that inform on the condition or treatment of new (ie out-of-distribution) subjects or measurements. A good example of this are clinical versus non-clinical group classification tasks: in many such studies, models are fitted and evaluated on a very narrow dataset, often on a subject-level basis or with both training and validation data recorded in the same measurement session. After this, a (usually very high) average decoding accuracy is reported. This average accuracy does not necessarily imply that these models are capable of performing out-of-distribution classification (eg, on new subjects) with a similar accuracy, since the neural and behavioral patterns on which these models are based can vastly differ between subjects and measurements.

Indeed, several studies have shown that many proposed procedures in schizophrenia prediction^{176–178} and other clinical predictions^{179,180} fail to generalize across different contexts, such as across subjects or across time ranges. Accordingly, when using statistical and machine learning approaches for clinical modeling, it is important to ensure the validity of the approach by carefully considering factors such as sample size, sample heterogeneity, temporal stability, as well as employing additional methods such as feature reduction or regularization^{178,181,182} to avoid overfitting the models to limited datasets.

Conclusions and Future Directions

Applying computational models to study the role of MMR in schizophrenia can yield several types of insights. First, connectivity models can help infer the putative neurophysiological mechanisms of MMR reduction (eg, disrupted long-range connectivity and/or local gain control). Since connectivity parameters (based on memory-related fMRI activity) have been

demonstrated to outperform direct measurements of behavioral or neural data in schizophrenia subgroup definition,¹⁸³ adopting a similar approach to MMRs could be equally promising. However, MMR-based connectivity studies so far have yielded heterogeneous results (Figure 2), and thus more systematic replications are required.

Second, decoding models and DNNs can help establish MMRs as diagnostic/risk biomarkers (via classification methods), but also simulate the mechanisms of MMR disruption (via neuroconnectionist approaches). These computational approaches may benefit from a higher sensitivity to subtle neural signatures of mismatch impairment, going beyond more traditional methods such as univariate analyses of MMRs.¹⁸⁴ While few studies have made use of decoding models for the purpose of inferring MMR representational content,⁶⁸ this approach could elucidate the type of sensory features whose processing may be selectively altered in schizophrenia. Furthermore, combining RSA with deep neural networks (Figure 3B) can help infer the network parameters whose disruption would approximate MMR effects found in schizophrenia.

Finally, cognitive models can quantify MMR attenuation in schizophrenia in terms of the underlying computational algorithms (eg, predictive processing), which may help disentangle related MMRs such as the MMN and P3 (Figure 4). Since this approach is relatively recent, it is yet to be extended to at-risk populations and related groups. In future work, cognitive models could also be combined with generative models of neural data to directly link disruptions of trial-by-trial predictive processing to the underlying neural mechanisms. In summary, applying complementary models to empirical data has the potential to elucidate both the pathophysiology and cognitive symptoms of schizophrenia.

Author Contributions

DCG, HHM, and MG contributed to design, drafted manuscript, and gave final approval. RA contributed to conception and design, drafted manuscript, and gave final approval. All authors agree to be accountable for all aspects of work ensuring integrity and accuracy.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: the Deutsche Forschungsgemeinschaft (grant number AU 423/2-1). We acknowledge support by the Open Access Publication Fund of Freie Universität Berlin.

ORCID iDs

Dirk C. Güttlin  <https://orcid.org/0000-0003-3822-896X>
Ryszard Aukstulewicz  <https://orcid.org/0000-0001-9078-3667>

References

- Näätänen R, Gaillard AW, Mäntysalo S. Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol.* 1978;42(4):313-329.
- Aukstulewicz R, Friston K. Repetition suppression and its contextual determinants in predictive coding. *Cortex.* 2016; 80:125-140.
- Garrido MI, Sahani M, Dolan RJ. Outlier responses reflect sensitivity to statistical structure in the human brain. *PLoS Comput Biol.* 2013;9(3):e1002999.
- Garrido MI, Kilner JM, Stephan KE, Friston KJ. The mismatch negativity: a review of underlying mechanisms. *Clin Neurophysiol.* 2009;120(3):453-463.
- Aukstulewicz R, Friston K. Attentional enhancement of auditory mismatch responses: a DCM/MEG study. *Cereb Cortex.* 2015;25(11):4273-4283.
- Sussman ES. A new view on the MMN and attention debate. *J Psychophysiol.* 2007;21(3-4):164-175.
- Sutton S, Braren M, Zubin J, John ER. Evoked-potential correlates of stimulus uncertainty. *Science.* 1965;150(3700): 1187-1188.
- Reed CL, Siqi-Liu A, Lydic K, et al. Selective contributions of executive function ability to the P3. *Int J Psychophysiol.* 2022;176:54-61.
- Schröger E, Marzecová A, SanMiguel I. Attention and prediction in human audition: a lesson from cognitive psychophysiology. *Eur J Neurosci.* 2015;41(5):641-664.
- Gijzen S, Grundei M, Lange RT, Ostwald D, Blankenburg F. Neural surprise in somatosensory Bayesian learning. *PLoS Comput Biol.* 2021;17(2):e1008068.
- Todd J, Salisbury D, Michie PT. Why mismatch negativity continues to hold potential in probing altered brain function in schizophrenia. *Psychiatry Clin Neurosci Rep.* 2023;2(3):e144.
- Todd J, Harms L, Schall U, Michie PT. Mismatch negativity: translating the potential. *Front Psychiatry.* 2013;4:171.
- Castro MK, Bailey DH, Zinger JF, Martin EA. Late electrophysiological potentials and emotion in schizophrenia: a meta-analytic review. *Schizophr Res.* 2019;211:21-31.
- Michie PT, Malmierca MS, Harms L, Todd J. The neurobiology of MMN and implications for schizophrenia. *Biol Psychol.* 2016;116:90-97.
- Balu DT. The NMDA receptor and schizophrenia: from pathophysiology to treatment. *Adv Pharmacol.* 2016;76:351-382.
- Umbricht D, Koller R, Vollenweider FX, Schmid L. Mismatch negativity predicts psychotic experiences induced by NMDA receptor antagonist in healthy volunteers. *Biol Psychiatry.* 2002;51(5):400-406.
- Higley MJ, Picciotto MR. Neuromodulation by acetylcholine: examples from schizophrenia and depression. *Curr Opin Neurobiol.* 2014;29:88-95.
- Kesby JP, Eyles DW, McGrath JJ, Scott JG. Dopamine, psychosis and schizophrenia: the widening gap between basic and clinical neuroscience. *Transl Psychiatry.* 2018;8(1):1-12.
- Moran RJ, Campo P, Symmonds M, Stephan KE, Dolan RJ, Friston KJ. Free energy, precision and learning: the role of cholinergic neuromodulation. *J Neurosci.* 2013;33(19):8227-8236.
- Leung S, Croft RJ, Baldeweg T, Nathan PJ. Acute dopamine D(1) and D(2) receptor stimulation does not modulate mismatch negativity (MMN) in healthy human subjects. *Psychopharmacology.* 2007;194(4):443-451.

21. Leung S, Croft RJ, Guille V, et al. Acute dopamine and/or serotonin depletion does not modulate mismatch negativity (MMN) in healthy human participants. *Psychopharmacology*. 2010; 208(2):233-244.
22. Jahshan C, Cadenhead KS, Rissling AJ, Kirihara K, Braff DL, Light GA. Automatic sensory information processing abnormalities across the illness course of schizophrenia. *Psychol Med*. 2012;42(1):85-97.
23. Näätänen R, Todd J, Schall U. Mismatch negativity (MMN) as biomarker predicting psychosis in clinically at-risk individuals. *Biol Psychol*. 2016;116:36-40.
24. Erickson MA, Ruffle A, Gold JM. A meta-analysis of mismatch negativity in schizophrenia: from clinical risk to disease specificity and progression. *Biol Psychiatry*. 2016;79(12):980-987.
25. Salisbury DF. Mismatch negativity to pitch and duration deviants is not a biomarker of psychosis risk. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2024;9(1):8-9.
26. David Redish A, Gordon JA. *Computational psychiatry: new perspectives on mental illness*. MIT Press; 2022.
27. Anticevic A, Murray JD. *Computational psychiatry: mathematical modeling of mental illness*. Academic Press; 2017.
28. Friston KJ, Stephan KE, Montague R, Dolan RJ. Computational psychiatry: the brain as a phantastic organ. *Lancet Psychiatry*. 2014;1(2):148-158.
29. Kriegeskorte N, Douglas PK. Cognitive computational neuroscience. *Nat Neurosci*. 2018;21(9):1148-1160.
30. Friston K, Brown HR, Siemerkus J, Stephan KE. The dysconnection hypothesis (2016). *Schizophr Res*. 2016;176(2-3):83-94.
31. Friston KJ, Frith CD. Schizophrenia: a disconnection syndrome? *Clin Neurosci*. 1995;3(2):89-97.
32. Chiarion G, Sparacino L, Antonacci Y, Faes L, Mesin L. Connectivity analysis in EEG data: a tutorial review of the state of the art and emerging trends. *Bioengineering*. 2023; 10(3):372.
33. Cao J, Zhao Y, Shan X, et al. Brain functional and effective connectivity based on electroencephalography recordings: a review. *Hum Brain Mapp*. 2022;43(2):860-879.
34. Li B, Razi A, Friston KJ. *Mapping psychopathology with fMRI and effective connectivity analysis*. Frontiers Media SA; 2017.
35. Grundei M, Schmidt TT, Blankenburg F. A multimodal cortical network of sensory expectation violation revealed by fMRI. *Hum Brain Mapp*. 2023;44:5871-5891. doi:10.1002/hbm.26482
36. Friston KJ. Functional and effective connectivity: a review. *Brain Connect*. 2011;1(1):13-36.
37. Kiebel SJ, Garrido MI, Moran RJ, Friston KJ. Dynamic causal modelling for EEG and MEG. *Cogn Neurodyn*. 2008;2(2):121-136.
38. Garrido MI, Friston KJ, Kiebel SJ, Stephan KE, Baldeweg T, Kilner JM. The functional anatomy of the MMN: a DCM study of the roving paradigm. *Neuroimage*. 2008;42(2):936-944.
39. Garrido MI, Kilner JM, Kiebel SJ, Friston KJ. Evoked brain responses are generated by feedback loops. *Proc Natl Acad Sci U S A*. 2007;104(52):20961-20966.
40. Phillips HN, Blenkmann A, Hughes LE, Bekinschtein TA, Rowe JB. Hierarchical organization of frontotemporal networks for the prediction of stimuli across multiple dimensions. *J Neurosci*. 2015;35(25):9255-9264.
41. Phillips HN, Blenkmann A, Hughes LE, et al. Convergent evidence for hierarchical prediction networks from human electrocorticography and magnetoencephalography. *Cortex*. 2016; 82:192-205.
42. Cooray GK, Garrido MI, Brismar T, Hyllienmark L. The maturation of mismatch negativity networks in normal adolescence. *Clin Neurophysiol*. 2016;127(1):520-529.
43. Garrido MI, Kilner JM, Kiebel SJ, Friston KJ. Dynamic causal modeling of the response to frequency deviants. *J Neurophysiol*. 2009;101(5):2620-2631.
44. Schmidt A, Diaconescu AO, Kometer M, Friston KJ, Stephan KE, Vollenweider FX. Modeling ketamine effects on synaptic plasticity during the mismatch negativity. *Cereb Cortex*. 2013; 23(10):2394-2406.
45. Boly M, Garrido MI, Gosseries O, et al. Preserved feedforward but impaired top-down processes in the vegetative state. *Science*. 2011;332(6031):858-862.
46. Shaw AD, Hughes LE, Moran R, Coyle-Gilchrist I, Rittman T, Rowe JB. In vivo assay of cortical microcircuitry in frontotemporal dementia: a platform for experimental medicine studies. *Cereb Cortex*. 2021;31(3):1837-1847.
47. Perrotelli A, Giordano GM, Brando F, Giuliani L, Mucci A. EEG-Based Measures in at-risk mental state and early stages of schizophrenia: a systematic review. *Front Psychiatry*. 2021; 12:653642.
48. Haaf M, Leicht G, Curic S, Mulert C. Glutamatergic deficits in schizophrenia—biomarkers and pharmacological interventions within the ketamine model. *Curr Pharm Biotechnol*. 2018; 19(4):293-307.
49. Gaebler AJ, Mathiak K, Koten JW Jr, et al. Auditory mismatch impairments are characterized by core neural dysfunctions in schizophrenia. *Brain*. 2015;138(Pt 5):1410-1423.
50. Gaebler AJ, Zweerings J, Koten JW, et al. Impaired subcortical detection of auditory changes in schizophrenia but not in major depression. *Schizophr Bull*. 2020;46(1):193-201.
51. Larsen KM, Madsen KS, van Themaat AHVL, et al. Children at familial high risk of schizophrenia and bipolar disorder exhibit altered connectivity patterns during pre-attentive processing of an auditory prediction error. *Schizophr Bull*. 2023;28. doi:10.1093/schbul/sbad092
52. Braeutigam S, Dima D, Frangou S, James A. Dissociable auditory mismatch response and connectivity patterns in adolescents with schizophrenia and adolescents with bipolar disorder with psychosis: a magnetoencephalography study. *Schizophr Res*. 2018;193:313-318.
53. Dima D, Frangou S, Burge L, Braeutigam S, James AC. Abnormal intrinsic and extrinsic connectivity within the magnetic mismatch negativity brain network in schizophrenia: a preliminary study. *Schizophr Res*. 2012;135(1-3):23-27.
54. Larsen KM, Mørup M, Birknow MR, et al. Altered auditory processing and effective connectivity in 22q11.2 deletion syndrome. *Schizophr Res*. 2018;197:328-336.
55. Larsen KM, Mørup M, Birknow MR, et al. Individuals with 22q11.2 deletion syndrome show intact prediction but reduced adaptation in responses to repeated sounds: evidence from Bayesian mapping. *Neuroimage Clin*. 2019;22:101721.
56. Rosch RE, Auksztulewicz R, Leung PD, Friston KJ, Baldeweg T. Selective prefrontal disinhibition in a roving auditory oddball paradigm under N-methyl-D-aspartate receptor blockade. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2019;4(2):140-150.
57. Larsen KM, Dzafic I, Darke H, et al. Aberrant connectivity in auditory precision encoding in schizophrenia spectrum disorder and across the continuum of psychotic-like experiences. *Schizophr Res*. 2020;222:185-194.

58. Dzafic I, Larsen KM, Darke H, et al. Stronger top-down and weaker bottom-up frontotemporal connections during sensory learning are associated with severity of psychotic phenomena. *Schizophr Bull.* 2021;47(4):1039-1047.
59. Taylor JA, Larsen KM, Dzafic I, Garrido MI. Predicting subclinical psychotic-like experiences on a continuum using machine learning. *Neuroimage.* 2021;241:118329.
60. Koshiyama D, Miyakoshi M, Joshi YB, et al. Abnormal effective connectivity underlying auditory mismatch negativity impairments in schizophrenia. *Biol Psychiatry Cogn Neurosci Neuroimaging.* 2020;5(11):1028-1039.
61. Todd J, Howard Z, Auksztulewicz R, Salisbury D. Computational modeling of oddball sequence processing exposes common and differential auditory network changes in first-episode schizophrenia-Spectrum disorders and schizophrenia. *Schizophr Bull.* 2023;49(2):407-416.
62. Ranlund S, Adams RA, Díez Á, et al. Impaired prefrontal synaptic gain in people with psychosis and their relatives during the mismatch negativity. *Hum Brain Mapp.* 2016;37(1):351-365.
63. Adams RA, Pinotsis D, Tsirlis K, et al. Computational modeling of electroencephalography and functional magnetic resonance imaging paradigms indicates a consistent loss of pyramidal cell synaptic gain in schizophrenia. *Biol Psychiatry.* 2022;91(2):202-215.
64. Todd J, Heathcote A, Mullens D, Whitson LR, Provost A, Winkler I. What controls gain in gain control? Mismatch negativity (MMN), priors and system biases. *Brain Topogr.* 2014;27(4):578-589.
65. Liu Y, Nour MM, Schuck NW, Behrens TEJ, Dolan RJ. Decoding cognition from spontaneous neural activity. *Nat Rev Neurosci.* 2022;23(4):204-214.
66. Grootswagers T, Wardle SG, Carlson TA. Decoding dynamic brain patterns from evoked responses: a tutorial on multivariate pattern analysis applied to time series neuroimaging data. *J Cogn Neurosci.* 2017;29(4):677-697.
67. Pisner DA, Schnyer DM. Support vector machine. In: *Machine learning.* Elsevier; 2020: 101-121.
68. An H, Ho Kei S, Auksztulewicz R, Schnupp JWH. Do auditory mismatch responses differ between acoustic features? *Front Hum Neurosci.* 2021;15:613903.
69. Taylor JA, Matthews N, Michie PT, Rosa MJ, Garrido MI. Auditory prediction errors as individual biomarkers of schizophrenia. *Neuroimage Clin.* 2017;15:264-273.
70. Barros C, Silva CA, Pinheiro AP. Advanced EEG-based learning approaches to predict schizophrenia: promises and pitfalls. *Artif Intell Med.* 2021;114:102039.
71. Cortes-Briones JA, Tapia-Rivas NI, Souza D, Estevez DC, A P. Going deep into schizophrenia with artificial intelligence. *Schizophr Res.* 2022;245:122-140.
72. Shim M, Hwang HJ, Kim DW, Lee SH, Im CH. Machine-learning-based diagnosis of schizophrenia using combined sensor-level and source-level EEG features. *Schizophr Res.* 2016;176(2-3):314-319.
73. Santos-Mayo L, San-Jose-Revuelta LM, Arribas JI. A computer-aided diagnosis system with EEG based on the P3b wave during an auditory odd-ball task in schizophrenia. *IEEE Trans Biomed Eng.* 2017;64(2):395-407.
74. Li F, Wang J, Liao Y, et al. Differentiation of schizophrenia by combining the spatial EEG brain network patterns of rest and task P300. *IEEE Trans Neural Syst Rehabil Eng.* 2019;27(4):594-602.
75. Luo Y, Zhang J, Wang C, et al. Discriminating schizophrenia disease progression using a P50 sensory gating task with dense-array EEG, clinical assessments, and cognitive tests. *Expert Rev Neurother.* 2019;19(5):459-470.
76. Chang Q, Liu M, Tian Q, et al. EEG-Based Brain functional connectivity in first-episode schizophrenia patients, ultra-high-risk individuals, and healthy controls during P50 suppression. *Front Hum Neurosci.* 2019;13:379.
77. Devia C, Mayol-Troncoso R, Parrini J, et al. EEG Classification during scene free-viewing for schizophrenia detection. *IEEE Trans Neural Syst Rehabil Eng.* 2019;27(6):1193-1199.
78. Zhang L. EEG Signals classification using machine learning for the identification and diagnosis of schizophrenia. *Conf Proc IEEE Eng Med Biol Soc.* 2019;2019:4521-4524.
79. Ahmedt-Aristizabal D, Fernando T, Denman S, et al. Identification of children at risk of schizophrenia via deep learning and EEG responses. *IEEE J Biomed Health Inform.* 2021;25(1):69-76.
80. Fernando T, Denman S, Ahmedt-Aristizabal D, et al. Neural memory plasticity for medical anomaly detection. *Neural Netw.* 2020;127:67-81.
81. Oh K, Kim W, Shen G, et al. Classification of schizophrenia and normal controls using 3D convolutional neural network and outcome visualization. *Schizophr Res.* 2019;212:186-195.
82. Oh J, Oh BL, Lee KU, Chae JH, Yun K. Identifying schizophrenia using structural MRI with a deep learning algorithm. *Front Psychiatry.* 2020;11:16.
83. Matsubara T, Tashiro T, Uehara K. Deep neural generative model of functional MRI images for psychiatric disorder diagnosis. *IEEE Trans Biomed Eng.* 2019;66(10):2768-2779.
84. Dakka J, Bashivan P, Gheiratmand M, Rish I, Jha S, Greiner R. Learning neural markers of schizophrenia disorder using recurrent neural networks. Published online 2017. doi:10.48550/ARXIV.1712.00512
85. Salvador R, Canales-Rodríguez E, Guerrero-Pedraza A, et al. Multimodal integration of brain images for MRI-based diagnosis in schizophrenia. *Front Neurosci.* 2019;13:1203.
86. Li G, Han D, Wang C, Hu W, Calhoun VD, Wang YP. Application of deep canonically correlated sparse autoencoder for the classification of schizophrenia. *Comput Methods Programs Biomed.* 2020;183:105073.
87. Patel P, Aggarwal P, Gupta A. Classification of Schizophrenia versus normal subjects using deep learning. In: Proceedings of the tenth Indian conference on computer vision, graphics and image processing. ACM; 2016. doi:10.1145/3009977.3010050
88. Niu YW, Lin QH, Qiu Y, Kuang LD, Calhoun VD. Sample augmentation for classification of schizophrenia patients and healthy controls using ICA of fMRI data and convolutional neural networks. In: 2019 tenth international conference on intelligent control and information processing (ICICIP). IEEE; 2019. doi:10.1109/icicip47338.2019.9012169
89. Qureshi MNI, Oh J, Lee B. 3D-CNN Based discrimination of schizophrenia using resting-state fMRI. *Artif Intell Med.* 2019;98:10-17.
90. Calhas D, Romero E, Henriques R. On the use of pairwise distance learning for brain signal classification with limited observations. *Artif Intell Med.* 2020;105:101852.
91. Phang CR, Noman F, Hussain H, Ting CM, Ombao H. A multi-domain connectome convolutional neural network for identifying schizophrenia from EEG connectivity patterns. *IEEE J Biomed Health Inform.* 2020;24(5):1333-1343.

92. Doerig A, Sommers RP, Seeliger K, et al. The neuroconnectionist research programme. *Nat Rev Neurosci.* 2023;24(7):431-450.
93. Kietzmann TC, Spoerer CJ, Sörensen LKA, Cichy RM, Hauk O, Kriegeskorte N. Recurrence is required to capture the representational dynamics of the human visual system. *Proc Natl Acad Sci U S A.* 2019;116(43):21854-21863.
94. Güçlü U, van Gerven MAJ. Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *J Neurosci.* 2015;35(27):10005-10014.
95. Cichy RM, Khosla A, Pantazis D, Torralba A, Oliva A. Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Sci Rep.* 2016;6:27755.
96. Lindsay GW. Convolutional neural networks as a model of the visual system: past, present, and future. *J Cogn Neurosci.* 2021;33(10):2017-2031.
97. Kriegeskorte N. Deep neural networks: a new framework for modeling biological vision and brain information processing. *Annu Rev Vis Sci.* 2015;1:417-446.
98. Ali A, Ahmad N, de Groot E, van Gerven J, Kietzmann MA, C T. Predictive coding is a consequence of energy efficiency in recurrent neural networks. *Patterns (N Y).* 2022;3(12):100639.
99. O'Reilly JA, Angsuwatanakul T, Wehrman J. Decoding violated sensory expectations from the auditory cortex of anaesthetised mice: hierarchical recurrent neural network depicts separate "danger" and "safety" units. *Eur J Neurosci.* 2022;56(3):4154-4175.
100. O'Reilly JA. Recurrent neural network model of human event-related potentials in response to intensity oddball stimulation. *Neuroscience.* 2022;504:63-74.
101. Shahnazian D, Holroyd CB. Distributed representations of action sequences in anterior cingulate cortex: a recurrent neural network approach. *Psychon Bull Rev.* 2018;25(1):302-321.
102. Lanillos P, Oliva D, Philippsen A, Yamashita Y, Nagai Y, Cheng G. A review on neural network models of schizophrenia and autism spectrum disorder. *Neural Netw.* 2020;122:338-363.
103. Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci U S A.* 1982;79(8):2554-2558.
104. Ghosh-Dastidar S, Adeli H. Spiking neural networks. *Int J Neural Syst.* 2009;19(4):295-308.
105. Wacongne C. A predictive coding account of MMN reduction in schizophrenia. *Biol Psychol.* 2016;116:68-74.
106. Wacongne C, Changeux JP, Dehaene S. A neuronal model of predictive coding accounting for the mismatch negativity. *J Neurosci.* 2012;32(11):3665-3678.
107. Valton V, Romaniuk L, Douglas Steele J, Lawrie S, Seriès P. Comprehensive review: computational modelling of schizophrenia. *Neurosci Biobehav Rev.* 2017;83:631-646.
108. Rolls ET, Loh M, Deco G, Winterer G. Computational models of schizophrenia and dopamine modulation in the prefrontal cortex. *Nat Rev Neurosci.* 2008;9(9):696-709.
109. Yamazaki K, Vo-Ho VK, Bulsara D, Le N. Spiking neural networks and their applications: a review. *Brain Sci.* 2022;12(7):863. doi:10.3390/brainsci12070863
110. Basu A, Deng L, Frenkel C, Zhang X. Spiking neural network integrated circuits: a review of trends and future directions. In: 2022 IEEE custom integrated circuits conference (CICC). IEEE; 2022. doi:10.1109/cicc53496.2022.9772783
111. Friston K. A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci.* 2005;360(1456):815-836.
112. Friston K. The free-energy principle: a unified brain theory? *Nat Rev Neurosci.* 2010;11(2):127-138.
113. Lecaigard F, Bertrand R, Brunner P, Caclin A, Schalk G, Mattout J. Dynamics of oddball sound processing: trial-by-trial modeling of ECoG signals. *Front Hum Neurosci.* 2021;15:794654.
114. Lecaigard F, Bertrand O, Caclin A, Mattout J. Neurocomputational underpinnings of expected surprise. *J Neurosci.* 2022;42(3):474-486.
115. Lieder F, Daunizeau J, Garrido MI, Friston KJ, Stephan KE. Modelling trial-by-trial changes in the mismatch negativity. *PLoS Comput Biol.* 2013;9(2):e1002911.
116. Maheu M, Dehaene S, Meyniel F. Brain signatures of a multi-scale process of sequence learning in humans. *Elife.* 2019;8. doi:10.7554/eLife.41541
117. Weber LA, Diaconescu AO, Mathys C, et al. Ketamine affects prediction errors about statistical regularities: a computational single-trial analysis of the mismatch negativity. *J Neurosci.* 2020;40(29):5658-5668.
118. Stefanics G, Heinzle J, Horváth AA, Stephan KE. Visual mismatch and predictive coding: a computational single-trial ERP study. *J Neurosci.* 2018;38(16):4020-4030.
119. Schlossmacher I, Lucka F, Peters A, Bruchmann M, Straube T. Effects of awareness and task relevance on neurocomputational models of mismatch negativity generation. *Neuroimage.* 2022;262:119530.
120. Ostwald D, Spitzer B, Guggenmos M, Schmidt TT, Kiebel SJ, Blankenburg F. Evidence for neural encoding of Bayesian surprise in human somatosensation. *Neuroimage.* 2012;62(1):177-188.
121. Grundei M, Schröder P, Gijzen S, Blankenburg F. EEG mismatch responses in a multimodal roving stimulus paradigm provide evidence for probabilistic inference across audition, somatosensation, and vision. *Hum Brain Mapp.* 2023;44(9):3644-3668.
122. Kolossa A, Fingscheidt T, Wessel K, Kopp B. A model-based approach to trial-by-trial p300 amplitude fluctuations. *Front Hum Neurosci.* 2012;6:359.
123. Kolossa A, Kopp B, Fingscheidt T. A computational analysis of the neural bases of Bayesian inference. *Neuroimage.* 2015;106:222-237.
124. Kopp B, Seer C, Lange F, et al. P300 amplitude variations, prior probabilities, and likelihoods: a Bayesian ERP study. *Cogn Affect Behav Neurosci.* 2016;16(5):911-928.
125. Wessel JR, Huber DE. Frontal cortex tracks surprise separately for different sensory modalities but engages a common inhibitory control mechanism. *PLoS Comput Biol.* 2019;15(7):e1006927.
126. MacPherson SE, Gillebert CR, Robinson GA, Vallesi A. *Intra- and inter-individual variability of executive functions: determinant and modulating factors in healthy and pathological conditions.* Frontiers Media SA; 2019.
127. Modirshanechi A, Kiani MM, Aghajani H. Trial-by-trial surprise-decoding model for visual and auditory binary oddball tasks. *Neuroimage.* 2019;196:302-317.
128. Mars RB, Debener S, Gladwin TE, et al. Trial-by-trial fluctuations in the event-related electroencephalogram reflect dynamic changes in the degree of surprise. *J Neurosci.* 2008;28(47):12539-12545.
129. Fletcher PC, Frith CD. Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nat Rev Neurosci.* 2009;10(1):48-58.

130. Sterzer P, Adams RA, Fletcher P, et al. The predictive coding account of psychosis. *Biol Psychiatry*. 2018;84(9):634-643.
131. Näätänen R, Paavilainen P, Rinne T, Alho K. The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin Neurophysiol*. 2007;118(12):2544-2590.
132. Summerfield C, Trittschuh EH, Monti JM, Mesulam MM, Egner T. Neural repetition suppression reflects fulfilled perceptual expectations. *Nat Neurosci*. 2008;11(9):1004-1006.
133. Summerfield C, Wyart V, Johnen VM, de Gardelle V. Human scalp electroencephalography reveals that repetition suppression varies with expectation. *Front Hum Neurosci*. 2011;5:67.
134. de Gardelle V, Waszczuk M, Egner T, Summerfield C. Concurrent repetition enhancement and suppression responses in extrastriate visual cortex. *Cereb Cortex*. 2013;23(9):2235-2244.
135. Tervaniemi M, Maury S, Näätänen R. Neural representations of abstract stimulus features in the human brain as reflected by the mismatch negativity. *Neuroreport*. 1994;5(7):844-846.
136. Czigler I, Weisz J, Winkler I. ERPs and deviance detection: visual mismatch negativity to repeated visual stimuli. *Neurosci Lett*. 2006;401(1-2):178-182.
137. Baldeweg T. Repetition effects to sounds: evidence for predictive coding in the auditory system. *Trends Cogn Sci*. 2006;10(3):93-94.
138. Baldeweg T, Klugman A, Gruzelier J, Hirsch SR. Mismatch negativity potentials and cognitive impairment in schizophrenia. *Schizophr Res*. 2004;69(2-3):203-217.
139. Meyniel F, Maheu M, Dehaene S. Human inferences about sequences: a minimal transition probability model. *PLoS Comput Biol*. 2016;12(12):e1005260.
140. Schröger E, Roeber U, Coy N. Markov chains as a proxy for the predictive memory representations underlying mismatch negativity. *Front Hum Neurosci*. 2023;17. doi:10.3389/fnhum.2023.1249413
141. Squires NK, Squires KC, Hillyard SA. Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalogr Clin Neurophysiol*. 1975;38(4):387-401.
142. Escera C, Corral MJ. Role of mismatch negativity and novelty-P3 in involuntary auditory attention. *J Psychophysiol*. 2007;21(3-4):251-264.
143. Tiitinen H, May P, Reinikainen K, Näätänen R. Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature*. 1994;372(6501):90-92.
144. Shannon CE. *A mathematical theory of communication*; 1948.
145. Büchel C, Geuter S, Sprenger C, Eippert F. Placebo analgesia: a predictive coding perspective. *Neuron*. 2014;81(6):1223-1239.
146. Meyniel F, Schlunegger D, Dehaene S. The sense of confidence during probabilistic learning: a normative account. *PLoS Comput Biol*. 2015;11(6):e1004305.
147. Meyniel F, Dehaene S. Brain networks for confidence weighting and hierarchical inference during probabilistic learning. *Proc Natl Acad Sci U S A*. 2017;114(19):E3859-E3868.
148. Meyniel F. Brain dynamics for confidence-weighted learning. *PLoS Comput Biol*. 2020;16(6):e1007935.
149. Mulders D, Seymour B, Mouraux A, Mancini F. Confidence of probabilistic predictions modulates the cortical response to pain. *Proc Natl Acad Sci U S A*. 2023;120(4):e2212252120.
150. Faraji M, Preuschoff K, Gerstner W. Balancing new against old information: the role of puzzlement surprise in learning. *Neural Comput*. 2018;30(1):34-83.
151. Modirshanechi A, Brea J, Gerstner W. A taxonomy of surprise definitions. *J Math Psychol*. 2022;110(102712):102712.
152. Seer C, Lange F, Boos M, Dengler R, Kopp B. Prior probabilities modulate cortical surprise responses: a study of event-related potentials. *Brain Cogn*. 2016;106:78-89.
153. Winkler I. Interpreting the mismatch negativity. *J Psychophysiol*. 2007;21(3-4):147-163.
154. Winkler I, Denham SL, Nelken I. Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends Cogn Sci*. 2009;13(12):532-540.
155. Itti L, Baldi P. Bayesian Surprise attracts human attention. *Vision Res*. 2009;49(10):1295-1306.
156. Baldi P, Itti L. Of bits and wows: a Bayesian theory of surprise with applications to attention. *Neural Netw*. 2010;23(5):649-666.
157. Weber LA, Tomiello S, Schöbi D, et al. Auditory mismatch responses are differentially sensitive to changes in muscarinic acetylcholine versus dopamine receptor function. *Elife*. 2022;11. doi:10.7554/eLife.74835
158. Hauke DJ, Charlton CE, Schmidt A, et al. Aberrant hierarchical prediction errors are associated with transition to psychosis: a computational single-trial analysis of the mismatch negativity. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2023;8(12):1176-1185. doi:10.1016/j.bpsc.2023.07.011
159. Fromm S, Katthagen T, Deserno L, Heinz A, Kaminski J, Schlagenhauf F. Belief updating in subclinical and clinical delusions. *Schizophr Bull Open*. 2022;14. doi:10.1093/schizbullopen/sgac074
160. Lawson RP, Mathys C, Rees G. Adults with autism overestimate the volatility of the sensory environment. *Nat Neurosci*. 2017;20(9):1293-1299.
161. Sevgi M, Diaconescu AO, Henco L, Tittgemeyer M, Schilbach L. Social Bayes: using Bayesian modeling to study autistic trait-related differences in social cognition. *Biol Psychiatry*. 2020;87(2):185-193.
162. Sapey-Triomphe LA, Weilhhammer VA, Wagemans J. Associative learning under uncertainty in adults with autism: intact learning of the cue-outcome contingency, but slower updating of priors. *Autism*. 2022;26(5):1216-1228.
163. Siegel JZ, Curwell-Parry O, Saunders PS, Crockett KEA, J M. A computational phenotype of disrupted moral inference in borderline personality disorder. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2020;5(12):1134-1141.
164. Rossi-Goldthorpe RA, Leong YC, Leptourgos P, Corlett PR. Paranoia, self-deception and overconfidence. *PLoS Comput Biol*. 2021;17(10):e1009453.
165. Suthaharan P, Reed EJ, Leptourgos P, et al. Paranoia and belief updating during the COVID-19 crisis. *Nat Hum Behav*. 2021;5(9):1190-1202.
166. Mathys C, Daunizeau J, Friston KJ, Stephan KE. A Bayesian foundation for individual learning under uncertainty. *Front Hum Neurosci*. 2011;5:39.
167. Mathys CD, Lomakina EI, Daunizeau J, et al. Uncertainty in perception and the hierarchical Gaussian filter. *Front Hum Neurosci*. 2014;8:825.
168. Fitzgerald K, Auksztulewicz R, Provost A, Paton B, Howard Z, Todd J. Hierarchical learning of statistical regularities over multiple timescales of sound sequence processing: a dynamic causal modeling study. *J Cogn Neurosci*. 2021;33(8):1549-1562.
169. Iglesias S, Mathys C, Brodersen KH, et al. Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. *Neuron*. 2019;101(6):1196-1201.

170. Aukstulewicz R, Friston KJ, Nobre AC. Task relevance modulates the behavioural and neural effects of sensory predictions. *PLoS Biol.* 2017;15(12):e2003143.
171. Palmer CE, Aukstulewicz R, Ondobaka S, Kilner JM. Sensorimotor beta power reflects the precision-weighting afforded to sensory prediction errors. *Neuroimage.* 2019;200:59-71.
172. Friston K, Kiebel S. Predictive coding under the free-energy principle. *Philos Trans R Soc Lond B Biol Sci.* 2009;364(1521):1211-1221.
173. Weber LA, Waade PT, Legrand N, Møller AH, Stephan KE, Mathys C. The generalized Hierarchical Gaussian Filter. *arXiv.* 2023;2305(10937).
174. Wehrman JJ, Casey C, Tanabe S, et al. Subanaesthetic doses of ketamine reduce, but do not eliminate, predictive coding responses: implications for mechanisms of sensory disconnection. *Br J Anaesth.* 2023;131(4):705-714. doi:10.1016/j.bja.2023.06.044
175. Sheffield JM, Suthaharan P, Leptourgos P, Corlett PR. Belief updating and paranoia in individuals with schizophrenia. *Biol Psychiatry Cogn Neurosci Neuroimaging.* 2022;7(11):1149-1157.
176. Chekroud AM, Hawrilenko M, Loho H, et al. Illusory generalizability of clinical prediction models. *Science.* 2024;383(6679):164-167.
177. Petzschner FH. Practical challenges for precision medicine. *Science.* 2024;383(6679):149-150.
178. Schnack HG, Kahn RS. Detecting neuroimaging biomarkers for psychiatric disorders: sample size matters. *Front Psychiatry.* 2016;7:50.
179. Winter NR, Blanke J, Leenings R, et al. A systematic evaluation of machine learning-based biomarkers for major depressive disorder. *JAMA Psychiatry.* 2024;10. doi:10.1001/jamapsychiatry.2023.5083
180. Wolfers T, Buitelaar JK, Beckmann CF, Franke B, Marquand AF. From estimating activation locality to predicting disorder: a review of pattern recognition for neuroimaging-based psychiatric diagnostics. *Neurosci Biobehav Rev.* 2015;57:328-349.
181. Kernbach JM, Staartjes VE. Foundations of machine learning-based clinical prediction modeling: Part II-generalization and overfitting. *Acta Neurochir Suppl.* 2022;134:15-21.
182. Varoquaux G, Raamana PR, Engemann DA, Hoyos-Idrobo A, Schwartz Y, Thirion B. Assessing and tuning brain decoders: cross-validation, caveats, and guidelines. *Neuroimage.* 2017;145(Pt B):166-179.
183. Brodersen KH, Deserno L, Schlagenhaut F, et al. Dissecting psychiatric spectrum disorders by generative embedding. *Neuroimage Clin.* 2014;4:98-111.
184. Dheerendra P, Grent-'t-Jong T, Gajwani R, et al. Intact mismatch negativity responses in clinical high risk for psychosis and first-episode psychosis: evidence from source-reconstructed event-related fields and time-frequency data. *Biol Psychiatry Cogn Neurosci Neuroimaging.* 2024;9(1):121-131.