Mismatch responses: Probing probabilistic inference in the brain

Dissertation

Zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften (Dr. rer. nat.)

am Fachbereich Erziehungswissenschaft und Psychologie
der Freien Universität Berlin

vorgelegt von

Miro Grundei, M. Sc.

Berlin, 2023

Gutachter*innen:

1. Prof. Dr. Felix Blankenburg
2. Prof. Dr. Arno Villringer
3. Prof. Dr. Uta Noppeney

Datum der Disputation: 22.11.2023

Acknowledgements

I would like to thank my advisors, colleagues, friends and family who provided the support without which this thesis would not have been possible.

First and foremost, I would like to thank my supervisor Felix Blankenburg for his immense support and helpful guidance during the years as a PhD student at the Neurocomputation and Neuroimaging Unit (NNU). He provided an intellectually stimulating and friendly working environment which allowed me to develop my own ideas and pursue my interests. Further, I am thankful for the ongoing support of Arno Villringer during the co-supervision of my PhD and for agreeing to examine this dissertation. I also thank Annette Kinder, Till Nierhaus und Ryszard Auksztulewicz for joining the doctoral committee. Moreover, I am very grateful for the financial support provided by the Berlin School of Mind and Brain as well as their academic infrastructure.

Thank you to all the members of the NNU who I have met and worked with over the years, to Alex, Christian, Daniela, Dirk, Evgeniya, Gian, Guido, Isil, Jo, Jona, Lisa, Marlon, Marianna, Pia, Robert, Ryszard, Sara, Till, Timo, Yuan-Hao and especially Sam, for discussing all the various challenges of our PhD research and for making the work in the lab so enjoyable.

Finally, I am very grateful to my wonderful friends who are always there to support me, especially Juli, Julius and Grischa, and most of all to Anna, for her contagious positivity, inspiration and love. And to my parents, Valeska and Martin, for supporting me along my journey.

Table of contents

Abstract

Sensory signals are governed by statistical regularities and carry valuable information about the unfolding of environmental events. The brain is thought to capitalize on the probabilistic nature of sequential inputs to infer on the underlying (hidden) dynamics driving sensory stimulation. Mismatch responses (MMRs) such as the mismatch negativity (MMN) and the P3 constitute prominent neuronal signatures which are increasingly interpreted as reflecting a mismatch between the current sensory input and the brain's generative model of incoming stimuli. As such, MMRs might be viewed as signatures of probabilistic inference in the brain and their response dynamics can provide insights into the underlying computational principles. However, given the dominance of the auditory modality in MMR research, the specifics of brain responses to probabilistic sequences across sensory modalities and especially in the somatosensory domain are not well characterized.

The work presented here investigates MMRs across the auditory, visual and somatosensory modality by means of electroencephalography (EEG) and functional magnetic resonance imaging (fMRI). We designed probabilistic stimulus sequences to elicit and characterize MMRs and employed computational modeling of response dynamics to inspect different aspects of the brain's generative model of the sensory environment. In the first study, we used a volatile roving stimulus paradigm to elicit somatosensory MMRs and performed single-trial modeling of EEG signals in sensor and source space. Model comparison suggested that responses reflect Bayesian inference based on the estimation of transition probability and limited information integration of the recent past in order to adapt to a changing environment. The results indicated that somatosensory MMRs reflect an initial mismatch between sensory input and model beliefs represented by confidence-corrected surprise (CS) followed by model adjustment dynamics represented by Bayesian surprise (BS). For the second and third study we designed a tri-modal roving stimulus paradigm to delineate modality specific and modality general features of mismatch processing. Computational modeling of EEG signals in study 2 suggested that single-trial dynamics reflect Bayesian inference based on estimation of uni-modal transition probabilities as well as cross-modal conditional dependencies. While early mismatch processing around the MMN tended to reflect CS, later MMRs around the P3 rather reflect BS, in correspondence to the somatosensory study. Finally, the fMRI results of study 3 showed that MMRs are generated by an interaction of modality specific regions in higher order sensory cortices and a modality general fronto-parietal network. Inferior parietal regions in particular were sensitive to expectation violations with respect to the cross-modal contingencies in the stimulus sequences. Overall, our results indicate that MMRs across the senses reflect processes of probabilistic inference in a complex and inherently multi-modal environment.

Zusammenfassung

Sensorische Signale sind durch statistische Regularitäten bestimmt und beinhalten wertvolle Informationen über die Entwicklung von Umweltereignissen. Es wird angenommen, dass das Gehirn die Wahrscheinlichkeitseigenschaften sequenzieller Reize nutzt um auf die zugrundeliegenden (verborgenen) Dynamiken zu schließen, welche sensorische Stimulation verursachen. Diskrepanz-Reaktionen ("Mismatch responses"; MMRs) wie die "mismatch negativity" (MMN) und die P3 sind bekannte neuronale Signaturen die vermehrt als Signale einer Diskrepanz zwischen der momentanen sensorischen Einspeisung und dem generativen Modell, welches das Gehirn von den eingehenden Reizen erstellt angesehen werden. Als solche können MMRs als Signaturen von wahrscheinlichkeitsbasierter Inferenz im Gehirn betrachtet werden und ihre Reaktionsdynamiken können Einblicke in die zugrundeliegenden komputationalen Prinzipien geben. Angesichts der Dominanz der auditorischen Modalität in der MMR-Forschung, sind allerdings die spezifischen Eigenschaften von Hirn-Reaktionen auf Wahrscheinlichkeitssequenzen über sensorische Modalitäten hinweg und vor allem in der somatosensorischen Modalität nicht gut charakterisiert.

Die hier vorgestellte Arbeit untersucht MMRs über die auditorische, visuelle und somatosensorische Modalität hinweg anhand von Elektroenzephalographie (EEG) und funktioneller Magnetresonanztomographie (fMRT). Wir gestalteten wahrscheinlichkeitsbasierte Reizsequenzen, um MMRs auszulösen und zu charakterisieren und verwendeten komputationale Modellierung der Reaktionsdynamiken, um verschiedene Aspekte des generativen Modells des Gehirns von der sensorischen Umwelt zu untersuchen. In der ersten Studie verwendeten wir ein volatiles "Roving-Stimulus"-Paradigma, um somatosensorische MMRs auszulösen und modellierten die Einzel-Proben der EEG-Signale im sensorischen und Quell-Raum. Modellvergleiche legten nahe, dass die Reaktionen Bayes'sche Inferenz abbilden, basierend auf der Schätzung von Transitionswahrscheinlichkeiten und limitierter Integration von Information der jüngsten Vergangenheit, welche eine Anpassung an Umweltänderungen ermöglicht. Die Ergebnisse legen nahe, dass somatosensorische MMRs eine initiale Diskrepanz zwischen sensorischer Einspeisung und Modellüberzeugung reflektieren welche durch "confidence-corrected surprise" (CS) repräsentiert ist, gefolgt von Modelanpassungsdynamiken repräsentiert von "Bayesian surprise" (BS). Für die zweite und dritte Studie haben wir ein Tri-Modales "Roving-Stimulus"-Paradigma gestaltet, um modalitätsspezifische und modalitätsübergreifende Eigenschaften von Diskrepanzprozessierung zu umreißen. Komputationale Modellierung von EEG-Signalen in Studie 2 legte nahe, dass Einzel-Proben Dynamiken Bayes'sche Inferenz abbilden, basierend auf der Schätzung von unimodalen Transitionswahrscheinlichkeiten sowie modalitätsübergreifenden bedingten Abhängigkeiten. Während frühe

Diskrepanzprozessierung um die MMN dazu tendierten CS zu reflektieren, so reflektierten spätere MMRs um die P3 eher BS, in Übereinstimmung mit der somatosensorischen Studie. Abschließend zeigten die fMRT-Ergebnisse der Studie 3 dass MMRs durch eine Interaktion von modalitätsspezifischen Regionen in sensorischen Kortizes höherer Ordnung mit einem modalitätsübergreifenden fronto-parietalen Netzwerk generiert werden. Inferior parietale Regionen im Speziellen waren sensitiv gegenüber Erwartungsverstoß in Bezug auf die modalitätsübergreifenden Wahrscheinlichkeiten in den Reizsequenzen. Insgesamt weisen unsere Ergebnisse darauf hin, dass MMRs über die Sinne hinweg Prozesse von wahrscheinlichkeitsbasierter Inferenz in einer komplexen und inhärent multi-modalen Umwelt darstellen.

List of abbreviations

ACC anterior cingulate cortex

AP alternation probability

ASA auditory scene analysis

BB Beta-Bernoulli (model)

BCI Bayesian causal inference

BMS Bayesian model selection

BS Bayesian surprise

CM cross-modal (inference model)

CS confidence-corrected surprise

DC Derichlet-Categorical (model)

DCM dynamic causal modeling

EEG electroencephalography

ERP event-related potentials

fMRI functional magnetic resonance imaging

GLM general linear model

HMM Hidden Markov model

IC insular cortex

IFG inferior frontal gyrus

IFJ inferior frontal junction

IPS inferior parietal sulcus

ISI inter stimulus interval

IT inferior temporal (cortex)

KL Kullback-Leibler (divergence)

LOC lateral occipital cortex

MEG magnetoencephalography

MMN mismatch negativity

MMR mismatch response

OP opercular cortex

PC predictive coding

PPI psycho-physiological interaction

PS predictive surprise

RS repetition suppression

SI primary somatosensory cortex

SII secondary somatosensory cortex

SMA supplementary motor area

sMMN somatosensory mismatch negativity

SP stimulus probability

STG superior temporal gyrus

STS superior temporal sulcus

TL train length

TP transition probability

TPJ temporo-parietal junction

UM uni-modal (inference model)

vMMN visual mismatch negativity

List of original research articles

Study 1

Gijsen*, S., **Grundei*, M.**, Lange, R. T., Ostwald, D. and Blankenburg, F. (2021). Neural surprise in somatosensory Bayesian learning. *PLoS computational biology, 17(2)*, e1008068.
*equal contribution

Study 2

Grundei, M., Schröder, P., Gijsen, S. and Blankenburg, F. (2023). EEG mismatch responses in a multimodal roving stimulus paradigm provide evidence for probabilistic inference across audition, somatosensation and vision. *Human Brain Mapping*, 1– 25.

Study 3

Grundei, M., Schmidt T. T., and Blankenburg, F. (2023). A multimodal cortical network of sensory expectation violation revealed by fMRI. *Human Brain Mapping, 44*(17), 5871-5891.

CHAPTER 1

Introduction

Our brain is constantly subjected to a vast amount of inputs acquired simultaneously by the sensory organs providing the interface to the world surrounding us. The sensory environment is highly structured and governed by statistical regularities at various scales (Ruderman and Bialek, 1994; Nelken et al., 1999; Bar, 2004; Bregman, 1994). Sequential sensory inputs thus contain valuable information about the probabilistic nature of environmental events (Dehaene et al., 2015; Armstrong et al., 2017; Santolin and Saffran, 2018; Sherman et al., 2020) which, in turn, can provide imperatives for the organism to successfully navigate the physical world (Lashley, 1951; Reber and Lewis, 1977; Sherman et al., 2020). One might consider the sensation of a raindrop on the skin before imminent rainfall or the sound of thunder seconds after observing a bright flash in the sky. The unfolding of multi-modal sensory events over time can be predictive of upcoming sensations within and across modalities, like looming clouds and increasing winds predicting the audio-visuo-tactile experience of a thunderstorm. Our brain possesses the remarkable capacity to capitalize on statistical regularities for behaviour and cognition (Sherman et al., 2020), a feat which is shared across species (Santolin and Saffran, 2018; Fitch and Hauser, 2004), develops at early infancy (Saffran et al., 1996; Perruchet and Pacton, 2006; Aslin and Newport, 2012) and provides the basis for language learning in humans (Chomsky, 2002; Hauser et al., 2002; Dehaene et al., 2015).

The extraction of regularities from the environment enables the brain to form expectations of future events (Winkler et al., 2009) which can result in more efficient processing of sensory input (Summerfield and de Lange, 2014; Ekman et al., 2017; Melloni et al., 2011; de Lange et al., 2018) and correspondingly lead to faster and more accurate reactions (Bertels et al., 2012; Kim et al., 2009). The brains of humans and other species have therefore been described as *anticipatory systems* (Rosen and Rosen, 2012) primarily engaged in matching incoming sensory inputs with top-down expectations (de Lange et al., 2018; Clark, 2013). Related hypotheses of general brain function and perception, dating back to von Helmholtz (1867), cast perception as a form

of hypothesis testing and (unconscious) inference (von Helmholtz, 1867; Gregory, 1980; Dayan et al., 1995; Knill and Pouget, 2004; Friston, 2005, 2010). The basic assumption of perceptual inference is that the brain infers on what is "out there" in the world by combining sense data with an internal generative model, i.e., the assumptions about how sensory patterns are generated by (hidden) states of the world. Given that the activation patterns at the sensory organs are its only source of information, the brain makes a "best guess" about the most likely causes of sensory inputs based on prior experience and assumptions of statistical regularities in the environment. These ideas have been furnished in the mathematical framework of Bayesian probability theory which provides a formalism to express this inferential process in terms of conditional probability distributions (Knill and Pouget, 2004; Friston, 2005, 2010). However, in its most general sense, perceptual probabilistic inference is cast as the assumption that the brain performs some form of predictive processing (Clark, 2013), capturing that the brain continuously infers the hidden structure of the environment and predicts incoming sensory data based on its generative model, without invoking assumptions of Bayesian computation or its neuronal implementation (Winkler et al., 2009; Winkler and Czigler, 2012).

1.1 Mismatch responses: MMRs

The sensory system automatically registers regularities in the input stream (Zhao et al., 2013; Chait, 2020) and detects corresponding changes rapidly (Sohoglu and Chait, 2016). Mismatch responses (MMRs) constitute specific brain signatures following a change in a sensory regularity. MMRs such as the mismatch negativity (MMN; Naatanen et al., 1978) and the P3 (Sutton et al., 1965) are among the earliest established EEG signatures. The P3 was the first specific component related to the detection of change in a sequential presentation of stimuli reported more than 50 years ago by Sutton et al. (1965) when computational approaches increasingly influenced neuroscientific and psychological research (Sutton, 1979). Thirteen years later, the mismatch negativity (MMN) was reported by Naatanen et al. (1978) and established as a second, earlier EEG component related to change detection in the auditory system. Both evoked responses were classically studied in the auditory modality, elicited by presenting a rare deviant sound in a stream of repeating standard sounds. In such an *oddball* paradigm, the frequent standard sound is occasionally replaced by a deviant sound differing in some property (Näätänen, 1992; Sams et al., 1983; Donchin, 1981). The rare presentation of deviants results in MMRs, typically recorded in the EEG, where the early MMN around 100-200ms is distinguished from the later P3 MMR around 250-500ms post stimulus presentation. Both components have been highly influential in various

branches of neuroscience such as perception, memory and disorders of the nervous system (Näätänen et al., 2019; Duncan et al., 2009; Polich, 2007) and are increasingly studied to gain insights into the brain's assumptions on the statistical contingencies underlying sequential inputs (Squires et al., 1975; Friston, 2005; Winkler et al., 2009; Winkler and Czigler, 2012; Heilbron and Chait, 2018). As such, MMRs provide a valuable tool to inspect probabilistic inference in the brain. In the following, we will introduce MMRs in the auditory, visual and somatosensory modality and highlight their equivalent reflection of the anticipatory nature of brain responses. Furthermore, the operationalization of the study of MMRs via the creation of model environments and computational modeling will be introduced before moving to a description of our empirical work on utilizing MMRs to assess the brain's generative model in the second section.

1.1.1 The Mismatch Negativity: MMN

The MMN was initially described by Naatanen et al. (1978) as reflecting a neuronal process of detecting changes in the auditory input stream. Its specific EEG signature is obtained by subtraction of the average standard response from the average deviant response, resulting in a difference wave that shows a negative peak between around 100-200ms post-stimulus onset at fronto-central electrodes with a corresponding polarity inversion at posterior electrodes around the mastoid bones (Giard et al., 1990). However, the exact peak latency depends on the relative stimulus change, shortening with increasing difference between stimuli (Sams et al., 1985; Tiitinen et al., 1994; Naatanen et al., 2007). While stimulation with the standard sound results in the auditory evoked potential corresponding to an ERP wave with a canonical sequence of peaks consisting of P1, N1 and P2 between 50 and 250ms, the presentation of a deviant sound results in stronger negativity within the N1-P2 complex which becomes best apparent in the signature MMN difference wave (Naatanen and Picton, 1987). The MMN can be elicited by small stimulus deviations in the absence of attention (Naatanen et al., 1978; Näätänen, 1992) and can be observed to almost any kind of feature differences in sound sequences, including frequency, intensity, duration, location and many others (Picton et al., 2000; Naatanen et al., 2007). Although the elicitation of the MMN is attention independent and usually not affected by task-related top-down influences (Alain and Woods, 1997; Naatanen et al., 1993; Ritter et al., 1999), its amplitude can be increased by attentional focus to the stimulation and decreased by attentional demands of a distracting primary task (Trejo et al., 1995; Woldorff et al., 1991, 1998). Inspite of its elicitation in the absence of attention, the MMN was shown to reflect participants' behavioral discrimination performance between stimuli (Brunia et al., 1993; Naatanen et al., 1993; Tiitinen et al., 1994) which makes it

an objective measure of perceptual stimulus discriminability and a highly valuable tool for clinical investigations of non-verbal, non-attentive, or even non-conscious patients (Naatanen, 2000; Naatanen et al., 2007; Atienza et al., 2001; Fischer et al., 1999; Wijnen et al., 2007).

The major neuronal generator of the auditory MMN is found in the auditory cortex where the exact location depends on the eliciting sound features and their complexity (Alho, 1995; Giard et al., 1990, 1995; Molholm et al., 2005; Sabri et al., 2004). Additionally, frontal cortex activation is generally assumed to contribute to the MMN (Deouell, 2007; Garrido et al., 2009b) as demonstrated by fMRI (Yucel et al., 2005; Opitz et al., 2002; Doeller et al., 2003; Molholm et al., 2005; Rinne et al., 2005) and intracranial recordings (Liasis et al., 2001; Rosburg et al., 2005). Although both generators are found independent of attentional focus, their activity can be modulated by attention (Yucel et al., 2005). Overall, a fronto-temporal network comprised of primary and non-primary auditory cortex and inferior frontal regions is canonically assumed to underlie the MMN component in the EEG (Deouell, 2007; Garrido et al., 2009b).

The MMN as a response to regularity violation

Over many years of MMN research, it has become increasingly evident that the MMN is not a response specific to the change in a stimulus *feature* but rather reflects a rule-based relationship between stimulus features (Paavilainen et al., 1999, 2007; Paavilainen, 2013; Naatanen et al., 2001, 2010). The interruption of repeating standard sounds with a deviant sound in the classic oddball paradigm turned out to be the most simple case of a regularity violation and the MMN is not restricted to stimulus *change* but can equivalently be found for rare *repetitions*. For instance, if a basic regularity is established by sequential presentation of two alternating tone pitches, the MMN is elicited if the continuous alternation is interrupted by occasional repetitions (Nordby et al., 1988; Ritter et al., 1992; Alain et al., 1994; Horvath and Winkler, 2004; Macdonald and Campbell, 2011). Similarly, the MMN was found in response to rare repetitions in an otherwise descending sequence of tones (Tervaniemi et al., 1994), indicating a form of abstraction from a detected rule or regularity underlying the sound presentation.

Saarinen et al. (1992) showed that the MMN can be elicted by rather complex patterns by presenting sequences of tone pairs with randomly differing frequency features across the pairs, such that no physically identical repeating standard stimulus was present in the sequence. A regularity was established by the fact that the tones within the tone pairs were strictly ascending, i.e. the second tone of the pairs was of higher frequency than the first, regardless of the (randomly set) frequency of the first stimulus. This abstract regularity of the direction of frequency change within

the tone pairs was occasionally violated by a presentation of tone pairs with descending frequency changes, eliciting an MMN to the second tone. This effect was replicated by Paavilainen et al. (1995) who additionally showed the MMN even if attention was drawn away to an entirely different sequence presented to the other ear. Moreover, using similar paradigms, the abstract feature MMN to two-tone patterns was also shown for infants (Carral et al., 2005; He et al., 2009) and rats (Ruusuvirta et al., 2007a,b), which provides evidence for a primitive mechanism underlying such auditory sensory pattern encoding (Naatanen et al., 2001).

More recently, highly abstract rules were shown to elicit MMN responses. In sequences used by Paavilainen et al. (2007), one stimulus-feature (length of the tone) was predictive of another stimulus-feature (frequency of the tone) such that long tones were followed by high tones and short tones by low tones. Importantly, the four different possible stimuli (four combinations of length and frequency) were presented an equal amount of times across the sequence (i.e. stimuli were overall equiprobable). An MMN was elicited when the transition rule was occasionally violated by presentation of a high tone following a short tone. This MMN to violations of a transition rule was replicated by Bendixen et al. (2008) who included additional change-points in the sequence at which the rule changed. Interestingly, only a few stimulus presentations following the change-point were necessary to elicit an MMN response to violations of the new transition rule, indicating a remarkable ability to quickly update the sensory rule representation. In addition to such single-feature regularities, the MMN can be elicited by violations of rules regarding feature combinations (e.g. intensity and frequency of a stimulus): In a study by Paavilainen et al. (2001), nine equiprobable standards of different frequencies were presented following the combinatory rule that tones of higher frequency have higher intensity. If such a feature combination was violated, e.g. by a high intensity tone of low frequency, the MMN was elicited. Such MMN responses to feature combinations have been shown in adults (Paavilainen et al., 2001; Gomes et al., 1997; Sussman et al., 1998), infants (Ruusuvirta et al., 2003, 2004) and rats (Astikainen et al., 2006), providing evidence for a ubiquitous role for abstract pattern encoding across stimulus features in basic auditory processing.

In studies showing the MMN to abstract rule violations, the regularities appear to be learned unconsciously and without explicit knowledge of the participants (Paavilainen et al., 2007; Schroger et al., 2007), even if participants were behaviourally able to detect the deviant (Paavilainen et al., 2001). Although their violations can draw attention away from a primary task (Schroger et al., 2007), abstract regularities can generally be picked up in the absence of attention (Bendixen and Schroger, 2008). Nevertheless, attentional focus can influence which aspects of a sequence are

considered to be the relevant reference patterns forming a standard stimulus trace. Experiments have shown that a regularly placed deviant in an oddball sequence (e.g. every fifth stimulus is a deviant; $p = 0.2$) only elicited the MMN if participants were asked to focus their attention on the pitch change, thereby ignoring the repeating pattern. In contrast, if participants were asked to attend to the repeating pattern, no MMN was observed (Sussman et al., 2002; Sussman, 2007, 2013). This demonstrated that standard formation can either be stimulus- or task-driven and attention to sequence features on a local (pitch) or global scale (pattern) might influence the MMN (Sussman, 2007).

The stimulus driven, automatic processing of the auditory input stream is studied in a closely related field of research, concerned with auditory scene analysis (ASA; Bregman, 1994). This line of research studies important basic properties of the sensory system such as the automatic integration of stimulus features to perceptual elements. For instance, the repeating pattern of four standard stimuli followed by a deviant might be integrated by the system to form a specific sequence (Sussman, 2007). Additionally, ASA also concerns the segregation of sensory inputs according to their sources of origin. In case of concurrent input streams, MMN responses have been used to indicate prior basic auditory object binding by inspecting if the MMN is restricted to deviants of a specific input stream during simultaneous presentation of different standard regularities in separate streams. Using the MMN, stream segregation was not only shown for simultaneous input streams presented separately to the ears (Ritter et al., 2000; McKenzie and Barry, 2006), but also for up to three streams from different positions in space (Nager et al., 2003), as well as for two streams from a single source differing only in their frequency ranges (Shinozaki et al., 2000; Winkler et al., 1992). Similarly, MMN can probe the properties of object formation in terms of stimulus grouping. For instance, if chunks of four standards followed by a deviant are successively presented, this elicits an MMN for the fifth stimulus (deviant) only as long as the inter stimulus interval (ISI) is sufficiently long. If rapidly presented ($<$100ms), the deviant is grouped with the standards without eliciting an MMN (Sussman et al., 1998). Moreover, Sussman (2005) used MMN responses to show that the auditory input is first segregated into different streams and that the stimuli are subsequently integrated to form single sound elements. Therefore, basic ASA, including input stream segregation and integration, has been suggested to be a prerequisite for the occurrence of the MMN (Winkler et al., 1992, 2003) and might provide the grounds on which a regularity can be formed and their violation detected.

Regularity violation as evidence for sensory expectation

Evidence for MMN in response to abstract regularity violations indicates a form of "primitive intelligence" in the auditory system (Naatanen et al., 2001, 2010). In order to detect differences between inputs, the system likely performs initial ASA (Bregman, 1994; Winkler et al., 1992, 2003; Sussman, 2005) and appears to have access to stimulus traces forming regularities over certain periods of time, potentially stored in sensory memory (Naatanen et al., 1978; Naatanen and Winkler, 1999; Tiitinen et al., 1994; Naatanen et al., 2005). Furthermore, MMN responses provide indications for an extrapolation of stimulus patterns and basic rule abstraction which suggest the formation of sensory *anticipation* of the upcoming stimulus (Tervaniemi et al., 1994; Naatanen, 2000; Winkler, 2007).

The examples presented above provide evidence for a role of stimulus expectancy and sensory anticipation. These indications are substantiated by findings related to omission responses which might be considered a special kind of MMR. Omission responses are responses to omitted stimuli from a sequence or an expected pattern without any physical input and are particularly well suited to isolate expectation related effects from other stimulus related processing (Yabe et al., 1997; Tervaniemi et al., 1994; Hughes et al., 2001; Bendixen et al., 2009; Salisbury, 2012; Chennu et al., 2016; SanMiguel et al., 2013a,b; Wacongne et al., 2011). The omission of an expected sound in a sequence of regularly presented sounds can result in an omission response time-locked to the expected occurrence of the regular stimulus which has been termed the omission MMN if it is observed within the MMN time window (e.g., Yabe et al., 1997; Tervaniemi et al., 1994; Hughes et al., 2001; Salisbury, 2012; Chennu et al., 2016; Prete et al., 2022). A recent study showed that the omission MMN to unexpected omissions was larger than the omission MMN to regular (thus rather expected) omissions (Prete et al., 2022). Moreover, using connectivity analyses, Chennu et al. (2016) indicated that the MMN as well as the omission MMN are both accompanied by increased forward and backward connectivity in a fronto-temporal network differing only in their driving input strengths, which indicates that similar recurrent processing underlies both responses. Additionally, the authors compared attended and unattended sequences and found that attentional focus decreased both mismatch responses mediated by increased feedback connectivity within the network, highlighting a top-down modulatory influence of attention. A slightly earlier omission response between 50-100ms has been described by Bendixen et al. (2009) who aimed to isolate the sensory expectation of specific tone features from an anticipation of upcoming input alone. The authors compared three types of omissions of sequential tones within a large range of varying frequencies, thus controlling for confounding effects of frequency differences. Strikingly, only in

the condition in which the stimulus omission concerned a predictable tone, an omission response was observed. Moreover, unlike the omission of an unpredictable tone, the omission of the predictable tone showed no signal difference in the first 50ms of the response compared to the actually presented (regular) tone. This indicated highly comparable initial processing of the physically presented and the expected sound in the auditory system until a detection of the omission occurred within 100ms (Bendixen et al., 2009). Taken together, the omission MMN demonstrates the anticipatory nature of the auditory system and indicates that the MMN reflects regularity violation with respect to expected stimuli.

The temporal specificity of the auditory system regarding the expectations of incoming sounds was probed by Grimm and Schroger (2007). The authors tested if frequency changes at different temporal positions within a single (short) tone sample were represented as separate events differentiable by their MMN responses, or if the MMN was rather sensitive to the general frequency deviance in the tone, regardless of the exact temporal position. In two basic oddball conditions, the authors presented short standard tones occasionally interrupted by deviant tones which differed from the standard in a frequency modulation *within* the tone at exactly 200ms of the tone sample but were otherwise identical. In one condition the deviant was presented with probability $p = 0.1$ in the sequence and in the other condition with $p = 0.3$. In the third key experimental condition, three different deviant tones were presented in the sequence, each with $p = 0.1$. The deviants each differed from the standard by a frequency modulation at different time points *within* the tone (at 100, 200 and 300ms). In this third condition, even though the overall probability of deviance was as high as in condition two ($p = 0.1+0.1+0.1 = 0.3$), the MMN actually resembled the response in the first condition ($p = 0.1$). This indicated that the MMN reflects a spectro-temporally specific expectation (exact point of modulation in the tone) rather than a feature specific one (any modulation of the tone) and has been suggested by the authors to reflect the point-wise comparison of current and upcoming inputs in the auditory system (Grimm and Schroger, 2007).

Temporally-specific stimulus expectancy often forms the basis of MMN responses to regularity violations. Research in humans (Takegata and Morotomi, 1999; Yabe et al., 1997), as well as rats (Jalewa et al., 2021) has shown that the MMN is dependent on the stability of stimulus onsets, indicating that temporal predictability is an important modulatory property of the MMN (Jalewa et al., 2021), comparable to the occurrence probability of a deviant and the magnitude of relative stimulus changes. Already Naatanen and Picton (1987) differentiated between "temporal" and "event" uncertainty, a notion which is similarly reflected in more recent distinctions between sensory expectations with respect to "when" a stimulus occurs and "what" features might

be perceived (Auksztulewicz et al., 2018). Specific omission responses, for instance, have been indicated to be related to the expectation of specific stimulus characteristics, rather than temporal expectancy (SanMiguel et al., 2013a,b). SanMiguel et al. (2013a) inspected omission responses to self-generated sounds, presenting participants with a specific sound following a button press in one condition, and a random sound following the button press in a second condition. In a few trials, the sound following the button press was omitted. Although the omission of the stimulus was unexpected and the stimulus was *temporally* predicted in both conditions, only the omission of the sound with predictable stimulus *characteristics* (i.e., condition one) resulted in an omission response. These results suggest that (at least some form of) omission responses reflect expectations regarding the stimulus identity.

Summary

Overall, the auditory MMN has proven to provide invaluable insights into the principles of auditory processing. Specifically, research has provided overwhelming evidence for a dependence of MMN responses on stimulus expectancy and a remarkable ability of the auditory system to extract sensory regularities from sequential input on different levels of complexity. The findings therefore strongly suggest a specialized role of the auditory and likely the sensory system in general for the processing of probabilistic contingencies in the sensory environment.

1.1.2 The visual MMN

The equivalent response to the MMN in the visual modality is commonly termed the vMMN. Although the specific characteristics and underlying mechanisms are the subject of ongoing debates (Czigler, 2007; Kimura et al., 2011; Stefanics et al., 2014; Male et al., 2020) the vMMN is a well researched EEG component. It has been reported for changes in stimulus color (Czigler et al., 2002, 2004; Kimura et al., 2006; Grimm et al., 2009; Czigler and Sulykos, 2010; Mo et al., 2011; Stefanics et al., 2011), shape (Maekawa et al., 2005; Tales et al., 1999, 2002; Tales and Butler, 2006), spatial frequency (Heslenfeld, 2003; Kenemans et al., 2003, 2010; Maekawa et al., 2005, 2009; Sulykos and Czigler, 2011), orientation (Astikainen et al., 2004, 2008; Czigler and Pato, 2009; Kimura et al., 2009; Czigler and Sulykos, 2010), location (Berti and Schroger, 2004, 2006), contrast (Stagg et al., 2004) and even motion (Kremlacek et al., 2006; Lorenzo-Lopez et al., 2004; Pazo-Alvarez et al., 2004a,b), among others.

The vMMN is reported for a more variable latency range and scalp distribution than its auditory counterpart, but can be characterized as a negative EEG component around 100-250ms at

posterior electrode sites such as occipital and temporo-parietal electrodes (Pazo-Alvarez et al., 2003; Czigler, 2007; Kimura et al., 2011). In correspondence to the auditory modality, the vMMN has been elicited in absence of attention during a variety of primary attentional tasks (Czigler, 2007). Although it is generally considered more difficult to overcome visual sensory dominance and to draw attention away from visual input compared to the auditory modality (Sinnett et al., 2007; Neumann et al., 1986; Czigler, 2007), the vMMN has been shown in absence of attention and can be considered an attention independent early mismatch response comparable to the auditory MMN (Czigler, 2007; Kimura et al., 2011; Stefanics et al., 2014). This view is supported by studies showing the vMMN during attentional focus on auditory stimulation (Astikainen et al., 2004) or during demanding primary visual tasks with task-irrelevant deviants outside the focus of attention (Czigler et al., 2002; Heslenfeld, 2003; Czigler et al., 2004; Winkler et al., 2005).

The neuronal generators underlying the vMMN are commonly found in the visual cortex (Pazo-Alvarez et al., 2003; Czigler, 2007; Kimura et al., 2011), with some studies specifically highlighting activation of hierarchically higher visual areas (Kimura et al., 2010a; Urakawa et al., 2010; Yucel et al., 2007; Stefanics et al., 2019) as well as frontal contributions (Yucel et al., 2007; Hedge et al., 2015; Pazo-Alvarez et al., 2003; Kimura et al., 2011). Using fMRI, Yucel et al. (2007) report activation in primary and non-primary visual as well as parietal and pre-frontal areas in response to unattended stimulus deviance. Although the activation of these regions was observed in the absence of attention to the stimulation, activation of visual regions was significantly reduced with increased attentional load during the primary task, equivalent to what the authors had previously reported for the auditory MMN in auditory regions (Yucel et al., 2005). In parallel to the auditory modality, inferior frontal cortex in particular has been suggested to contribute to the vMMN and mismatch processing across sensory modalities (Hedge et al., 2015; Downar et al., 2000; Iglesias et al., 2013; Kim, 2014).

vMMN reflects expectation violation in response to sequence regularities

As described above, it is commonly assumed that the auditory MMN depends on the expectancy of a stimulus based on a previous sequence regularity, independent of stimulus properties. Similarly, the vMMN is found to be independent of the direction of change (e.g., in luminance intensity from low to high or v.v.) and to be particularly sensitive to the probability of stimulus deviance (Stagg et al., 2004). Interestingly, based on its dependence on stimulus probability, the vMMN was dissociated from an early posterior positivity around 100ms which reflected stimulus *change* as such (Kimura et al., 2006, 2005b,a; Fu et al., 2003): In contrast to the earlier positivity which

was elicited by stimulus changes even if they were equiprobable to repetitions, the vMMN around 100-200ms only followed low probability changes and was additionally shown to increase with the amount of prior standard repetitions before the deviant (Kimura et al., 2006).

In parallel to the auditory modality, the vMMN has been shown in response to sequence regularities with different levels of complexity and abstraction. In fact, many of the above presented experiments on the auditory MMN have been similarly used to probe the vMMN, indicating highly comparable sensitivity to stimulus expectancy. For example, the vMMN can also be elicited by unexpected repetitions interrupting sequences of regularly alternating visual features such as colors (Czigler et al., 2006a) or emotional faces (Kimura et al., 2012). Moreover, the elicitation of the vMMN depends on a certain level of *unexpectedness* of the deviant. In striking similarity to the auditory response, the vMMN in response to predictable oddball sequences with regularly placed deviants (e.g. every fifth stimulus, $p = 0.2$) was only elicited if participants were asked to attend to the local properties of the stimulus sequence (e.g. the luminance of the stimuli) while ignoring the global sequence regularity (repeating pattern) of the repeating deviant (Kimura et al., 2010b). If the global regularity was attended, the vMMN response broke down, indicating that standard pattern formation can be modulated by top-down demands, especially during global regularity violation, which had been previously indicated for the auditory MMN (Sussman et al., 2002; Sussman, 2007, 2013).

Stefanics et al. (2011) tested the vMMN for a set of more complex regularities specifying transition rules of stimulus pairs. Participants were shown red and green stimuli presented in temporal pairs of different combinations, such that two stimuli were presented in rapid succession (300ms ISI), followed by a slightly longer break (800ms ISI) before the next pair. Transition rules of the form *"if the first stimulus in a pair is red (green), then the second stimulus is likely green (red)"* were introduced to manipulate stimulus expectancy in three types of sequences. One random sequence contained equiprobable stimulus pairs of the same (e.g. red followed by red) or different colors (e.g. red followed by green). One sequence included 30% of different-color pairs and one sequence included 10% different-color pairs, such that in both of these sequences the second stimulus of a different-color pair was unexpected given the first stimulus of the pair (with two levels of probability: $p = 0.1$ and $p = 0.3$). Although the colors within each sequence were overall equiprobable, the observed EEG responses to the second stimulus of a pair were different between the three sequences: In the random sequence without any regularity (or conditional rule), the visual changes only elicited the posterior change positivity described above (Kimura et al., 2006, 2005b,a; Fu et al., 2003). In both other conditions, the second stimulus of a different-color

pair elicited the vMMN and the most pronounced vMMN was observed in the low probability ($p = 0.1$) condition (although with later latency). These results demonstrated the sensitivity of the vMMN to stimulus expectancy based on conditional rules in striking similarity to the auditory MMN.

Visual expectation violation to sequence regularities has additionally been demonstrated with an omission response comparable to the one found in the auditory modality. Occasional omissions in a regular presentation of visual stimuli resulted in responses with the modality specific topography of the vMMN between 150-200ms which was shown to be independent of attention and task-relevance (Czigler et al., 2006b), indicating the existence of an omission vMMN. The authors additionally observed a similar omission vMMN in response to the offset of rapid trains of visual stimuli. In order to exclude influences of steady state responses to repeating stimuli, the trains of stimuli were followed by longer breaks. Therefore this offset response indicated expectancy related omission effects based on the sensory regularity built up by the stimulus trains.

Summary

Taken together, ample evidence suggests a comparable dependence of the vMMN on stimulus expectancy and a sensitivity of the visual system to sequence regularities on different levels of complexity. Although much less researched in this respect, the similarity of the key properties of the vMMN to its auditory equivalent suggests similar specialization to probabilisitic sensory inputs.

1.1.3 The somatosensory MMN

The corresponding response to the MMN in the somatosensory modality (sMMN) was first reported by Kekoni et al. (1997). Although much less researched than its counterparts in audition and vision, the sMMN has since been shown in response to stimulus deviance in a variety of properties, including spatial location (Kekoni et al., 1997; Shinozaki et al., 1998; Akatsuka et al., 2005; Restuccia et al., 2007, 2009; Chen et al., 2008; Hu et al., 2013; Strommer et al., 2014; Naeije et al., 2016, 2018; Fardo et al., 2017; Shen et al., 2018b,a; Hautasaari et al., 2019; He et al., 2020; Xu et al., 2021b), stimulus frequency (Kekoni et al., 1997; Spackman et al., 2007; Butler et al., 2012), stimulus duration (Akatsuka et al., 2005, 2007a,b; Spackman et al., 2007, 2010) and stimulus intensity (Ostwald et al., 2012; Zhang et al., 2019). In parallel to the other modalities, the sMMN (and its equivalent response in MEG) is observed primarily between 100-250ms post stimulus onset, although a considerable number of studies have reported an additional early com-

ponent between 50-100ms (Shinozaki et al., 1998; Akatsuka et al., 2005, 2007a,b; Ostwald et al., 2012; Naeije et al., 2016; Xu et al., 2021b).

In contrast to the MMN, the EEG scalp distribution of the sMMN is less consistent across studies and conflicting evidence has been reported. Nevertheless, evidence clearly converges to a negative deflection contralateral to the stimulation site, with some studies reporting more fronto-central (Kekoni et al., 1997; Shinozaki et al., 1998; Akatsuka et al., 2005; Restuccia et al., 2009; Fardo et al., 2017; He et al., 2020), more central (Spackman et al., 2010; Ostwald et al., 2012; Shen et al., 2018b), more temporal (Hu et al., 2013; Strommer et al., 2014; Zhang et al., 2019; Shen et al., 2018a), or more parietal (Restuccia et al., 2007) extension of the effect in EEG. Given the (lateralized) central location of the somatosensory system in the brain it is likely that the exact extension of a lateral activation (i.e., from frontal to parietal) of the scalp electrodes varies strongly with the dipole orientation and location within the somatosensory system. Such an interpretation is in line with the large variety of stimulus properties used to elicit the sMMN which possibly activate different dipolar generators in the somatotopic organisation. Evidence from fMRI (Chen et al., 2008), ECoG (Spackman et al., 2010; Butler et al., 2011) and dipole modeling (Spackman et al., 2010; Ostwald et al., 2012; Naeije et al., 2016, 2018; Fardo et al., 2017; Andersen and Lundqvist, 2019; Hautasaari et al., 2019; Xu et al., 2021b) indicate the somatosensory system as the primary driver of the sMMN response. Moreover, in correspondence with the auditory and visual modality, inferior frontal generators have been reported to contribute to the sMMN (Chen et al., 2008; Spackman et al., 2010; Allen et al., 2016; Fardo et al., 2017). Spackman et al. (2010) showed via intracranial recordings that the sMMN response over somatosensory cortex was followed slightly delayed by activity over the pre-frontal cortex, indicating a comparable signal cascade as suggested for auditory and visual modality. In an fMRI experiment Chen et al. (2008) suggested that SII activity related to the sMMN response was entirely attention independent, while activations in a fronto-parietal network were modulated by attention. On the other hand, Fardo et al. (2017) additionally suggested a modulatory role for top-down attention on SI and SII network activity during somatosensory mismatch processing.

Nevertheless, in correspondence with the pre-attentive auditory MMN, the somatosensory equivalent response can be elicited in absence of attention, as shown by most sMMN studies (Kekoni et al., 1997; Shinozaki et al., 1998; Akatsuka et al., 2005, 2007a,b; Spackman et al., 2007; Restuccia et al., 2007, 2009; Chen et al., 2008; Spackman et al., 2010; Butler et al., 2012; Hu et al., 2013; Strommer et al., 2014; Fardo et al., 2017; Shen et al., 2018b,a; Andersen and Lundqvist, 2019; Zhang et al., 2019; He et al., 2020; Xu et al., 2021b), although attentional focus has been

shown to increase the sMMN (Hu et al., 2013; He et al., 2020) and to facilitate mismatch processing in the somatosensory system (Fardo et al., 2017; He et al., 2020). The attention-independent elicitation of the sMMN highlights its similarity to the auditory and visual modality and indicates a comparable role of the sMMN in early automatic change detection in the somatosensory system. Moreover, it also makes the response a useful tool for clinical investigations of sensory function in unresponsive or non-verbal individuals (Naatanen, 2009), comparable to the auditory MMN and potentially with additional sensory specific properties. The mature functioning of the sMMN has been shown in healthy children (Restuccia et al., 2009; Spackman et al., 2010) and infants (Shen et al., 2020). Therefore, as suggested by Naatanen (2009), an altered sMMN might serve as a comparative measure for developmental disorders and potentially even for neurodegenerative diseases and psychiatric conditions as these are all known to be accompanied by modulations of the auditory MMN.

In an interesting line of research, Shen et al. (2018b,a, 2020) have recently highlighted the potential of sMMN for insights specific to the somatosensory modality. Shen et al. (2018b) demonstrated the use of the sMMN to study body part segmentation by showing a smaller sMMN for location deviants when the stimuli were placed on the forearm within the wrist boundary (which is thought to be a body-part segmentation boundary; e.g. de Vignemont et al., 2009), compared to an sMMN for equidistant stimulation points separated by the wrist boundary. The authors argued that the increased sensory mismatch for deviant stimulation across the wrist boundary suggests a different representation of the hand as a separate body part. With the same study protocol, Shen et al. (2020) indicated that infants have such a structured body representation already at a few months of age. Moreover, Shen et al. (2018a) showed that the sMMN reflects the somatotopic organization in the brain. The authors presented sMMN to location deviants of different distances in somatotopic space compared to distances in physical space on the body. In a classic oddball protocol, stimulation on the lip functioned as the standard stimulus and the sMMN to deviant stimulation on the neck (which has comparably closer bodily distance and farther somatotopic distance) was compared to sMMN to deviants on the hand (farther bodily distance, closer somatotopic distance). Despite being closer in physical space, the neck-deviant elicited larger sMMN than the hand-deviant, suggesting that early somatosensory mismatch processing is influenced by the somatotopic representation of body parts. As such, the sMMN might provide a unique tool for the investigation of sensory expecations with regards to body representations.

The sMMN as a reflection of expectation violation

While unequivocally seen as a key feature of MMRs such as the auditory and visual MMN (Näätänen et al., 2019; Wacongne et al., 2012; Kimura et al., 2011), to date, only very few studies have specifically investigated the relationship of the sMMN to sensory expectancy. Basic sensitivity of the sMMN to stimulus probability has been shown by Akatsuka et al. (2007b), showing increases in the early (50ms) and later (150-250ms) responses to deviants presented with probability $p = 0.1$ compared to $p = 0.3$ and $p = 0.5$. Moreover, initial evidence for a somatosensory omission response has been provided by Yamashiro et al. (2008) who showed a somatosensory P100 and N140 response-complex time-locked to the offset of a train of repetitive stimuli. The responses were identical for attended and unattended stimulus trains and, although not framed as such by the authors, this offset-response provided first indications for a somatosensory equivalent to an omission MMN as it corresponded to an expectation violation response to a physically absent stimulus. More recently, direct evidence for an omission sMMN was indeed provided by Naeije et al. (2018) and Andersen and Lundqvist (2019), combining MEG measurements with source modeling. Both studies show an omission sMMN between 100-200ms, generated in SII. These results strongly suggest a specific sensitivity of the sMMN to stimulus expectancy.

In additional analyses, Andersen and Lundqvist (2019) showed differential frequency modulations for the omission sMMN and stimulus repetitions. While the omission response, reflecting expectation violation, was accompanied by increases in gamma oscillations in the insular cortex, stimulus repetitions showed increased power in the beta-band frequency localized to a parietal cortex source. The authors interpreted stimulus repetitions as reflective of expectation increase since no habituation tendencies were otherwise observed in the MEG signal. Strikingly, these observations are in line with previous suggestions for an asymmetric representation of expectations and their violation by beta and gamma oscillations, respectively (Arnal and Giraud, 2012; Bastos et al., 2015; Friston, 2019).

Apart from these initial findings, almost no work has been done in the somatosensory modality to explicitly investigate effects of stimulus expectancy. Xu et al. (2021b) recently presented a study in which they placed double-deviants in an otherwise classic oddball paradigm, such that the second deviant was rendered fully predictable. The authors showed early (30-100ms) and later (130-230ms) MMRs in MEG for both deviants, with combined activation of SI and SII underlying these components. Only the early response was significantly increased for the first (unpredictable) compared to the second (predictable) deviant, whereas the later response did not show a difference. The authors interpreted their finding as evidence for a sensitivity of the earlier mismatch

component to stimulus predictability. However, several issues with the experimental setup complicate an interpretation of these results. The first deviant consisted of a physically stronger input than the standard which can fully explain the increased early response, known to be strongly modulated by stimulus intensity. Moreover, the first deviant was a double stimulus (accompanied by the standard) and the second deviant was presented alone (no standard). In addition to reducing the physical input for the second deviant, this could even have introduced an additional underlying omission response to the standard regularity (being only interrupted during the second deviant presentation) which might counteract any attenuation of the later mismatch response by stimulus predictability. Therefore, overall, the study by Xu et al. (2021b) can only provide initial indications of expectancy effects on somatosensory mismatch responses

Summary

Studies in the auditory and to a lesser extend the visual modality have seen many replications of the key findings corresponding to elicitation of mismatch signatures by complex regularity violations and stimulus expectancy and their results are often supported with control experiments. Although the sMMN has been established as a mismatch response comparable to its counterparts in the auditory and visual modality, research on the sMMN only provides initial evidence for its equivalence with respect to the role of stimulus expectancy. Nevertheless, particularly the omission sMMN might be regarded as compelling evidence for an important role of expectation during sMMN generation.

1.1.4 The P3

The P3 is another long known mismatch signature in response to deviations from stimulus regularities at a later latency than the MMN (Sutton et al., 1965; Squires et al., 1975; Donchin, 1981; Duncan-Johnson and Donchin, 1982; Picton, 1992). The large positive going component peaks about 250-500ms after stimulus onset over central electrodes and can be further divided into two sub-components, the earlier P3a with central to fronto-central scalp distribution and the later P3b with centro-parietal extension (Polich, 2007). The P3 is generally considered to be an attention-dependent response (Duncan-Johnson and Donchin, 1982; Näätänen and Gaillard, 1983; Polich, 2007; Duncan et al., 2009). However, while the earlier P3a is task-independent, drawing observers' attention to rare or novel stimuli (Escera et al., 2000; Friedman et al., 2001; Knight and Scabini, 1998), the later P3b is rather sensitive to task-related target stimuli (Polich, 2007; Duncan et al., 2009).

Although primarily researched in the auditory modality, this large central positivity with late latency of around 300ms is also found for visual (Conill, 1998; Picton, 1992; Duncan et al., 2009) and somatosensory deviance (Yamaguchi and Knight, 1991, 1992; Ostwald et al., 2012; Shen et al., 2018b,a; Andersen and Lundqvist, 2019) and the P3 largely shows modality independent characteristics (Polich, 2007). Correspondingly, the generating sources underlying the P3 are found in a distributed network, involving inferior frontal, temporo-parietal and anterior cingulate regions, while contributions from sensory regions such as superior temporal and occipital cortex are thought to be sensory specific (Linden, 2005; Polich, 2007). While the task-independent P3a is indicated to predominantly arise from activation in pre-frontal and inferior parietal regions, activations in anterior cingulate and partly in parietal cortex are involved during task-related P3b responses to target stimuli (Linden, 2005; Polich, 2007).

The P3 as a reflection of expectation violation

As elaborated in the previous sections, MMN responses across modalities signal violations of regularities in the sensory environment, largely independent of attentional focus. As such, the MMN is thought to reflect a first stage of novelty processing, registering unexpected sensory input. It has been suggested that the P3a subsequently signals an involuntary attention switch to the registered sensory event and thus corresponds to a second stage of novelty processing (Escera et al., 2000; Friedman et al., 2001; Wetzel and Schroger, 2014). Although the P3a has been associated with the process of attention switching, the exact role of the response within the second stage of novelty processing is debated (Dien et al., 2004; Horvath et al., 2008; Schroger et al., 2014; Wacongne et al., 2011). In particular, it has been shown that the P3a can either be elicited by stimulus driven novelty chained to the MMN or by the assumed significance of an event, without prior elicitation of MMN (Horvath et al., 2008; Rinne et al., 2006; Schroger et al., 2014). As such, the response might relate to expectation specific processes and evaluation of stimulus significance for the system rather than attention switching itself (Horvath et al., 2008; Schroger et al., 2014).

The idea that the P3a constitutes a response to *unexpected* stimuli is found already in its earliest descriptions (Sutton et al., 1967; Squires et al., 1975; Donchin, 1981), and Sutton et al. (1967) even demonstrated its elicitation by the offset of expected stimulus presentation, similar to an omission response. More recently, studies have again focused on the relation of the P3 to stimulus expectancy. Novel sounds have been demonstrated to capture attention due to the violation of perceptual expectations based on sensory regularities and implicit learning of conditional probabilities rather than the rarity of the stimulus *per se* (Parmentier et al., 2011). Moreover, in experiments

where stimulus sequences were self-generated, the P3a has been shown to be modulated by the predictability of the upcoming stimulus (von Carlowitz-Ghori et al., 2011; SanMiguel et al., 2013a,b). In an experiment by von Carlowitz-Ghori et al. (2011), participants could either self-generate standards and deviants with button presses (fully predictable) or self-generate a stimulus which would turn out to be either standard or deviant (timing predictable, content unpredictable). Compared to normal oddball sequences, the authors found a reduced P3a for fully predictable deviants. In the second case, although the timing of the self-generated stimulus was fully predictable, an increased P3a was observed in response to the unpredictable content of the stimulus. This was interpreted as highlighting its sensitivity to a mismatch between actual and anticipated effect of an action (button press). In the experiments by SanMiguel et al. (2013a,b), already described in section 1.1.1, only unpredictable omissions of self-generated sounds elicited MMN as well as P3 responses, while predictable omissions did not. These effects are in line with recent findings of an omission P3a during passive observation of classic oddball sequences (Prete et al., 2022). The authors provided first evidence for a clear distinction between expected and unexpected omissions, showing an increased P3a for unexpected stimulus omissions compared to omissions at regular (predictable) intervals. Moreover, Chang et al. (2018) showed that in predictable oddball sequences (where the deviant occured every fifth stimulus; $p = 0.2$), P3a amplitude reduction to predictable deviants correlated with beta frequency modulation before the deviant presentation, while no beta modulation was observed during the unpredictable sequences. Their results are in line with the suggested role for beta-band frequency modulations during stimulus prediction (Bastos et al., 2015; Friston, 2019). Taken together, these results suggest that the P3a explicitly reflects novelty in terms of expectation violation.

The P3 in response to global regularity violation

An important experimental paradigm in current MMR research is the *local-global* paradigm (Bekinschtein et al., 2009) which has been used to inspect expectation violation to stimulus regularities on two levels of complexity within a single sequence. The paradigm is a variant of a predictable oddball sequence, in which the deviant is presented regularly (e.g. every fifth stimulus; $p = 0.2$). In such a sequence, on the local level the deviant interrupts a train of repeating standards while simultaneously forming a regularly repeating pattern on a global level. Generally, such a sequence is known to elicit an MMN (and often P3) in response to the local deviant, although it was shown that the MMN can be suppressed by explicit attention to the global regularity (Sussman et al., 2002; Sussman, 2007, 2013; Kimura et al., 2010b) and the P3 can be reduced if participants are

given information of the regularity prior to the experiment (Max et al., 2015), indicating influences of top-down expectancy modulation of both the MMN and the P3. In their new variant of the paradigm, Bekinschtein et al. (2009) created a global regularity by presenting sequences of chunks consisting of five stimuli in which the last stimulus was a deviant. Therefore, locally, the fifth stimulus was a deviant, and globally it was an expected stimulus (standard). They added a violation of the global regularity by occasionally presenting chunks in which the fifth stimulus was a standard (instead of a deviant). The authors show that the repetition of a standard sound at the position where the deviant was (globally) expected elicited a P3 response, whereas no MMN was observed. The finding of a P3 response to such a regularity violation indicated that the P3 reflects expectation violation on a more global scale of sequence perception, suggesting a hierarchy of mismatch processing with increasing levels of information integration in line with an interpretation of MMN and P3 reflecting two stages of novelty processing.

The effects of the local-global paradigm have been replicated in several studies providing different insights into the putative mismatch processing hierarchy reflected by MMN and P3 (Wacongne et al., 2011; Chennu et al., 2013, 2016; El Karoui et al., 2015; Niedernhuber et al., 2022; Chao et al., 2018; Uhrig et al., 2014). Importantly, Niedernhuber et al. (2022) have recently shown that the P3 in response to global regularity violation can be equivalently observed across senses in the auditory, visual and somatosensory modality in accordance with the suggested modality independent role of the P3. Chennu et al. (2013) showed that the (auditory) MMN to local deviants was only observed if the local sequence features were explicitly attended whereas the P3 in response to global deviants was contingent on an attentional focus to the global regularity, as suggested by earlier findings on the predictable oddball sequence described in previous sections (Sussman et al., 2002; Sussman, 2007, 2013; Kimura et al., 2010b). Moreover, complementing source reconstruction with intracranial recordings, Chennu et al. (2013) localized the underlying generators of the P3 to global deviants primarily in temporo-parietal and pre-frontal regions. These results were confirmed by El Karoui et al. (2015) who additionally provided a systematic description of the processes underlying auditory local and global responses with event-related, time-frequency and connectivity analyses of intracranial recordings. Their findings suggest that local and global MMRs represent a cascade of two distinct neural events. The response to local deviants was found over the auditory cortex in superior temporal lobe around 135ms, reflecting an auditory MMN which was accompanied by power increases in the high gamma frequency band comparable with localized sensory processing (Edwards et al., 2005), and by early connectivity increases in temporal cortex around 80ms. In contrast to the focalized early response to local deviants, the global

deviants elicited a later response (around 360ms) distributed across temporal and frontal electrode sites, reflective of a P3. Additionally, the late response was accompanied by modulations in beta band synchronization over frontal cortex as well as increased fronto-temporal connectivity starting from 160ms. Similarly, studies in monkeys have corroborated these findings (Chao et al., 2018; Uhrig et al., 2014): Early mismatch responses to local deviants were found around auditory cortex accompanied by oscillations in the gamma-band and later mismatch responses to global deviants were found in temporal and frontal cortex accompanied by alpha/beta oscillations, along with increased feedback connectivity from frontal cortex (Chao et al., 2018). Uhrig et al. (2014) additionally report parietal involvement during global regularity violations. Taken together, these findings are highly suggestive of a hierarchy of mismatch processing in which violations of local sequence regularities are reflected in an early MMN generated primarily in sensory regions, whereas violations of global sequence regularities are reflected in a later P3 response recruiting a distributed network of sensory and modality independent frontal and parietal regions (Niedernhuber et al., 2022).

The hypothesized central role for stimulus expectancy during mismatch processing in response to local and global regularity violations was further supported by studies complementing the local-global paradigm with unexpected omissions (Wacongne et al., 2011; Chennu et al., 2016). For instance, Wacongne et al. (2011) tested the responses to different sequences of stimulus chunks (as in the classic local-global paradigm). In one sequence, chunks of five standards were repeated with occasional local deviants at the fifth position. If the fifth stimulus in a chunk was omitted, the omission MMN was observed. In a second sequence, chunks of five stimuli with a local deviant at the fifth position were repeated, forming a global regularity. As expected, the occasional presentation of a standard at the fifth position resulted in the global P3 MMR as it formed a globally unexpected repetition. An omission of the fifth stimulus in these sequences, however, elicited an increased and extended omission MMN. The authors reason that the omission violated two predictions simultaneously: The local regularity predicted a standard repetition and the global regularity predicted a deviant, none of which actually occurred (omission). The authors interpreted the increased omission response as evidence for two overlapping top-down predictions in the absence of any bottom-up sensory signals.

Summary

The P3 is a late MMR, reported across the senses and related to stimulus expectancy comparable to the earlier MMN responses. In contrast to the MMN, the P3 has been shown to be sensitive

global sequence regularities, potentially reflecting increasing information integration in a second stage of mismatch processing.

1.2 Theories on MMR generation

Despite being among the most well-studied EEG components, the neuronal generation of MMRs, and the MMN in particular, remain subject of ongoing debate (Naatanen et al., 2005; May and Tiitinen, 2010; Garrido et al., 2009b; Heilbron and Chait, 2018). In its original description, the MMN was hypothesized to result from a comparison of the sensory input against a memory trace of the recent past within the sensory system, signaling a mismatch if the memory representation is not matched by the input (Naatanen et al., 1978; Naatanen and Winkler, 1999; Tiitinen et al., 1994).

The insights discussed in the previous sections regarding the dependence of mismatch responses on regularity violation at various scales have led to further developments of memory-based accounts of MMN generation into model- and expectation-based interpretations (Winkler, 2007; Winkler et al., 2009; Winkler and Czigler, 2012). In contrast to previous formulations who have sometimes hinted at a notion of predictability within the memory representation (Tervaniemi et al., 1994), Winkler (2007) explicitly suggested that the expectation of an extracted sensory regularity lies at the core of MMN generation. The view assumes a perceptual model which abstracts representations of sensory regularities from previous stimulus exposure and forms expectations about incoming sensory inputs. Whenever inputs diverge from such model-based expectations, the MMN is elicited and the representation requires updating (Winkler, 2007; Winkler et al., 2009). In contrast, previous memory-based formulations suggested a comparison of current sensory input with a stored memory trace of the recent past, rather than a model-based abstraction.

As described before, ASA (e.g., Bregman, 1994) underlies basic object formation, feature integration and stream segregation and is thought to precede the sensory registration of a mismatch. Schroger et al. (2014) proposed that an auditory event representation system might encompass ASA and deviance detection alike under the assumptions of a generative model of sensory input which maintains an abstraction of the sensory regularity and assumptions of its generation and is concerned with the anticipation of future events (prediction). As argued by Winkler and Czigler (2012), many of the characteristics of ASA are similarly found in basic visual processing. As such, auditory and visual MMN might rely on similar formations of basic perceptual object representations and a model-based interpretation of MMN generation might equivalently encompass the visual MMN (Winkler and Czigler, 2012). As we have seen in the previous sections, increasing

evidence supports the view that many of the key properties of the MMN in relation to sensory regularities and stimulus expectancy are shared by visual and similarly, although less researched, by somatosensory MMRs. Therefore, in line with the suggestions by Winkler and Czigler (2012), comparable principles might underlie these responses across sensory modalities.

Expectation or adaptation?

The assumption that the MMN corresponds to a *mismatch* response resulting from a comparison between an incoming sensory signal with some form of neuronal representation (e.g. representation of a memory trace or model-based expectation) is challenged by adaptation-based accounts of MMN generation. Proponents of the view argue that any observed differences between standard and deviant responses can be explained by neuronal attenuation (or fatigue) leading to stimulus specific adaptation (May et al., 1999; Jaaskelainen et al., 2004; May and Tiitinen, 2010). Neuronal populations tuned to the standard stimulus are thought to adapt to its repeated presentation. The presentation of a deviant stimulus subsequently activates neighboring populations, not fatigued at the synaptic level (sometimes called "fresh afferents"). These populations thus elicit a stronger response which results in alterations of the sensory evoked potential (e.g. the N1 in audition) ultimately underlying any deflections in the MMN difference wave.

In animals, stimulus specific adaptation has been shown to cause response patterns similar to the MMN (e.g., Ulanovsky et al., 2003, 2004) and response reduction to repeated stimuli, more broadly termed repetition suppression (RS), is a widely studied phenomenon in neuroscience (Grill-Spector et al., 2006). However, no consensus exists regarding the exact mechanisms underlying RS (Grill-Spector et al., 2006; Auksztulewicz and Friston, 2016; Todorovic and de Lange, 2012; May and Tiitinen, 2010) while strong evidence speaks against its exhaustive explanatory power for the MMN (Naatanen et al., 2005; Garrido et al., 2009b; Summerfield et al., 2008; Wacongne et al., 2012; Auksztulewicz and Friston, 2016; Todorovic et al., 2011; Todorovic and de Lange, 2012; Heilbron and Chait, 2018). Nevertheless, an increasing MMN with prior standard repetition is a well known feature of the response (Sams et al., 1983; Näätänen, 1992; Imada et al., 1993; Javitt et al., 1998) and RS of the standard trace likely plays a role during such modulations (Baldeweg, 2006; Auksztulewicz and Friston, 2016). A theory of MMN generation ultimately has to account for the empirical features of the response and repetition sensitivity is a prominent example. Other features reviewed in the previous sections include mismatch effects in response to stimulus omissions, violations of abstract feature rules and complex or global scale sequence regularities as well as deviant responses to unexpected repetitions. These findings provide com-

pelling evidence for a role of stimulus expectancy in MMR generation and potentially during RS, although their interpretations remain disputed (May and Tiitinen, 2010). The elicitation of a deviant response to unexpected repetitions, for instance, speaks against the basic assumption of adaptation accounts that repeated stimuli induce response reduction. However, such mismatch effects in response to repetitions have been argued under adaptation assumptions to result from RS by recruitment of (high level) neuronal populations tuned specifically to alternating stimulus pairs. Sudden repetitions would thus activate a new population tuned to repetition pairs, resulting in a stronger response to repetition than alternation. Given that Wacongne et al. (2012) have shown an elicitation of the MMN to unexpected repetitions by using sequences of stimulus pairs which were separated by long ISIs of 10s, however, any adaptation effects should have been eliminated by long recovery times for the involved neuronal populations.

Strikingly, Todorovic et al. (2011) and Todorovic and de Lange (2012) showed differences for RS under expected and unexpected conditions, showing stronger RS for expected repetitions. Moreover, Todorovic and de Lange (2012) indicated a double dissociation between response suppression by stimulus repetition and by expectation, coining the term *expectation suppression* in contrast to RS: Early signals were attenuated by repetition (but not expectation), mid-latency signals (MMN window) were attenuated by expectation (but not repetition) and late signals (P3 window) were attenuated by both. Finally, Costa-Faidella et al. (2011a) showed that RS effects to repetitions with randomly varying ISIs were reduced compared to those with regular ISI, indicating that RS to repeated stimuli was modulated by temporal expectancy. Overall, while studies support a non-negligible influence of RS on evoked responses in general, and stimulus repetition is a modulatory feature of MMRs, current research is predominantly in support of the view that stimulus expectancy takes on the underlying modulatory role during MMR generation (Heilbron and Chait, 2018).

Perception as inference

The above expectation- and model-based formulations of MMN generation (Winkler, 2007; Winkler et al., 2009; Winkler and Czigler, 2012; Schroger et al., 2014) draw heavily on the general ideas of perceptual inference (von Helmholtz, 1867; Gregory, 1980; Dayan et al., 1995; Knill and Pouget, 2004). More specific formulations of a Bayesian brain hypothesize that the brain performs probability estimation corresponding to (approximate) Bayesian inference (Dayan et al., 1995; Knill and Pouget, 2004; Friston, 2005, 2010). In accordance with Bayes' rule, the brain thus combines the conditional probability of an observation y given a hypothesis about a hidden

state of the world θ, i.e., the *likelihood* $p(y|\theta)$, with the *prior* probability of the state $p(\theta)$ to obtain a *posterior* probability estimate $p(\theta|y)$ reflecting how likely this state actually drives the observations:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \tag{1.1}$$

where $p(y)$ corresponds to the probability of observing y across all possible (i.e., considered) states of the world. The exact computation of the latter quickly results in a combinatorial explosion depending on the number of considered states. Thus, as the exact Bayesian computation is often intractable, the brain is generally assumed to approximate Bayesian inference (Friston, 2010). Importantly, any iterative process of Bayesian inference uses the current posterior estimate of $p(\theta|y)$ as the prior for the next computation and thus continuously updates the probability of future observations.

An influential process theory for an implementation scheme of perceptual (Bayesian) inference in the brain is the theory of *predictive coding* (PC; Mumford, 1992; Rao and Ballard, 1999; Lee and Mumford, 2003; Friston, 2005; Friston and Kiebel, 2009). According to PC, the brain's generative model is continuously updated by comparing incoming sensory information with model predictions on all levels of hierarchical cortical organization. The brain is hypothesized to be fundamentally hierarchically organized (Felleman and Van Essen, 1991) such that neurons at each processing level receive and integrate reciprocal feedforward and feedback inputs of adjacent levels, resulting in increasing spatio-temporal scales of information representation along the cortical hierarchy (Friston, 2005; Kiebel et al., 2008). At each hierarchical level, the higher level "abstractions" project predictions (priors) to lower levels where the difference between bottom-up input and top-down predictions corresponds to a prediction error which is sent further up the hierarchy. This continuous and simultaneous procedure across all hierarchical processing stages ultimately furnishes a most likely estimate of perceptual content. Canonically, in the laminar cortical structure, feedforward connections originate in superficial layers and project to deep layers where feedback connections originate and project to superficial layers (Felleman and Van Essen, 1991). Under PC, the asymmetry between feedforward and feedback connection pathways is thought to carry prediction errors and predictions in the layered structure of the cortex, respectively (Friston, 2005). Moreover, it has been suggested that this asymmetry is reflected in a specific oscillatory profile (Arnal and Giraud, 2012; Bastos et al., 2015; Friston, 2019; van Kerkoerle et al., 2014): Feedforward prediction errors are thought to be conveyed by higher frequencies (gamma band, > 30 Hz) and top-down predictions by lower frequencies (alpha/beta band, $10 - 30$ Hz).

PC can elegantly account for many phenomena in psychology and neuroscience (Knill and Pouget, 2004; Rao and Ballard, 1999; Friston, 2005, 2010; Clark, 2013; Heilbron and Chait, 2018; Millidge et al., 2021), including mismatch responses (Friston, 2005; Garrido et al., 2009b; Heilbron and Chait, 2018; Stefanics et al., 2014; Winkler and Czigler, 2012; Wacongne et al., 2012). Under PC, neuronal responses correspond to prediction errors and the MMN reflects an increased prediction error in response to the unexpected deviant stimulus in contrast to an attenuated response to the expected standard stimulus. Therefore, PC provides a common framework for adaptation-based views highlighting the importance of RS dynamics for MMN generation as well as model-based views suggesting a comparator mechanism between the current sensory input and the recent stimulus trace. As such, the theory has the potential to unify previously opposing theories of MMN generation (Garrido et al., 2009b) while accounting for the key empirical features of the MMN discussed in the previous sections (Wacongne et al., 2012; Heilbron and Chait, 2018).

Moreover, PC provides a mechanistic explanation of processes such as attention and RS and their corresponding influences on MMRs (Auksztulewicz and Friston, 2015; Garrido et al., 2009b; Auksztulewicz and Friston, 2016). Attention corresponds to a top-down modulatory weighting of bottom-up error signals by their estimated precision, i.e., the inverse variance of the estimated distribution (Feldman and Friston, 2010). Generally, under PC, stimulus predictability of sensory regularities is accompanied by top-down modulations of synaptic gain (Auksztulewicz et al., 2017). Increased precision weighting (by means of synaptic gain modulation) has been shown to correspond to the MMN increase during attentional focus (Auksztulewicz and Friston, 2015) as well as RS responses to repeated stimuli (Auksztulewicz and Friston, 2016). Overall, although PC is critically discussed (Egner and Summerfield, 2013) and its empirical evaluation has been questioned (Kogo and Trengove, 2015), the theory offers a principled framework under which many different aspects of MMR dynamics can be accounted for, including their ubiquitous occurrence across sensory modalities.

Summary

Taken together, it is generally concluded that MMRs are a reflection of the brain's sensitivity to the predictable structure of the sensory environment and that the brain capitalizes on stimulus expectancy by efficient processing of redundant (expected) sensory inputs. While the neuronal implementation and exact computation is subject of ongoing research and debate, the brain is broadly hypothesized to perform some form of predictive processing and probabilistic inference under which MMRs correspond to signatures of mismatch between sensory inputs and the brain's

generative model of the sensory environment.

1.3 Utilizing MMRs to assess the brain's generative model

1.3.1 The roving stimulus paradigm - A probabilistic model environment

In order to probe brain responses to probabilistic input and assess the underlying expectations of the system, sequential stimulation can be used to form sensory regularities. As we have seen, MMRs elicited by sequential stimuli can carry information about the degree to which these inputs were anticipated by the system.

The classic paradigm to study MMRs is the oddball paradigm in which the deviation from a regularity is directly coupled with a specific stimulus property since the item frequencies of the presented stimuli are unequal. In order to decouple the response to regularity violation from the response to a specific stimulus feature many studies have switched the roles of the standard and deviant features for each experimental block and average the deviants and the standards across blocks to obtain the MMR signatures. However, although indeed reducing confounds of stimulus features in the deviant response, it remains that the item frequencies within each sequence are biased, potentially increasing influences of long term RS to the more frequent stimulus. The roving stimulus paradigm makes use of this principle of switching roles of standard and deviant features while additionally keeping the item frequency equal (Cowan et al., 1993; Baldeweg et al., 2004; Garrido et al., 2008). In the standard version of the paradigm consisting of binary sequences, stimuli alternate between two different features (e.g. low and high intensity with overall $p = 0.5$) such that the deviant is defined as the first stimulus that breaks a train of repeating (standard) stimuli. With repetition, the deviant subsequently becomes the new standard, defining a new train of stimulus repetitions.

Roving stimulus sequences can be created by defining a generative process probabilistically producing observations. In contrast to a generative model (e.g., of the brain), a generative process describes the true states of the world governing observations[1]. A generative process might result in observations o at time t which evolve according to a Markov chain, such that each observation depends on the previous observation and the current state of the world s:

$$p(o_t|o_{t-1}, s_t) \tag{1.2}$$

The resulting sequence of observations is thus goverened by transition probabilities specifying

[1]Therefore, a different notation will be used for the formulations of generative models and generative processes.

the repetitions and alternations of observations. The dynamics of the hidden state s evolving over time with $p(s_t|s_{t-1})$ might either be specified such that s remains constant over time (reflecting a static hidden state, e.g. by a state transition matrix corresponding to the identity matrix), or such that occasional changes in the probability distributions governing observations are introduced. If s is static, sequences are fully described by the probability of observation transitions. If s is dynamic, occasional change-points in the observation sequence might reflect changes in the underlying probabilities.

Correspondingly, an observer of such sequences might infer on the probabilistic structure underlying sequential observations. As such, the study of MMRs to probabilistic observations embeds in the broader range of the neuroscience of probabilistic inference and decision making under uncertainty (Behrens et al., 2007; Nassar et al., 2010, 2012; Yu and Dayan, 2005) and roving stimulus sequences become a special case of model environments employed in these fields (Modirshanechi et al., 2022). Therefore, probabilistic sequences can be used to study the brain's generative model during exposure with the sensory (model) environment. For example, sequences of observations can model volatile environments if a dynamic hidden state governs occasional change points at which the generative process switches to a different set of probabilities specifying the sequential observations. Similarly, a multi-modal environment might be specified by defining a generative process that produces multiple observation sequences governed by cross-modal interdependence to probe probabilistic inference across senses (see chapter 2).

Overall, the principle of probabilistic sequence generation provides an elegant way to create a) suitable sequences for the roving stimulus paradigm by defining the probability of stimulus transitions independently from stimulus features and b) sensory environments governed by different probabilistic regularities. Correspondingly, MMRs might be investigated as signatures of a passive observation of such probabilistic environments and can thus provide insights into the brain's generative model.

1.3.2 Computational modeling of brain responses

In the past decade, the investigation of MMRs and their underlying mechanisms has moved from average-based analyses of mismatch signatures to modeling approaches capturing more nuanced dynamics of brain responses. In their seminal work, Garrido et al. (2008, 2009b, 2007a,b, 2009a) used dynamic causal modeling (DCM) to show that adaptation-based and model-based theories alone are not sufficient to account for the dynamics of the MMN. DCM is concerned with the comparison of different implementations of biologically informed network models in order to

draw conclusions on the underlying connectivity dynamics. The authors were able to show that influences of both MMN hypotheses underlie mismatch processing in the auditory system under predictive coding implementations of a fronto-temporal network.

Other modeling approaches concern the comparison of model signatures corresponding to an hypothesized underlying process, in order to evaluate their contribution to account for brain dynamics. As we have seen, MMRs such as the MMN and the P3 have been proposed to reflect a neuronal expression of error or mismatch between the current and the expected (or predicted) sensory input *given* the brain's generative model. As such, the responses are thought to carry information about the brain's inference process on the statistical structure of the environment. A growing branch of MMR research employs computational models which perform iterative (Bayesian) inference on sequential input to account for the dynamics of brain responses during sequence perception. Bayesian modeling of EEG data has been used to inspect the single-trial dynamics of the MMN in the auditory (Lieder et al., 2013; Weber et al., 2020; Maheu et al., 2019), visual (Stefanics et al., 2018; Schlossmacher et al., 2022) and somatosensory modality (Ostwald et al., 2012) and the P3 across different senses (Kolossa et al., 2012, 2015; Kopp et al., 2016; Wessel and Huber, 2019; Seer et al., 2016; Modirshanechi et al., 2019; Mars et al., 2008; Maheu et al., 2019; Ostwald et al., 2012). In the following, the basic principles of Bayesian inference models will be briefly introduced and different approaches to the inference on sequential input will be highlighted and categorized.

Inference models

MMRs are classically investigated in paradigms using sequences with two different stimuli, such as the oddball paradigm or the roving stimulus paradigm. If we consider a sequence with binary observations $Y = \{0, 1\}$, the inference process about the hidden state governing the evolution of observations from y_{t-1} to y_t might be modeled with a Beta-Bernoulli (BB) model which estimates probability $\theta \in [0, 1]$. Given a sequence of samples $y_1, ..., y_t$ the model combines the likelihood $p(y_t|\theta)$ with the prior belief $p(\theta)$ in order to refine a posterior estimate over the latent variable $p(\theta|y_t)$. The BB model corresponds to a *conjugate* Bayesian model, with conjugacy referring to the fact that the posterior distribution is in the same probability distribution family as the prior distribution which allows iterative combination of prior and likelihood over time in accordance with Bayes rule. In the Beta-Bernoulli case, the prior is modeled as a Beta distribution $Beta(\theta; \alpha, \beta)$ with α and β corresponding to concentration parameters keeping count of the occurrences of observation categories ($y = 0; y = 1$). The likelihood of observations is modeled as a Bernoulli

distribution $Bern(y;\theta)$. With each new observation, the parameters of the belief distribution over θ are updated and over time the iterative process converges to the true hidden state probability distribution.

The Dirichlet-Categorical (DC) model is a conjugate extension of the BB model for sequences with more than two unique observations ($Y = \{1, ..., M\}$). It models the likelihood of the observations with the Categorical distribution with $\{1, ..., M\}$ different possible realizations per sample y_t such that $y \sim Cat(\theta_1, ..., \theta_M)$. The prior over hidden states is modeled by the Dirichlet distribution with $\theta_1, ..., \theta_M \sim Dir(\alpha_1, ..., \alpha_M)$, where α corresponds to the concentration parameters keeping count of observation categories.

In models such as BB or DC, the representation of the latent variable θ is modeled as a discrete distribution over categories (of unique samples) and usually remains static over time. As such, with increasing amount of considered samples over sequence-time, the converging posterior becomes inflexible. In case of volatile environments with sudden changes in the hidden state, these models are not well adaptive. Therefore, such models might account for changing environments by "forgetting" the recent past, without explicit representation of a volatile hidden state. A temporal integration parameter $\tau \in [0, 1]$ can implement a gradual discounting of observation-samples the further in the past they occurred by means of exponential down-weighting (thereby functioning as window of observation integration). Implementations of forgetting or "leaky integration" have been widely used in the modeling of brain function (Farashahi et al., 2017; Glaze et al., 2015; Ossmy et al., 2013) and the dynamics of MMRs (Ostwald et al., 2012; Maheu et al., 2019).

The Hidden Markov model (HMM) explicitly represents a hidden state θ as evolving according to a Markov chain with $p(\theta_t|\theta_{t-1})$ described by a transition matrix $A \in \mathbb{R}^{N\times N}$ with N being the number of modeled hidden states. In contrast to BB or DC models, the HMM provides an explicit representation of environmental volatility by maintaining separate beliefs about discrete hidden states governing the sequence statistics. At each time-point, observations are modeled as being emitted from the hidden state with $p(y_t|\theta_t)$. The parameters of the HMM are commonly inferred using an Expectation-Maximisation algorithm (Dempster et al., 1977; Baum et al., 1970).

Representation of sequence statistics

Given a specific sequence of observations $o_1, ..., o_T$, a Bayesian model might consider different statistics of the sequence for making inference:

Stimulus probability (SP) inference does not capture any Markov dependencies between observations and models the observation identity as input samples with $y_t = o_t$ for $t = 1, ..., T$. In

case of the DC model, SP therefore corresponds to counting the observation frequency. Sensitivity of MMRs to SP has been investigated in terms of the oddball paradigm in which the deviant is defined by the rarity of a feature (Naatanen et al., 2007; Garrido et al., 2009b) as well as in terms of computational models specifically implementing SP inference (Mars et al., 2008; Lieder et al., 2013; Kolossa et al., 2012).

Alternation probability (AP) inference captures a limited form of first order Markov dependency by modeling observation alternations and repetitions as input samples $y_t = a_t$ for $t = 2, ..., T$ with $a_t = 1_{o_t \neq o_{t-1}}$, where a_t indicates an alternation and takes on the value 1 if the current observation o_t differs from the previous observation o_{t-1}. As discussed in section 1.1, MMRs are known to be sensitive to unexpected stimulus repetitions (Tervaniemi et al., 1994; Czigler et al., 2006a) which might be reflective of AP inference. Moreover, studies which manipulated AP independent of SP indicated some specificity of brain responses to unexpected stimulus alternations in particular (Summerfield et al., 2008, 2011; de Gardelle et al., 2013).

Transition probability (TP) inference captures first order Markov dependencies by modeling each possible transition of observations with a set of parameters $\theta^{(i)}$ for each unique transition from $o_{t-1} = i$ where input samples correspond to $y_t = o_t$ for $t = 1, ..., T$. The TP model can be extended to capture second order Markov dependencies with a second set of parameters $\theta^{(j)}$ specifying transitions from $o_{t-2} = j$. Converging findings from modeling studies indicate that the perceptual system specifically estimates the probability of transitions between observations in a sequence rather than, e.g., SP or AP (Meyniel et al., 2016; Maheu et al., 2019). This is in line with research showing that MMN and behavioural responses were suggestive of an estimation of TP (Mittag et al., 2016; Koelsch et al., 2016) and that the processing of sensory sequences primarily relies on their transitional structure (Dehaene et al., 2015; Meyer and Olson, 2011; Meyer et al., 2014). However, only few studies directly compared the estimation of TP with other sequence statistics, such as SP or AP which have similarly been shown to be relevant statistics (Mars et al., 2008; Lieder et al., 2013; Wang et al., 2015) and can be subsumed under TP estimation.

Model read-out

As we have seen, Bayesian models can be utilized to implement different approaches to incorporate sequential information to infer the probability of new observations. The models provide a read-out of estimated quantities as the sequence unfolds over time and the read-out can be used to compare the model dynamics with brain data, e.g., by means of regression analyses. Different read-out functions have been implemented by modeling studies highlighting different aspects of

the inference process (Lieder et al., 2013):

Novelty detection formulations quantify the hypothesis that MMRs reflect the degree to which a new observation is surprising or novel (Squires et al., 1975; Escera and Corral, 2007; Tiitinen et al., 1994; Lieder et al., 2013) in contrast to basic change detection assumptions. In the information theoretic concept of surprise introduced by Shannon (1948), an observation y_t is surprising and therefore informative if it is associated with low probability of occurrence. Shannon's surprise or predictive surprise (PS) is thus defined as the negative logarithm of the probability of observing y_t given the prior belief about observations:

$$PS(y_t) = -ln\, p(y_t|\theta_t) = -ln\, p(y_t|y_{t-1}, ..., y_1) \tag{1.3}$$

Novelty detection in terms of surprise reflects a basic form of mismatch between sensory input and internal model belief and, as such, is closely related to the concept of prediction error (Friston, 2005). Previous modeling approaches have used the PS formulation of novelty to describe MMRs primarily in terms of the P3 (Mars et al., 2008; Kolossa et al., 2015; Kopp et al., 2016). In contrast to formulations of prediction error in predictive coding inference schemes, surprise is an undirected quantity. Directed formulations of prediction errors scaled by the precision of the model (inverse variance of the belief distribution) are implemented in modeling frameworks such as the hierarchical Gaussian filter (Mathys et al., 2011, 2014) and have been shown to fit the dynamics of visual MMN responses better than formulations of categorical change detection (Stefanics et al., 2018) or adaptation (Schlossmacher et al., 2022).

Belief commitment formulations additionally introduce quantifications of confidence into surprise formulations of novelty detection similar to precision weighting of prediction errors. Confidence-corrected surprise (CS) quantifies the concept that a low-probability observation is more surprising if the commitment to the belief is high (Faraji et al., 2018): CS has been shown to be similar to PS as it corresponds to a surprise formulation of an initial mismatch between input and model belief but additionally considers the commitment of the generative model to a belief and scales with the negative entropy of the prior distribution (which corresponds to a quantification of the variability or redundancy of the distribution). CS is defined as the Kullback-Leibler (KL) divergence quantifying the difference between the (informed) prior distribution at the current time step t and a flat (i.e., uninformed) prior distribution $\hat{p}(\theta_t)$ which is updated only with the most recent observation y_t:

$$CS(y_t) = KL(p(\theta_t)||\hat{p}(\theta_t|y_t)) \tag{1.4}$$

As a relatively new surprise formulation (Modirshanechi et al., 2022; Faraji et al., 2018), CS has only seen little empirical evaluation (Modirshanechi et al., 2019). However, confidence has been shown to be an important computational concept shaping probabilistic inference (Meyniel et al., 2015; Meyniel and Dehaene, 2017; Heilbron and Meyniel, 2019).

Model update formulations quantify the hypothesis that MMRs reflect the degree to which a new observations triggers an adjustment of the generative model (Winkler, 2007; Winkler et al., 2009; Lieder et al., 2013). Bayesian surprise (BS) quantifies this update of the generative model of a Bayesian observer to incorporate new observations (Itti and Baldi, 2009; Baldi and Itti, 2010) and is defined as the KL divergence between the belief distribution prior and posterior to the model update:

$$BS(y_t) = KL(p(\theta_{t-1}|y_{t-1}, ..., y_1)||p(\theta_t|y_t, ..., y_1)) \tag{1.5}$$

BS has been shown to account for single-trial dynamics of the P3 (Kolossa et al., 2015; Mars et al., 2008; Seer et al., 2016; Ostwald et al., 2012) as well as the auditory (Lieder et al., 2013) and somatosensory MMN (Ostwald et al., 2012), and provides a well-established quantification of model-adjustment dynamics.

Descriptive formulations quantify the observed behaviours of MMRs at a phenomenological level, as termed by Lieder et al. (2013), without the implementation of probabilistic inference processes. These might include implementations of novelty detection or neuronal adaptation without aspirations of biological plausibility. An adaptation model might be designed as the decay of response to an input, thus capturing the well-known property of the MMN to increase in amplitude as a function of the number of prior standard repetitions (train-lengths; Haenschel et al., 2005; Cowan et al., 1993). Lieder et al. (2013) for instance showed that such an adaptation model was outperformed by models performing Bayesian inference. Similarly, the authors used variants of basic novelty-detection models to capture the occurrence of a stimulus change in a binary fashion to test it against more dynamic Bayesian models of novelty detection. Alternatively, a descriptive novelty-detection model might be designed such that the "prediction error" increases proportional to the repeated standard stimulus trace (Lecaignard et al., 2022).

In our work, we designed a descriptive *train-length* (TL) model which directly reflected our empirical observation that MMRs of all senses increase with train-lengths (see chapter 2) which we showed to be driven by a reciprocity in the response to standards (becoming more positive) and the response to deviants (becoming more negative). Therefore, the TL model was designed to capture repetition suppression as decay of responses to repeated stimuli as well as the increase

in prediction error proportional to the repeated standard train. As such, the TL model takes into account these observed dynamics of MMRs without invoking any assumptions of underlying inference processes. It tracks the stimulus train lengths c for a given sensory input by counting observation repetitions: $c_t = r_t(c_t + r_t)$, where $r_t = 1_{y_t=y_{t-1}}$ takes on the value 1 whenever the current observation is a repetition of the previous observation y_{t-1} and $r_t = 0$ resets the current train length to zero. To form single-trial predictors of the EEG data, the model outputs values that increase linearly with train length and have opposite signs for standards and deviants:

$$TL(y_t) = \begin{cases} -c_t & \text{if } r_t = 1 \\ c_{t-1} & \text{if } r_t = 0 \end{cases} \tag{1.6}$$

Empirical model selection

The computational models described above take sequences of observations (trials) as inputs and provide trial-wise estimates which are read out, e.g., as surprise values. These model values over trials correspond to the regressors which are ultimately fit to single-trial brain data obtained by, e.g., EEG to draw inferences on the explanatory power of a specific model implementation for the dynamics of brain responses. Model fitting can be implemented as variational inversion schemes (Beal, 2003; Penny et al., 2003, 2005; Flandin and Penny, 2007) providing model evidences in terms of Free Energy approximation to (log) model evidence for each participant and model (as well as data samples). Free Energy quantifies a trade-off between model complexity and model fit (Beal, 2003) and has been shown to outperform other approaches used for model comparison (Penny, 2012). The comparison of different models across participants can subsequently be performed in group level Bayesian model selection procedures (BMS; Stephan et al., 2009; Rigoux et al., 2014):

Two approaches to BMS can be performed, depending on the underlying assumption of the model distribution in the population. In fixed-effects BMS, it is assumed that a given model underlies the data generation in *all* participants (and the general population), whereas random-effects BMS treats the underlying models as random effects which can vary across the set of participants and correspond to a sample of the fixed (unknown) set of models in the general population. As suggested by Stephan et al. (2010), a fixed-effects BMS procedure might only be selected if the model of interest concerns a basic physiological mechanism. Since it is unknown if the brain relies on a fixed implementation of a specific generative model to infer the structure of sensory regularities or if the model might vary between model environments or participants, an empirical

approach of random-effects BMS is advisable to draw inferences on the model used by the brain. The approach provides (protected) exceedance probabilities quantifying the frequency of a model in the participant sample over all other considered models. Moreover, within the random-effects BMS framework, model evidences can be grouped into families of models to compare families against each other and draw inferences on specific features shared across the model family (Penny et al., 2010).

Summary

Computational modeling provides a principled framework for the assessment of computational strategies and inference processes underlying sequence perception. Different model implementations allow to test diverging hypotheses about (properties of) the generative model underlying mismatch processing in the brain. Model comparison procedures can be employed to select the most likely model in the tested population.

1.4 Aim of the thesis

The aim of the thesis was to inspect MMRs across sensory modalities in response to probabilistic stimulus sequences to compare their response dynamics and probe the brain's generative model of its sensory environment. To this end, we used the roving stimulus paradigm to create model environments and recorded MMRs by means of EEG and fMRI. Given that the somatosensory system has received the least attention of the modalities reviewed here, in study 1, we first aimed for a thorough investigation of somatosensory MMRs by use of computational modeling of single-trial EEG data in sensor and source space. As investigations of the underlying computational principles of somatosensory MMRs are rare, we employed a volatile roving stimulus sequence to compare a variety of signatures of Bayesian inference and assess the brain's generative model in response to a changing sensory environment. In studies 2 and 3, for a comparative inspection of MMRs in the auditory, visual and somatosensory modality, we developed a tri-modal version of the roving stimulus paradigm to delineate modality specific and modality general features of mismatch processing. The paradigm served to inspect effects of deviance detection and repetition suppression as well as additional responses to sequence features with cross-modal predictive properties. In study 2, we used EEG to compare the spatio-temporal dynamics of mismatch processing across the senses. In addition to a dissection of the early MMN and the later P3 responses across modalities, we employed computational modeling to inspect MMRs on a single-trial basis and compare their dynamics to signatures of Bayesian inference models implementing different computational

strategies of uni- and cross-modal sequence processing. In study 3, we used fMRI to differentiate and locate MMRs of each sensory modality to the tri-modal sequences and investigate the neuronal generators underlying modality-specific and -general processing as well as the processing of expectation violations with respect to the cross-modal contingencies in the sequences. In the next chapter, I will briefly introduce the specific paradigms and methodology employed for our study of MMRs across the senses and summarize the empirical work and most important findings.

CHAPTER 2

Summary of empricial studies

2.1 Methodology

For the empirical evaluation of brain responses to probabilistic input, we implemented two different roving stimulus paradigms, depicted in figure 2.1 and further described below. The corresponding sequences reflect different sensory environments and enable the inspection of MMRs to unexpected stimulus transitions independent of their equiprobable features. Study 1 makes used of volatile roving stimulus sequences and for studies 2 and 3 we designed a tri-modal version of the roving stimulus paradigm. Generally, these paradigms serve to investigate the probabilistic representation of stimulus transitions which is hypothesized to underlie sequence perception and statistical learning (Dehaene et al., 2015; Mittag et al., 2016; Meyniel et al., 2016; Maheu et al., 2019). Moreover, the resulting stimulus sequences provide suitable discrete inputs for Bayesian inference models which operate on sequential observations to iteratively infer their underlying probability. Finally, on a higher level of description, the paradigm produces trains of repeating stimuli of variable lengths (see fig. 2.1A) which are well suited to investigate repetition effects on MMRs, deviant and standard responses and aid the comparison with the existing body of basic MMR research (Baldeweg et al., 2004; Baldeweg, 2006; Haenschel et al., 2005; Costa-Faidella et al., 2011a; Cooper et al., 2013).

Roving stimulus paradigms

Volatile roving stimulus sequences

In study 1 we designed a volatile roving stimulus paradigm to inspect MMRs in the somatosensory modality. The generative process for the stimulus sequences (see fig.2.1A) corresponds to a state s evolving according to a Markov chain with $p(s_t|s_{t-1})$. The state transition matrix $\mathbf{A} \in \mathbb{R}^{2\times2}$ was set such that a switch in sequence statistics occurred with $p = 0.01$. Observations $o \in O = \{0, 1\}$ were emitted with $p(o_t|o_{t-1}, o_{t-2}, s_t)$ such that each observation depended on the previous two

observations (second order Markov dependency; not shown in the graphical model of 2.1) as well as the current hidden state. Observations were used for binary somatosensory stimulus sequences with low and high intensity $y \in Y = \{0, 1\}$. This additional layer (Y) simply represents the fact that observations generated by the Markov process can be used for different forms of empirical stimulus sequences (such as other modalities or other stimuli). The two possible latent states[1] corresponded to a fast switching regime which produced observations with relatively frequent alternations and a slow switching regime which resulted in longer trains of repeated stimuli in the sequence. In the slow switching regime, transition probabilities of observations associated with repetitions were higher than transition probabilities specifying a change: Depending on the previous two stimulus configurations, in the slow switching regime the repetition probabilities corresponded to $p(repetition) = 0.65$ and $p(repetition) = 0.85$, while in the fast switching regime they corresponded to $p(repetition) = 0.3$ and $p(repetition) = 0.75$. The resulting volatile roving stimulus sequences were used to probe the sensitivity of the brain's generative model to adapt their estimated statistics to sudden changes in the environment. Overall, the stimulus probability of low and high stimuli was balanced ($p = 0.5$) and at the start of each sequence the hidden states had equal probability ($p = 0.5$).

Tri-modal roving stimulus sequences

For studies 2 and 3 we designed a novel tri-modal roving stimulus paradigm to inspect MMRs in the auditory, visual and somatosensory modality. The generative process for the creation of our tri-modal roving stimulus paradigm[2] is depicted in fig.2.1B and corresponds to a single state s which remained static over time (state transition matrix corresponding to the identity matrix $\mathbf{A} = \mathbf{I}$). The observations $o \in O = \{0, ..., 7\}$ were emitted with emission probability $p(o_t|o_{t-1}, s_t)$ exhibiting first order Markov dependency. For the tri-modal sequences, the set of possible observations was used to create $N = 2^3 = 8$ different stimulus combinations for binary sequences $\{y^A, y^V, y^S\} \in Y = \{0, 1\}$, corresponding to roving stimulus sequences with low and high intensity stimuli for the auditory (A), visual (V) and somatosensory (S) modality. The observation transitions were governed by an transition matrix $\mathbf{B} \in \mathbb{R}^{N \times N}$. Therefore, although corresponding to separate sensory input streams, stimulus changes and repetitions in each modality depended on the preceding tri-modal observation. This embedding of cross-modal probabilitic dependencies across the sequences enabled the inspection of mismatch processing with respect to global

[1] The specific implementation in the publication included a third hidden state deterministically emitting a catch trial which is not described here.

[2] The description used here differs to the publication in order to enable the presentation of both roving stimulus paradigms with the same notation.

multi-modal sequence statistics. As each stimulus transition in each modality was conditional on the previous tri-modal stimulus configuration, the uni-modal sequences exhibit a global property in terms of the multi-modal context of stimulus transitions. The probabilistic model was defined such that, in each modality, repetitions ($p(repetition) = 0.825$) were more likely than changes ($p(change) = 0.175$) and at each time-step only one modality could change. Moreover, to balance the sequences in terms of stimulus transitions in each modality, identical probabilities were specified across modalities and balancing of stimulus features across the sequence was ensured by keeping the observation transition matrix symmetric ($\mathbf{B} = \mathbf{B}^T$). The global sequence property was used to manipulate the cross-modal predictability of stimulus transitions in the sequences as follows:

At each time-point, from the perspective of any one of the three modalities, the other two modalities can be congruent (both of low or high intensity) or incongruent (one is low and the other high). A transition is sampled according to $p(y_t^A, y_t^V, y_t^S | y_{t-1}^A, y_{t-1}^V, y_{t-1}^S, s_t)$. For example, let a transition expressed as $[100|000]$ indicate a unimodal auditory change ($y_t^A = 1, y_{t-1}^A = 0$) with repeating visual and somatosensory modalities ($y_t^V = y_{t-1}^V = 0$ and $y_t^S = y_{t-1}^S = 0$). Three different types of stimulus sequences were generated with different settings specifying the transition probabilities of each modality given the arrangement of the stimuli in the other two modalities (i.e., either congruent or incongruent). One setting defined lower change probability if the other two modalities were *congruent* (e.g. for any auditory (y^A) change from $t-1$ to t, y^S and y^V were congruent with $p(\textit{change}|\textit{congruent}) = p(100|000) = p(000|100) = p(111|011) = p(011|111) = 0.025$ and y^S and y^V were incongruent with $p(\textit{change}|\textit{incongruent}) = p(101|001) = p(001|101) = p(110|010) = p(010|110) = 0.15$)). Correspondingly, the second setting defined lower change probability if the other two modalities were *incongruent* ($p(\textit{change}|\textit{congruent}) = 0.15$, $p(\textit{change}|\textit{incongruent}) = 0.025$). The third setting defined equal change probabilities whether the other two modalities were *congruent* or *incongruent* ($p(\textit{change}|\textit{congruent}) = 0.0875$, $p(\textit{change}|\textit{incongruent}) = 0.0875$). Per definition, the repetition probability follows the same principle with $p(\textit{repetition}) = 1 - p(\textit{change})$. With a terminology inspired by Arnal and Giraud (2012), the resulting stimulus transitions for each modality within the different sequences can therefore be defined as being either *predicted* (here: higher transition probability conditional on congruency/incongruency of the other modalities), *mispredicted* (here: lower transition probability) or *unpredictable* (here: equal transition probability). Fig.2.1B shows the principle of trial definition for an examplary sequence.

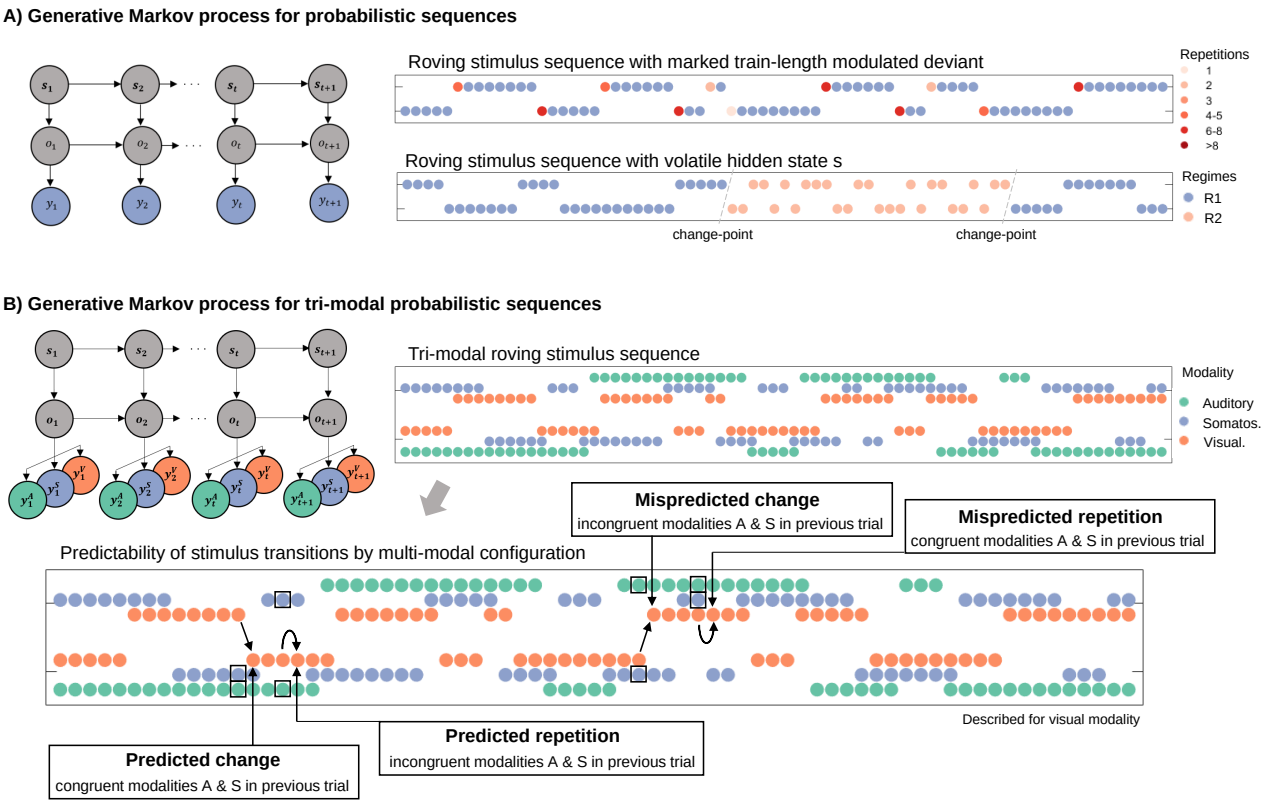


Fig. 2.1 Probabilistic roving stimulus sequences. A) Generative Markov process for probabilistic sequences. **Left:** Graphical model. Arrows reflect to conditional dependence. First order Markov dependence is depicted. **Right:** Corresponding roving stimulus sequences consisting of trains of repeating stimuli which alternate between two different features. Trains of different lengths can be used to categorize stimuli depending on the number of prior repetitions (here, deviant colored in shades of red). If the Markov process includes change-points, these can be used to introduce volatility in the sequence, here by phases of faster stimulus alternations resulting in shorter train lengths (Regime 2; light pink). **B)** Generative Markov process for tri-modal probabilistic sequences. **Left:** Graphical model. Arrows reflect to conditional dependence. **Right:** Corresponding roving stimulus sequences in each modality (Auditory: green; Visual: purple; Somatosensory: orange) consist of trains of repeating stimuli alternating between two different features. **Bottom:** Specification of the probabilities governing the Markov process can be used to introduce predictability of stimulus transitions by the multi-modal configuration. Exemplary shown for a sequence of somatosensory modality (orange) with specific probability setting: Transition probabilities determine that for each modality a change is less likely if the other two modalities are incongruent (and more likely if they are congruent) and v.v. for repetitions ($p(repetition) = 1 - p(change)$).

Computational models

In a uni- (study 1) and a multi-modal (study 2) EEG study we employed computational modeling and model comparison procedures in order to draw inferences on the computational principles underlying single-trial response dynamics and ultimately the generation of MMRs. For any of the Bayesian model implementations we utilized and compared three different surprise functions as read-out for the model parameters capturing different hypotheses on MMR generation: Predictive surprise (PS) reflecting (probabilistic) novelty detection, confidence-corrected surprise (CS) reflecting novelty detection scaled by model commitment and Bayesian surprise (BS) reflecting model updating.

Inference in a volatile environment

In study 1, utilizing the somatosensory roving stimulus paradigm with occasional change-points, we aimed for a thorough investigation of different aspects of the generative model. First, we compared different strategies to infer on observations in a volatile environment by comparing a Dirichlet-Categorical (DC) model which implemented flexibility with regards to the estimated probabilities by means of varying degrees of forgetting with a Hidden Markov model (HMM) which directly represented the volatility of the hidden state. Additionally, we compared different modes of inference on the sequence statistics: Stimulus probability (SP) inference reflecting sensitivity to the occurrence frequency of stimulus features, alternation probability (AP) inference reflecting sensitivity to stimulus changes and repetitions as well as transition probability (TP) inference reflecting sensitivity to the transitional structure of input sequences.

Inference in a multi-modal environment

In study 2, utilizing the tri-modal roving stimulus paradigm, we first compared a descriptive train-length (TL) model capturing the observed dynamics of MMRs with Bayesian DC models in a family-wise model comparison. Further, we compared a generative model performing uni-modal (UM) inference which was restricted to the binary sequential inputs of a single modality with a model performing cross-modal (CM) inference which was informed by the multi-modal context.

The CM model corresponds a modified version of the DC model in order to capture the cross-modal dependencies embedded in the tri-modal roving stimulus sequences, i.e., the alternation and repetition probabilities conditional on the tri-modal stimulus configuration. The dependencies in the sequences were balanced between stimulus identities but provided information about the probability of repetition or alternation conditional on the congruency of the other modalities. The CM

model operates on sequence input from a specific modality while taking into account the configuration on the other two modalities (i.e., the multi-modal context): The model estimates alternation probabilities (as described in section 1.3.2) with separate sets of parameters $\theta^{(i)}$ when the other modalities are congruent and $\theta^{(j)}$ when the other modalities are incongruent. Therefore, while the uni-modal implementation estimates the probability of stimulus alternations within a modality, the CM model estimates this probability within a modality conditional on the congruency of the other modalities. As such, the CM model provides a minimal implementation of a Bayesian model that captures the cross-modal dependencies in the tri-modal sequences.

2.2 Study 1

Gijsen*, S., **Grundei*, M.**, Lange, R. T., Ostwald, D. and Blankenburg, F. (2021). **Neural surprise in somatosensory Bayesian learning.** *PLoS computational biology, 17(2)*, e1008068.
*equal contribution

With study 1 we aimed to dissect the computational dynamics reflected in *somatosensory* MMRs. We employed volatile roving stimulus sequences representing a changing environment in order to draw conclusions on the underlying generative model. To this end, we recorded EEG from 40 participants and investigated MMRs in terms of ERP analyses and single-trial modeling in sensor and source space. We focused on the somatosensory domain as it is the least studied sensory system with regards to MMRs in general and especially in terms of their relation to perceptual inference. A previous study has shown that a Bayesian observer model estimating SP (observation frequencies) can account for the dynamics of sMMN and P3 indicating that these responses reflect signatures of model updating in terms of Bayesian surprise (Ostwald et al., 2012). However, as the study did not compare different probabilistic models with respect to other read-out functions or other estimation of sequence statistics such as AP or TP, the conclusions about the specific computational principles underlying somatosensory MMRs were limited. Therefore, we employed a comprehensive model comparison scheme to describe the cortical dynamics of the somatosensory system using different implementations of a generative model capturing environmental volatility (DC and HMM) as well as neural surprise signatures (PS, CS, BS).

By means of ERP analyses we identified three time windows of MMRs, an early positive MMR peaking around 60ms, an sMMN response peaking around 120ms and a P3 MMR peaking around 360ms. The three responses showed differential modulation by stimulus repetition. While the early MMR was not sensitive to the amount of prior stimulus repetition, the standard and deviant

responses in the sMMN time-window both showed repetition modulation in opposite directions and the late P3 MMR showed deviant increases with prior standard repetition. Dipole modeling was used to further investigate responses in the somatosensory system during sMMN processing (up to 200ms) in SI and bilateral SII, indicating SI as the primary driver for the earliest MMR and a combined response of SI and SII for the sMMN.

Using Bayesian model selection, we compared observer models which explicitly represented the environmental volatility by modeling a latent hidden state which governs the sequential observations (HMM) with a non-hierarchical observer model with a static hidden state (DC model). Our results showed that single-trial dynamics were best described by the DC model using a restricted window of integration over past observations (50% weighting within the last 25 to 100 observations). Thus, the system appeared to capture a changing environment, represented by the volatile roving stimulus sequences, using a static latent state model that integrates evidence on a restricted timescale. Moreover, within the model space, we compared differences in the estimation of sequence statistics in terms of SP, AP and TP. The modeling results indicated that the system performs TP inference (i.e., tracking transitions between observations) and likely does not go beyond first-order TP estimation. Furthermore, we compared different signatures of surprise computation in sensor and source space. We provided a dissociation of signatures corresponding roughly to belief inadequacy as represented by CS and belief updating as represented by BS. At the electrode level, from around 70ms on, early surprise effects represented by CS best explained the EEG data on electrodes contralateral and subsequently ipsilateral to the stimulation site up to around 200ms, corresponding to the sMMN response. Additionally, some weaker indications were provided for BS around 140ms within the sMMN time-window and around 300ms corresponding to the P3 MMR. The dipole responses at the source level indicated that bilateral SII represented signatures of model-inadequacy as reflected by CS from 70ms on and SI showed weak indications for model updating signatures as reflected by BS from around 140ms. This dissociation is compatible with the idea that early surprise signals may control subsequent model update rates.

In conclusion, we show that signals of early somatosensory processing can be accounted for by (surprise) signatures of Bayesian perceptual inference. The system appears to capture a changing environment using a static latent state model that integrates evidence on a relatively local timescale and estimates transition probabilities of observations using first order dependencies. In turn, we provide evidence that the estimated statistics are used to signal early model inadequacy scaled by confidence (CS) in SII and weak indications for model updating (BS) in SI as well as a tendency for the P3 MMR to reflect model updating.

2.3 Study 2

Grundei, M., Schröder, P., Gijsen, S. and Blankenburg, F. (2023). **EEG mismatch responses in a multi-modal roving stimulus paradigm provide evidence for probabilistic inference across audition, somatosensation and vision.** *Human Brain Mapping*, 1– 25.

MMRs such as the MMN and P3 have provided valuable insights into the brain's processing of regularities and the generation of corresponding sensory predictions (Naatanen et al., 2001; Paavilainen, 2013; Schroger et al., 2014; Yaron et al., 2012; Squires et al., 1975; Bekinschtein et al., 2009; Wacongne et al., 2011; Dehaene et al., 2015; Heilbron and Chait, 2018). However, direct comparisons of MMR between modalities are rarely provided and the extent to which the MMN and P3 reflect uni- and multi-modal regularity violations are not well understood. Therefore, in study 2, we recorded EEG from 34 participants undergoing the novel tri-modal roving stimulus paradigm. We aimed to investigate the (spatio-)temporal response profiles of MMRs across the auditory, visual and somatosensory modality during probabilistic multi-modal sequence processing and compare how different computational models can account for single-trial EEG dynamics.

First, we were able to extract modality specific and modality independent aspects of mismatch responses in all three modalities from the EEG recorded during simultaneous input streams of the modality specific sequences. Using basic ERP analyses, we found comparable but modality specific MMN signatures between 100-200ms, corresponding to the fronto-central auditory MMN, the occipito-temporal vMMN and the temporal sMMN. Although with variable latencies, across sensory modalities a central P3 response was identfied between 200-600ms and a conjunction analysis revealed an overlap at central electrodes from around 300ms onward, reflecting a modality general P3 response. Source reconstructions were employed to identify the underlying neuronal generators of the MMRs: The modality specific MMN responses were source localized to the respective sensory cortices of each modality and additional shared right lateralized (pre-)frontal sources. The sources contributing to the modality general P3 MMR were found primarily in a shared network with frontal dominance. With exception of the vMMN, all MMRs showed parametric modulation by stimulus repetition, indicating highly comparable dynamics of mismatch processing across modalities. Specifically, longer stimulus repetitions led to increased MMN amplitudes. This effect was driven by a reciprocity in the response to standards (which became more positive) and the response to deviants (which became more negative). In line with its polarity, responses around the P3 were found to show the opposite pattern. The evaluation of the cross-modal

stimulus predictability revealed that standard and deviant responses within the modality-general P3 cluster were sensitive to predictive information carried by the tri-modal stimulus configuration. Across the senses, the modality-general P3 was increased to mispredicted stimulus transitions, indicating sensitivity to the global multi-modal sequence context.

In addition to the ERP analysis, the study aimed to compare different computational strategies of sequence processing potentially underlying neuronal generation of MMRs. To this end, we generated regressors from different Bayesian observer models and a model reflecting the combined dynamics of repetition suppression and change detection, all making different predictions for the dynamics of single-trial EEG data. The latter TL model reflected the reciprocal dynamics of standard and deviant responses to stimulus repetitions observed in the ERP analyses and previously reported in the (auditory) MMN literature (Baldeweg, 2006; Haenschel et al., 2005; Lieder et al., 2013), without invoking assumptions of probabilistic inference. Comparisons of computational models indicated that Bayesian models, tracking transitions between observations (i.e., estimating TP), captured the dynamics of single-trial responses to the roving stimulus sequences better than the static model reflecting the combined dynamics of repetition suppression and change detection (TL model). Moreover, the cross-modal Bayesian model (CM) which additionally captured cross-modal conditional dependence of stimulus alternation outperformed the purely uni-modal Bayesian model (UM). The CM model outperformed the UM model at some central electrodes during both the MMN (120ms) and the P3 (330) time windows. Notably, the superiority of the CM model at central electrodes suggest a correspondence of the modeling results with the cross-modal P3 modulation found in the average-based results and additionally suggest some multi-modal contribution during earlier times, likely from similar regions underlying the P3. Finally, the comparison of different read-out functions for the Bayesian models provided tentative evidence that the early MMN may signal model inadequacy scaled by confidence as reflected by CS, whereas the later P3 MMR seems to rather reflect model updating dynamics in terms of BS.

Overall, the results showed that highly comparable dynamics underlie MMRs in response to stimulus transitions in all three modalities and suggest an interaction of sensory specific cortices with modality-general (pre-frontal) areas as generators. Moreover, these dynamics were best explained by Bayesian observer models tracking uni-modal stimulus transitions as well as cross-modal conditional dependencies. Therefore, the results provided evidence for a role of the MMN and P3 as markers of probabilistic sequence processing at increasing scales of complexity.

2.4 Study 3

Grundei, M., Schmidt T. T., and Blankenburg, F. (2023). **A multimodal cortical network of sensory expectation violation revealed by fMRI.** *Human Brain Mapping, 44*(17), 5871-5891.

Previous research has rarely investigated MMRs to multi-modal sequences and very few studies allow direct comparisons of MMRs between modalities, particularly using fMRI. Nevertheless, seminal work has provided indications for a multi-modal mismatch network of modality-specific sensory cortices and modality-general fronto-parietal regions (Downar et al., 2000, 2001, 2002), in line with converging evidence from the auditory modality (Chennu et al., 2016; El Karoui et al., 2015; Uhrig et al., 2014; Phillips et al., 2015, 2016; Durschmid et al., 2016). Moreover, it has been suggested that the abstract structure of auditory and visual sequence patterns might be encoded in inferior frontal and intraparietal cortex (Wang et al., 2015, 2019; Dehaene et al., 2015; Planton and Dehaene, 2021) and current research on statistical learning points to domain-general computations underlying associative and probabilistic learning across different senses (Frost et al., 2015; Saffran and Thiessen, 2007). Therefore, as previous research suggests a role for modality-general activations of fronto-parietal regions during mismatch processing and sequence perception, we aimed to characterize the putative mismatch network in response to multi-modal sequence regularities with fMRI in more detail. To this end, we presented participants (N=29) with our tri-modal version of the roving stimulus paradigm in fMRI to elicit MMRs in the auditory, visual and somatosensory modality.

With GLM analyses we showed that modality specific activation to the stimulus sequences were found in hierarchically higher sensory cortices corresponding to STG for the auditory, IT and LOC for the visual and SII and IC for the somatosensory modality. Modality general activation was found in a fronto-parietal network involving the IFJ, TPJ, IC and SMA/ACC. Using PPI-connectivity analyses, we demonstrated connectivity increases during mismatch processing from seed regions in each higher-order sensory cortex, based on the MMR effects, to the common multi-modal network hubs, including clusters in the extended TPJ area encompassing IPS. Moreover, we showed that deviant responses within the network were modulated by local stimulus repetition, increasing with the amount of prior standard presentations. These effects were most pronounced in higher-order sensory cortices of each modality and indicate highly comparable dynamics of expectation violation across the senses. Additionally, our novel experimental manipulation of cross-modal stimulus predictability revealed that hierarchically higher (modality

general) regions of the mismatch network were sensitive to cross-modal expectation violation: We found increased activity for mispredicted stimulus transitions with lower cross-modal conditional probability, compared to predicted transitions, which was most pronounced in the IPS, indicating that the intraparietal cortex might keep track of global multi-modal sequence regularities. Taken together, our findings shed light on mismatch responses to multi-modal probabilistic sequences in a shared cortical network of expectation violation.

CHAPTER 3

Discussion

The aim of the thesis was to investigate neuronal mismatch responses and utilize their signatures to probe processes of probabilistic inference in the brain. To this end, three empirical evaluations of brain responses to roving stimulus sequences were performed. The sequences represented sensory model environments, designed to inspect different aspects of mismatch processing. A somatosensory roving stimulus paradigm was used to study somatosensory MMRs in response to a volatile environment, allowing for a thorough investigation of the underlying computational principles by means of single-trial EEG analyses. Furthermore, a novel tri-modal version of the roving stimulus paradigm was designed to inspect MMRs to probabilistic multi-modal input sequences in EEG and fMRI. Single-trial modeling of EEG data was complemented with conventional fMRI analyses to specify the mismatch dynamics of the auditory, visual and somatosensory modality.

In order to provide the basis for an evaluation of MMRs as signatures of the brain's underlying generative model of the sensory environment, the discussion will commence with an overview of the MMRs identified across the three empirical studies. The corresponding modality specific and modality general aspects of mismatch processing will be highlighted, where modality general refers to comparable response dynamics or a shared cortical network. Further, the different aspects of the corresponding computational dynamics which have been dissected by means of model comparison in studies 1 and 2 will be jointly discussed. Finally, the results of the empirical evaluations will be merged to conclude with a hypothesized processing hierarchy of probabilistic inference in the brain as reflected by MMRs.

3.1 Domain specificity of MMRs

The auditory MMN

The auditory MMN, as the classic signature of mismatch processing was identified in our multimodal EEG study as a negativity at fronto-central electrodes and corresponding positivity at temporo-parietal sites, ranging around 80-200ms in line with the MMN literature (Naatanen et al.,

1978, 2007). Within the MMN cluster, deviants deflected from the standard ERP around the peak of the auditory N1 component, reaching their maximum difference around the peak of the subsequent P2 component. Our source reconstruction analyses suggested activity in bilateral STG and right lateralized pre-frontal cortex to underlie the MMN in correspondence with the large body of prior MMN research (Doeller et al., 2003; Downar et al., 2000; Molholm et al., 2005; Opitz et al., 2002; Rinne et al., 2005; Yucel et al., 2005; Deouell, 2007) as well as with our fMRI results which showed auditory mismatch effects primiarly in bilateral STG and IFJ with right lateralization. Interestingly, the STG activation identified in our fMRI study corresponded to higher order auditory processing areas with minimal overlap with primary auditory cortex, which is often specified in EEG source modeling as a separate source in Heschl's gyrus (e.g., Garrido et al., 2009b). In contrast, we found sensory specific activations in secondary auditory areas across the STG (TE3) and in the upper (STS1/TE4) and lower bank of the STS (STS2/TE5) which are considered high level auditory processing regions (Zachlod et al., 2020). This result is in accordance with a recent comparative fMRI study showing STG (TE3) activation for MMRs to intensity changes among other stimulus features (Zvyagintsev et al., 2020). Therefore, while it is obvious that primary auditory cortex contributes to the MMR as it receives sensory input signals, our fMRI results suggest that the primary generator signaling sensory specific auditory mismatch lies in non-primary auditory regions in STG. This indication is in line with DCM studies modeling the STG as the intermediate stage of mismatch processing, receiving feedforward input from primary auditory and feedback input from pre-frontal cortex (Auksztulewicz and Friston, 2015; Chennu et al., 2016; Garrido et al., 2008, 2007a, 2009b; Phillips et al., 2015, 2016). As such, our results are in support of the interpretations put forward by such modeling studies of MMRs as reflections of prediction error signals in response to violation of top-down predictions.

The visual MMN

The visual MMN (vMMN) identified in our multi-modal EEG study developed from occipital to parieto-temporal electrodes between around 130-200ms in line with previous vMMN studies (Pazo-Alvarez et al., 2003). We localized the vMMN to the visual cortex as well as right lateralized pre-frontal areas in source reconstruction analyses as suggested by prior research (Kimura et al., 2011; Czigler, 2007; Pazo-Alvarez et al., 2003; Hedge et al., 2015; Downar et al., 2000). With the fMRI study, we localized the visual MMR more specifically to LOC and IT in hierarchically higher brain areas of visual processing such as V4, V5 and the fusiform gyrus as well as the IFJ region, thus confirming some indications of previous research for higher order visual areas

underlying visual MMRs (Urakawa et al., 2010; Stefanics et al., 2019; Yucel et al., 2007; Egner et al., 2010). Stefanics et al. (2019) identified highly similar higher order visual regions even though they used vastly different roving stimulus features (faces as opposed to flashes). The authors suggested corresponding vMMN responses to reflect precision-weighted prediction errors in line with hierarchical predictive coding in the visual system (Stefanics et al., 2018, 2019, 2014). Such interpretations would be supported by visual research indicating that top-down projections from higher cortical processing stages modulate responses particularly in non-primary visual areas (Buffalo et al., 2010; Johnson et al., 2007; Kastner et al., 1998; Mohr et al., 2009) and highlighting the modulatory effects of sensory expectations in hierarchical visual processing (de Lange et al., 2018; Ferrari et al., 2022; Summerfield and de Lange, 2014; Summerfield and Egner, 2009).

It should be noted that we observed a discrepancy between the EEG and fMRI result in terms of the strength of the visual MMRs. The vMMN as recorded with EEG only showed relatively small difference between standard and deviant responses. This observation might be reflective of the variable reports in the literature indicating large variance in the reported topographies and time windows for the vMMN in contrast to the auditory MMN, even failing to identify the response entirely (Male et al., 2020). The sensory specific effects in the fMRI study, however, were found to be comparable across the senses. While this divergence might be driven by the inability to differentiate the contributions of MMN and P3 in fMRI, it might also be reflective of a different placement of the visual stimuli in the fMRI study. The visual stimuli consisted of bilateral flashes with different intensities. To my knowledge, no other study has used similar visual flash stimuli to elicit vMMN and the results are therefore not directly comparable to previous research. However, in the EEG study, the stimuli were placed further in the periphery than in the fMRI study where they were presented much more central in the visual field. Due to the retinotopic organisation of the visual cortex (Sereno et al., 1995; Horton and Hoyt, 1991), a "far peripheral" placement (i.e., $>$ 60° in the EEG study; Strasburger et al., 2011) of the stimuli results in the activation of (primary) visual areas folded deep inside the cortex, in the calcarine sulcus between the hemispheres. It is therefore possible that the vMMN was not weaker per se but was merely harder to detect by means of EEG.

The somatosensory MMN

The somatosensory MMN (sMMN) was investigated in two EEG studies and identified as a (fronto-) temporal negativity, resulting primarily from increased N140 components, as previously suggested by Kekoni et al. (1997). Most previous sMMN studies used oddball paradigms where

some critical discussion revolves around the distinction of the sMMN from an N140 modulation by stimulus properties alone. With our two studies we provide further evidence for an sMMN around the N140 component which can be assumed to be independent of stimulus confounds due to the reversed roles of standard and deviant stimuli in the roving paradigm. Given a comparatively small amount of research as well as conflicting results regarding the exact components that may constitute a somatosensory equivalent to the auditory MMN, our results constitute an important contribution by providing evidence for a temporally centred negativity with corresponding central positivity as the primary sMMN around 140ms in line with previous suggestions (Ostwald et al., 2012). However, in both EEG studies, we identified an additional early somatosensory MMR at around 60ms which was followed by the sMMN around 90-190ms with a corresponding positivity reversal at central electrodes. We identified the somatosensory system as the primary driver of these MMRs in both studies and a contribution of right lateralized pre-frontal cortex was highlighted in our multi-modal study. In the uni-modal somatosensory study, dipole modeling of the MMRs suggested that the early MMR ($\sim$60ms) originated in SI while the sMMN appeared to be a combined response of SI and SII, supporting similar indications of previous research (Akatsuka et al., 2007a,b; Andersen and Lundqvist, 2019; Butler et al., 2012; Naeije et al., 2016, 2018; Ostwald et al., 2012; Spackman et al., 2010).

Our fMRI study confirmed the EEG source modeling results, showing mismatch related activation in SI and SII specific to the somatosensory input sequences as well as activation in IC and the IFJ region in line with prior fMRI research on somatosensory MMRs (Allen et al., 2016; Chen et al., 2008; Downar et al., 2000). Most pronounced activation was found in SII and activation in SI was found in BA1 and BA2 which are considered non-primary input regions (Purves et al., 2008). A similar network of SII, IC and fronto-parietal activations has been previously suggested in an fMRI study by Chen et al. (2008) to underlie somatosensory MMRs to attended and unattended uni-modal deviants during similar stimulation, showing that SII activation was independent of the attentional focus, in contrast to the higher-level processing stages. This supports the idea that SII is a primary driver for an early stage of mismatch processing reflected in the sMMN response. The IC, on the other hand, has been suggested to be involved in the integration of ascending sensory information with descending signals from higher level regions in prefrontal cortex during somatosensory processing (Cerliani et al., 2012; Lovero et al., 2009; Seth et al., 2011) and might play a role in the interaction with higher level processing stages. Correspondingly, in a DCM study using somatosensory roving stimulus sequences Allen et al. (2016) identified modulations in a network comprised of SI, IC and MFG during somatosensory mismatch processing and

provided evidence for a role of IC in the coordination of hierarchical predictive interactions. In correspondence with the results of our connectivity analysis in the fMRI study, the authors showed increasing feedforward connectivity from somatosensory cortex to insular and prefrontal cortices for unexpected stimuli, while feedback projections were found between IC and somatosensory cortex, suggesting that the IC mediates the reciprocal exchange with hierarchically higher areas (Allen et al., 2016). Similarly, Fardo et al. (2017) demonstrated in a DCM study that expectation violation reflected in somatosensory MMRs was accompanied by intrinsic modulations within the somatosensory system in SI and SII as well as extrinsic recurrent connectivity modulations between somatosensory, pre-frontal and parietal regions.

3.2 Domain generality of MMRs

The P3 - a reflection of the modality general mismatch network

P3 responses were identified in form of a large late MMR at central electrodes in response to sequences of auditory, visual and somatosensory stimuli in our EEG studies. The exact latency of the P3 differed between modalities with an early latency for the auditory modality from around 250ms, later latency for the visual modality from around 300ms and a continuous evolution was observed in the somatosensory modality in both studies from the central positive sMMN reversal-component to a P3 around 300ms. Despite these latency specifics, a common P3 mismatch cluster was identified in the multi-modal study from 300-350ms, followed by a more posterior cluster extending from 400-600ms, in line with a large body of research showing P3 responses in each sensory modality (Duncan-Johnson and Donchin, 1982; Conill, 1998; Yamaguchi and Knight, 1991, 1992) and suggestive of an earlier P3a and a later P3b component (Polich, 2007).

Across studies we identified fronto-parietal activation to underlie MMRs in all three modalities. Particularly our fMRI results provide compelling evidence for a shared mismatch network involving IFJ, TPJ, IPS and pre-SMA/ACC. The fronto-sensory activation likely reflects primarily the first stage of mismatch processing. This is supported by our EEG results showing pre-frontal and sensory cortices to underlie MMN responses across the senses as well as by previous research in the auditory (Deouell, 2007; Garrido et al., 2009b), and to a lesser extent the visual (Hedge et al., 2015; Tse et al., 2021) and somatosensory modality (Allen et al., 2016; Fardo et al., 2017; Huang et al., 2005). The parietal cortex, on the other hand, is likely involved at later stages of mismatch processing and, in interaction with pre-frontal cortex, corresponds to modality general processing reflected in the P3 (Linden, 2005; Bekinschtein et al., 2009; Phillips et al., 2016; Uhrig et al., 2014). As recently highlighted by Doricchi et al. (2022), the larger TPJ area with its dorsal

extension into IPS (Igelstrom and Graziano, 2017), is thought to code matches (left lateralized) and mismatches (right lateralized) between expected and actual events across sensory, motor and cognitive operations and keeps track of their statistical contingencies (Doricchi et al., 2022; Parr et al., 2023). Our fMRI results provided further evidence for such wider TPJ activation as a common signature of mismatch processing across the senses. Moreover, previous MMR studies indicated connectivity modulations within and between the identified regions of the mismatch network and highlighted their accordance with hierarchical predictive processing (Allen et al., 2016; Fardo et al., 2017; Auksztulewicz and Friston, 2015; Chennu et al., 2016; El Karoui et al., 2015; Phillips et al., 2015, 2016; Uhrig et al., 2014). We complemented these findings by additionally showing connectivity modulations from sensory regions to the fronto-parietal mismatch network during expectation violation, highlighting its domain generality.

Local expectation violation - a shared feature across modalities

An important feature of the MMN which theories of its generation have aimed to account for is its sensitivity to stimulus repetition. The MMN is long-known to increase with prior repetition of standards (Sams et al., 1983; Näätänen, 1992; Imada et al., 1993; Javitt et al., 1998). Correspondingly, across all investigations we found modulations of MMRs by stimulus repetition. We showed that the auditory and the somatosensory MMN increase with the length of the preceding stimulus train and found indications for a comparable tendency in the visual MMN (although non-significant in the EEG study). Moreover, we showed that these increases were driven by a reciprocal negative modulation of deviant and positive modulation of standard responses, suggesting a combined influence of a repetition-scaling of change detection (i.e., error signaling) as well as repetition suppression of expected input. Strikingly, the same repetition effect was identified for the sMMN in both EEG studies while only few previous studies have described similar modulations outside the auditory modality. For the sMMN, Kekoni et al. (1997) reported an attenuated standard response restricted to the first repetition of a stimulus after which it plateaued and Naeije et al. (2018) showed an increased response for deviants which followed four standards compared to one. We extended these indications by showing very robust modulation of somatosensory standards and deviants across a longer range of at least the first 8 stimulus repetitions.

For the auditory MMN, these repetition dynamics have been previously described under the framework of predictive coding: Repetition positivity of the standard trace during MMN generation has been interpreted as a reflection of repetition suppression, resulting from fulfilled prediction (Auksztulewicz and Friston, 2016; Baldeweg, 2006; Costa-Faidella et al., 2011a,b). A cor-

responding negative modulation of deviant responses thus signals a failure to suppress prediction error after violation of the local regularity established by the current stimulus train. Under such a view, longer trains of repetitions lead to higher precision in the probability estimate which in turn results in a scaling of the prediction error in response to prediction violation (Friston, 2005, 2010; Auksztulewicz and Friston, 2016). In our fMRI study we showed that the deviant increase with train length found for all three modalities was most pronounced in the modality-specific regions of the mismatch network, in higher order sensory cortices. This finding of expectancy modulated mismatch responses in higher order sensory cortices aligns well with research in rats showing that prediction error responses to regularity violations increase along the auditory processing pathway (Parras et al., 2017) with error responses predominantly found in hierarchically higher auditory regions (Parras et al., 2021; Luo et al., 2023) and adaptation dominating in primary auditory cortex (Parras et al., 2021). Additionally, although much less pronounced, we found some repetition modulation in the fronto-parietal mismatch network which might indicate further projections of remaining prediction errors, in line with the connectivity increases to these areas. Given that we identified the same pattern of repetition modulation also for the P3 responses of each modality in the EEG, these parametric activations might also correspond to repetition effects specific to the P3. Overall, the observed repetitions effects indicate highly comparable dynamics of MMRs based on stimulus expectancy across modalities and might reflect a sensitivity of MMRs to local stimulus transitions (Dehaene et al., 2015).

Global expectation violation - sensitivity of the P3 to cross-modal dependencies

In both of our studies investigating MMRs in multi-modal sensory environments by means of EEG and fMRI, we identified novel context sensitivity to global cross-modal predictive information in the stimulus sequences. We used tri-modal probabilistic sequences to create cross-modal regularities by defining the transitions in one modality conditional on the configuration of the other two modalities. As such, stimulus transitions were more or less likely based on the multi-modal context and thus were rather predicted or mispredicted.

In the multi-modal EEG study, we found increased P3 responses to such mispredicted stimulus transitions, regardless of their role as standards or deviants. This indicates that the P3 partly reflects the violation of global, cross-modal predictions extracted from multi-modal context information, in addition to local regularity violations driven by repetitions. In recent MMR research, the local-global paradigm has revealed that the early MMN and the later P3 reflect two hierarchical stages of mismatch processing which signal regularity violation on increasing levels of sequence complexity

(Bekinschtein et al., 2009; Chennu et al., 2013, 2016; Dehaene et al., 2015; Durschmid et al., 2016; El Karoui et al., 2015; King et al., 2014; Wacongne et al., 2011). While the MMN is primarily sensitive to local regularities, such as basic stimulus repetition, the P3 is additionally sensitive to global regularities such as a repeating pattern over an extended period of time. Importantly, this dichotomy for MMRs was recently shown to hold for the auditory, visual and somatosensory modality (Niedernhuber et al., 2022). Our finding of a sensitivity of the P3 response of each modality to cross-modal regularity violations additionally suggests that the P3 MMR might signal violation of sequence contingencies on multiple *spatio-temporal* global scales as our manipulation concerned the stimulus configuration in a tri-modal sensory space.

Evidence converges to the view that the MMN, induced by local deviants primarily activates sensory regions, while the P3 MMR after global deviance is accompanied by frontal (Chao et al., 2018; Chennu et al., 2013; El Karoui et al., 2015) and fronto-parietal activations (Bekinschtein et al., 2009; Uhrig et al., 2014), in line with the neuronal sources generally thought to underlie the P3 (Linden, 2005; Polich, 2007). In our fMRI study, across sensory modalities, we identified the dorsal part of the larger TPJ area around IPS (Doricchi et al., 2022; Igelstrom and Graziano, 2017) to be activated by cross-modal expectation violation. This suggests a correspondence between the fMRI and EEG results, since the parietal cortex has been shown to accompany P3 activation during global regularity violations in humans (Bekinschtein et al., 2009) and in macaque monkey IPS (Uhrig et al., 2014). Moreover, the finding is directly in line with the well-characterized role of the posterior parietal cortex as a major connection hub for different senses (Damasio, 1989; Hagmann et al., 2008; Tomasi and Volkow, 2011), mapping multi-modal inputs in both human and non-human primates (Avillac et al., 2007; Sereno and Huang, 2014). This *convergence-zone* for multi-modal information integration (Damasio, 1989) has been proposed to provide a critical gateway to transform sensory information into cognitively relevant functions (Mesulam, 1998). Thus, the extended TPJ area (including the IPS) forms the major parietal network hub for multi-modal integration and higher-order cognition (Igelstrom and Graziano, 2017) and is particularly known for coding unexpected events across a variety of sensory and cognitive processes (Doricchi et al., 2022).

It has been argued that Bayesian inference (in terms of a predictive coding implementation) can account for and unify the large amount of different functions and sub-divisions of the parietal convergence-zone (Seghier, 2013): The area provides an amodal interface between bottom-up sensory inputs and hierarchically higher cortical processing levels and is ultimately engaged in optimizing probabilistic representations of what caused sensory inputs. This suggests that the

area forms an integrative processing hub in the mismatch network with IPS specifically signaling expectation violation based on cross-modal contingencies. Such an interpretation is in line with previous suggestions of IPS involvement in a modality-general representation of probabilistic sequences (Planton and Dehaene, 2021) which is based on findings of IPS activity during regularity violations across different modalities and presentation formats (Planton and Dehaene, 2021; Wang et al., 2015, 2019).

3.3 MMRs as signatures of Bayesian inference

To investigate the computational principles underlying MMRs, we employed single-trial modeling of the signal dynamics during mismatch processing in two EEG studies. While both studies provided different comparisons of specific implementations of Bayesian inference models, in the multi-modal study we additionally performed an overarching comparison of a descriptive train-length (TL) model with a family of Bayesian models. The TL model captured the dynamics of repetition suppression and error response scaled by the length of the preceding stimulus train, derived from our empirical observations of these dynamics described above as well as the literature (e.g., Haenschel et al., 2005; Baldeweg, 2006; Lieder et al., 2013). We aimed to clarify if the observed dynamics reflect signatures of Bayesian computation beyond the phenomenological level. Using family-wise model comparison, we showed that models performing Bayesian inference based on stimulus transitions indeed outperformed the descriptive TL model supporting the view that MMRs are markers of probabilistic sequence processing in the brain. In the following, the specific insights we gained from the different implementations of Bayesian inference models across the two EEG studies will be discussed.

MMRs reflect belief mismatch and model updating

A comparison which we investigated in both EEG studies concerns different surprise read-out functions for the Bayesian models. The results provided insights into the dynamics specific to MMN and P3 responses and allowed to draw some conclusions on their potential computational roles. Interestingly, across both studies, the results showed the tendency to converge to a key distinction: While early responses reflected signatures of probabilistic novelty detection scaled by belief commitment as implemented by confidence-corrected surprise (CS), later responses reflected signatures of model-adjustment or belief updating as implemented by Bayesian surprise (BS).

We found evidence for CS encoding from around 70ms in both EEG studies. Focusing on the

the somatosensory system, in study 1 we localized CS to be encoded in bilateral SII until around 170ms which is broadly in line with our fMRI findings identifying most pronounced sensory specific mismatch processing in SII. Similarly, CS was generally dominant for MMN responses in the multi-modal study until around 250ms. This reflection of CS in early mismatch processing is in line with its computational role to signal an initial mismatch between the current belief of the generative model and a new observation, therefore corresponding to a probabilistic novelty detection response (termed puzzlement surprise by Faraji et al. (2018)). We showed that CS outperformed predictive surprise (PS) as another measure of probabilistic novelty detection which corresponds to an unsigned prediction error (Lieder et al., 2013). In contrast to basic prediction error formulations, CS is scaled by the negative entropy of the belief distribution serving as a quantification of model commitment (Faraji et al., 2018; Modirshanechi et al., 2022): CS in response to an unexpected event is high if the precision of the belief distribution is high, thus reflecting the idea that prediction errors are scaled by precision weighting to encode belief uncertainty (Friston, 2005; Mathys et al., 2014; Iglesias et al., 2013; Stefanics et al., 2018). Previous studies have similarly included confidence as a weighting factor on beliefs and have shown that it plays an important role in probabilistic inference (Buchel et al., 2014; Meyniel et al., 2015; Meyniel and Dehaene, 2017; Meyniel, 2020; Mulders et al., 2023). Therefore, our results are in line with previous research and support interpretations of early MMRs as reflecting initial novelty detection as suggested by Escera and Corral (2007) which, importantly, we show to be scaled by belief uncertainty. This encoding of uncertainty supports the view that the signal indicating a mismatch between the current observation and the model belief is a reflection of probabilistic inference processes (Friston, 2005).

It should be noted that the multi-modal study was not designed to differentiate response signatures between the senses as the regressors of the different modalities were fit in a single model. Nevertheless, we found clear evidence for surprise signatures across all relevant electrodes (involved in each MMN of the three senses). Moreover, the beta maps of model regressors of each modality were directly reflective of their respective sensory specific MMN topography suggesting that the model regressors indeed captured modality specific signal behaviours. Given the similarity of MMR dynamics discussed above, a comparison of different surprise signatures under the assumption of the same generative model underlying the signals of all modalities can be assumed to be warranted. Thus, although the primary aim of the model comparison of the multi-modal study was to distill differences between uni- and multi-modal inference, it nevertheless provided valuable indications for the computations underlying MMRs in general which are particularly

suggestive given the direct correspondence between both independently performed EEG studies.

A second indication given by both EEG studies is that the initial signaling of belief inadequacy is followed by signatures of model updating as reflected by BS (Lieder et al., 2013; Modirshanechi et al., 2022). Both studies show a dominance of BS around 300ms, corresponding to P3 responses. This finding is in line with previous studies suggesting that the P3 reflects BS dynamics (Ostwald et al., 2012; Kolossa et al., 2015; Mars et al., 2008; Seer et al., 2016) and correspondingly indicates that the P3 is reflective of model-adjustment at later stages of mismatch processing. Interestingly, O'Reilly et al. (2013) showed in an fMRI study that belief updating represented by BS was encoded in ACC activation during processing of unexpected visual events and was additionally associated with modulations of IPS, although IPS was also generally reflective of error signaling in terms of PS. Given that both of these areas are part of the mismatch network identified in our fMRI study, this suggests a correspondence of our finding of model updating signatures in the P3 with these modality general hubs of mismatch processing. Moreover, in terms of our mutli-modal study, these findings suggest a role for IPS in signaling mismatching beliefs about multi-modal inputs to trigger updating processes coordinated by the ACC (O'Reilly et al., 2013; Debener et al., 2005).

In the somatosensory study, we additionally found indications for BS at centro-posterior electrodes at earlier times, between 140-190ms localized to SI in the somatosensory system. This finding is in line with previous somatosensory research in humans (Ostwald et al., 2012) and rodents (English et al., 2023), although these studies did not compare BS with other surprise signatures. Given that we restricted our dipole model to the somatosensory system, the updating signature in SI might also reflect feedback processes from other cortical regions (such as ACC), which we did not explicitly model. However, these interpretations are rather speculative, especially since the early involvement of BS was not replicated in the multi-modal work. Nevertheless, the fact that CS is reflected in signals already from around 70ms onward in both of our modeling studies indicates a cascade of an initial mismatch informing subsequent belief updating which potentially already starts prior to 200ms.

On a broader scale, our results converge to the view that earlier MMRs may be considered markers of prediction error scaled by confidence which inform subsequent updates of the predictive model reflected in the P3. As phrased by Meyniel (2020): confidence-weighting sets the balance between predictions and new data to inform updating of the current belief and represents a useful strategy whenever multiple sources of information must be integrated.

MMRs reflect transition probability inference

A generative model hypothesized to underlie mismatch processing might implement different approaches to incorporate new observations such as the estimation of stimulus probability (SP) which reflects the sensitivity of MMRs to the frequency of items (Naatanen et al., 2007; Mars et al., 2008; Lieder et al., 2013; Kolossa et al., 2012), alternation probability (AP) which reflects sensitivity to alternation and repetition (Tervaniemi et al., 1994; de Gardelle et al., 2013; Summerfield et al., 2008; Todorovic and de Lange, 2012) as well as transition probability (TP) which explicitly tracks all possible transitions between observations (Dehaene et al., 2015; Meyniel et al., 2016; Maheu et al., 2019; Meyer and Olson, 2011; Meyer et al., 2014). In both modeling studies, we show that TP inference, which subsumes SP and AP computation, is well suited to account for the single-trial dynamics of EEG signatures during mismatch processing, in line with the suggestions by Meyniel et al. (2016) who proposed TP as a likely core model of probabilistic inference in the brain.

In the somatosensory study, we directly compared TP with SP and AP inference approaches, showing clear superiority of the TP model. Moreover, as we included second order Markov dependency in the somatosensory roving sequences we aimed to test whether the brain additionally tracks possible transitions from the previous *two* observations (TP_2). However, we were not able to conclusively differentiate between TP_1 and TP_2 models. Thus it is possible that the brain perfoms TP inference during perception of sequences on a higher level than just the first order Markov property. In any case, our modeling work provided further evidence for a sensitivity of MMRs to transitional structure of sequences which has previously been indicated for the auditory MMN (Mittag et al., 2016; Koelsch et al., 2016; Dehaene et al., 2015) as well as visual processing of probabilistic sequences more generally (Meyer and Olson, 2011; Meyer et al., 2014). These latter studies suggested that higher order visual processing stages such as area IT in monkeys might be reflective of TP inference which would be in line with our fMRI finding of these regions during visual mismatch processing. Other research has indicated a sensitivity to TP over SP inference in the modality general areas ACC and IFJ (Wang et al., 2017).

While evidence generally seems to converge to TP inference underlying MMRs, an open question concerns the different representations of local and global regularities reflected in MMN and P3 responses. Maheu et al. (2019) showed that both mid-latency (MMN window) and late MMRs (P3 window) were best explained by a model performing TP, whereas earliest responses ($<$100ms) by a model performing SP inference. Although we similarly found MMRs prior to 100ms, we did not confirm such a reflection of SP dynamics in earliest MMRs with our comparison of TP and SP models for the somatosensory system, but rather found evidence for TP throughout the full

post-stimulus time window. Maheu et al. (2019) additionally indicated that the different identified processing phases might resort to diverging time-scales of information integration from the input sequences. Earliest responses, which showed sensitivity to rarely occurring stimulus features (SP model) also showed very long time-scales of integration and thus likely reflect sensory adaptation dynamics. In contrast, later responses showed increasingly shorter time-scales of sequence integration which might rather indicate preferred pattern processing: Mid-latency and late mismatch signatures showed medium and short local integration windows, with observation half-lives of (i.e., 50% observation weighting at) around 10 and 5 observations, respectively. Interestingly, in our multi-modal study, we similarly found evidence for integration windows with half-lives of around 10-20 observations for earlier MMN responses and around 5-10 observations for later P3 responses. While these findings appear to contradict the distinction of local pattern sensitivity reflected in MMN and global pattern sensitivity in P3 responses, Maheu et al. (2019) argue that this difference might in fact relate to a relevance of basic transitional structure during earlier MMRs and rather chunk-based pattern inference during later MMR processes potentially operating on chunks of around 5 stimuli (e.g., Bekinschtein et al., 2009; Wacongne et al., 2011). Moreover, chunks of stimuli (i.e. groups of stimulus combinations) have been previously proposed to be an important quantity during sequence processing (Dehaene et al., 2015) and, by building on TP learning, are thought to underlie the processing of more complex sequence structure which ultimately gives rise to language learning (Saffran et al., 1996; Leonard et al., 2015).

Inference in a volatile environment

In the uni-modal EEG experiment we used somatosensory roving stimulus sequences to create a volatile sensory environment and investigate different implementations of a generative model performing inference on the probabilistic processes governing incoming observations. One comparison concerned the representation of environmental volatility. To adequately adapt one's beliefs in the face of changes to environmental statistics, hidden state dynamics may be explicitly represented by estimating different sets of statistics for a discretized latent state (HMM). Alternatively, without modeling the hidden state, a Bayesian observer might account for changing statistics in terms of leaky integration, favoring more recent observations over past ones (DC model). Both approaches, hierarchical models with hidden state representations (Iglesias et al., 2013; Lawson et al., 2017; Meyniel and Dehaene, 2017) and non-hierarchical (flat) models without representation of higher order change-points (Summerfield et al., 2011; Bell et al., 2016; Farashahi et al., 2017) have been shown to be appropriate learning strategies in model comparison studies.

Our comparison of these two approaches to draw inferences in a volatile environment provided strong evidence in favor of the DC model class over the HMM for all electrodes and post-stimulus time-points (which were found to be above the control null-model). As such, we showed that leaky integration provided sufficient flexibility for the model to adapt to the changing sequence statistics. The window of integration corresponded to a half-life of around 25 to 100 recent observations corresponding to the past 20 to 70 seconds in our experiment. Generally, while a very short time-scale promotes sensitivity to noise, a timescale that is too large prevents flexible adaptation following a change-point. In the context of the specific experiment, the observed timescales are local enough to adapt the estimated statistics to the occurence of change-points (occurring every 100 stimuli on average). Nevertheless, a wide range of integration windows have been reported in previous studies and might even be variable depending on the employed model environment (Ossmy et al., 2013). Overall, evidence appears to converge to local (such as a half-life of $<$20 observations) rather than global time-scales during mismatch processing (Ostwald et al., 2012; Kolossa et al., 2012; Meyniel et al., 2016; Maheu et al., 2019; Rubin et al., 2016).

It is possible that the brain would resort to more complex inference processes with probabilistic representations of hidden states if the state transitions would be more explicit and states would produce more diverging observation sequences. This might similarly be the case if the environment would contain incentives to learn the hidden states in terms of reward based inference (e.g., Behrens et al., 2007; Rouhani and Niv, 2021; Niv et al., 2015; Xu et al., 2021a). In our experiment, participants were neither aware of the hidden states in the data generation process, nor was their dissociation or any tracking of sequence statistics required to perform the behavioural task. Therefore, EEG signatures most likely reflect a form of unconscious, implicit learning of environmental statistics which has been previously suggested to underlie the MMN (van Zuijen et al., 2006; Koelsch et al., 2016) and sequence processing more general (Atas et al., 2014). It has been argued that explicitly modeling change-points in the environment is both difficult and potentially unnecessary for efficient Bayesian inference (Ryali et al., 2018). Instead, a form of restricted sensory evidence accumulation (or "leaky integration") might rather be performed by the brain as it is suggested to be biologically plausible (Ryali et al., 2018; Yu and Cohen, 2008; Farashahi et al., 2017; Glaze et al., 2015). Our results were in support of such a view by providing evidence that the somatosensory system resorts to an inference strategy of leaky integration in a volatile environment, as previously suggested for the somatosensory modality by Ostwald et al. (2012).

With respect to our multi-modal roving paradigm, although not explicitly designed as such,

the different sequences used over experimental blocks can also be viewed as a volatile sensory environment since the underlying statistics governing the observations changed over time during the experiment (although not within each experimental block). In light of these changes, although unnoticed by the participants, it can be argued that the generative model continuously required updating to adequately predict future observations. As described above, the P3 showed sensitivity to cross-modal prediction violation and tended to carry signatures of model updating, which suggests that the P3 likely reflects model updates with respect to the multi-modal context. As such, our results might be related to recent findings which suggest that later MMRs, such as the P3, reflect belief updates about the volatility of the underlying (hidden) statistics governing sensory observations (e.g., Weber et al., 2020, 2022). Additionally, other findings indicate that activation of areas which we identified as part of the modality general mismatch network (particularly IFG and IPS) might underlie the updating of expectations in changing sensory contexts (e.g., Ferrari et al., 2022; Niv et al., 2015; Leong et al., 2017). Therefore, our results might similarly suggest that later modality general MMRs are reflective of model updating processes in light of changing global (e.g., multi-modal) sequence statistics.

Inference in a multi-modal environment

With our tri-modal probabilistic sequences we created sensory model environments to investigate multi-modal Bayesian inference. We showed that a cross-modally informed model (CM), tracking conditional dependencies between modalities in addition to uni-modal transitions, outperformed a purely uni-modal transition probability model (UM) at central electrodes within an early ($\sim$120-250ms) and a late time-window ($\sim$300-350ms).

The fact that the CM model accounted for EEG dynamics in a late time-window around 300ms was directly reflective of the sensitivity of the P3 to cross-modal predictability and supports an interpretation of the P3 as a signature of Bayesian inference with respect to global (multi-modal) regularities. Given that cross-modal learning was not explicitly instructed or task-relevant, the results are compatible with the view that the brain is sensitive to multi-modal information by default (Ghazanfar and Schroeder, 2006; Driver and Noesselt, 2008) and that processing of such information might be appropriately captured by Bayesian inference (Noppeney, 2021; Shams and Beierholm, 2022; Kording et al., 2007). The earlier effects of the CM model prior to 300ms were not reflected in the average-based results which suggests that potential modulations of MMN signatures by cross-modal predictability manifested in single-trial dynamics but not in significant mean differences between predictability conditions. Nevertheless, given that this earlier CM effect

was primarily confined to central electrodes it may similarly be related to activity of the modality general mismatch network underlying the central topography of the P3, albeit at earlier latency. Given our fMRI results, the generators of these effects most likely include the parietal cortex (Linden, 2005; Polich, 2007) which has been additionally shown in response to global regularity violations, specifically (Bekinschtein et al., 2009; Uhrig et al., 2014).

Recent advances in research on Bayesian causal inference (BCI) suggest a central role for posterior parietal cortex, and IPS in particular, during multi-modal inference. BCI concerns the assumption that the generative model of the brain explicitly models the potential causal structure that might have generated sensory signals (Noppeney, 2021; Kording et al., 2007; Shams and Beierholm, 2022). As highlighted in a review by Noppeney (2021), a series of studies has convincingly demonstrated the spatio-temporal evolution of corresponding multi-modal inference processing in the brain. Initial sensory segregation ($<$100ms) of input signals, concerned with the computation of uni-modal estimates, was found in modality specific sensory cortices. Automatic ("forced") sensory fusion, forming multi-modal estimates, was found encoded in (posterior) IPS around 100-250ms and subsequent BCI computations were associated with (anterior) IPS around 350-450ms, inferring the causal structure of the input signal based on these estimates. These studies indicate that reliability-weighted sensory estimates are integrated in IPS prior to 300ms and used subsequently in interaction with frontal areas to infer the hidden causes of sensory inputs (Noppeney, 2021; Cao et al., 2019; Rohe and Noppeney, 2015). Our fMRI findings of activation of IPS (and in part IFJ) by cross-modal prediction violation might be reflective of similar inference processes. Taken together, our EEG and fMRI results of the multi-modal experiment might be suggestive of a comparable inference hierarchy for mismatch processing in multi-modal environments. We found that early processing ($\sim$70-120ms) was concerned with uni-modal transition probability inference, primarily driven by activation in (higher order) sensory regions. Multi-modal inference, tracking (cross-modal) conditional dependency, was found to become relevant during mid-latency and late mismatch processing (at central electrodes) and likely involved the IPS as a sensory convergence zone in the modality general mismatch network which we found to be sensitive to cross-modal conditional probability.

Overall, our results further indicate that the IPS keeps track of global cross-modal sequence regularities in terms of Bayesian inference, likely by estimation of transition probabilities across multi-modal observations or, potentially, by inference based on multi-modal stimulus chunks (analogous to the suggestions by Dehaene et al., 2015; Maheu et al., 2019). Moreover, as part of the network of executive control, the IPS has been suggested to coordinate the selection of sensory

expectations in multi-dimensional environments with co-existing regularities (Niv et al., 2015; Leong et al., 2017) as well as in volatile environments with changing regularities depending on the sensory context as mentioned above (Ferrari et al., 2022). Thus, converging evidence suggests a key role for the posterior parietal cortex, and IPS in particular, during probabilistic inference in multi-modal environments and our results across empirical investigations provide compelling evidence to substantiate this view.

3.4 General conclusion - A spatio-temporal hierarchy of mismatch processing

With our investigations of MMRs in audition, vision and somatosensation, we have provided further evidence that mismatch signals are ubiquitous during perceptual processing in the brain with highly similar response profiles across the senses. We identified an underlying network of modality specific (higher order) sensory cortices and modality general fronto-parietal regions. We have shown that shared features of MMRs such as repetition sensitivity of standard and deviant responses are well accounted for by Bayesian models performing probabilistic inference. Our comparison of different implementations of such models have provided insights into the computations involved in the inference processes most likely underlying MMRs. In the following conclusive remarks, I will suggest a common framework of a hierarchy of mismatch processing based on the main results and the current literature as discussed in the previous sections. Figure 3.1 provides an overview of the hypothesized evolution of cortical areas during mismatch processing post stimulus presentation, informed by the temporal specificity of our EEG results and the identified computational dynamics as well as the spatial specificity of fMRI results and dipole modeling.

Generally, our results indicated that MMN and P3 responses might be viewed as reflections of continuous processing during which the classic MMR signatures are peak markers of roughly differentiable phases. Early mismatch processing starts prior to 100ms and, around the time of the MMN, primarily invovles interactions between sensory cortices and right lateralized prefrontal regions around the IFJ. The early processing phase is concerned with modality specific sequence features and the estimation of their transitional structure. Unexpected stimulus transitions trigger a mismatch between the current sensory input and the model belief, scaled by confidence in the belief (which is based on the previously observed evidence). This initial signal of model inadequacy is further projected to modality general processing stages which involve a fronto-parietal network consisting of IFJ, TPJ, IPS as well as pre-SMA and ACC and underlie the generation of the P3. Processing around the time of the P3 is concerned with model adjustment based on prediction errors (signaling belief inadequacy) from earlier processing stages, most likely specifically

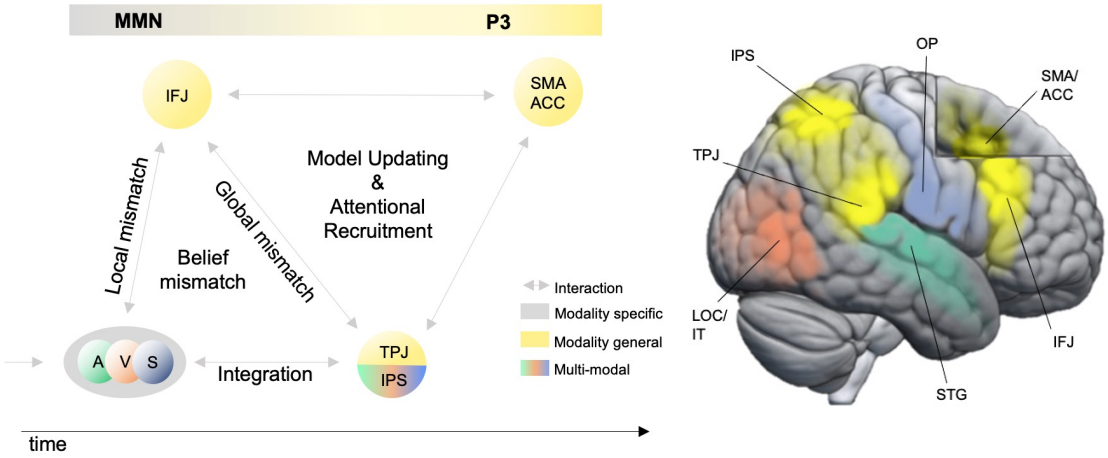


Fig. 3.1 The spatio-temporal hierarchy of mismatch processing. Left: The involvement of areas of the mismatch network from left to right over post-stimulus time. Early mismatch processing (inkl. the MMN) is primarily based on interactions (depicted by bilateral grey arrows) between sensory cortices and IFJ and reflects a mismatch between model beliefs and current sensory input, primarily concerning local sequence properties (e.g., local stimulus transitions). Later mismatch processing additionally involves parietal areas such as TPJ and IPS and ultimately gives rise to the P3 as a global mismatch response. The IPS in particular is primarily concerned with sensory integration (triple-colored) and inference based on cross-modal information. Later modality general mismatch processing involves interactions between SMA/ACC, TPJ/IPS and IFJ which underlie model updating computation. Finally, attentional resources are recruited and deployed for future evidence accumulation. Right: Modality specific areas primarily correspond to higher order sensory cortices: Auditory (green) along the superior temporal gyrus (STG); Visual (orange) across the lateral occipital cortex (LOC) and inferior temporal cortex (IT); Somatosensory (purple) in the opercular cortex (OP). Modality general areas (yellow) involve a fronto-parietal network consisting of inferior frontal junction (IFJ), the anterior portion of the supplementary motor area (SMA) and the anterior cingulate gyrus (ACC), the temporo-parietal junction (TPJ) and intraparitetal sulcus (IPS).

involving anterior cingulate regions during belief updating. Moreover, at the level of the P3, information from multiple sources is integrated and available for inference based on global sequence properties. In case of multi-modal sequence processing, the parietal cortex becomes particularly relevant as a major processing area for multi-modal statistical contingencies. While the TPJ (with pronounced right lateralization) is signaling mismatches for each modality, the IPS integrates uni-modal sequence information and specifically signals mismatching expectations with respect to cross-modal relations, e.g., by estimation of transition probability conditional on multi-modal environmental properties. Although processes in IPS, and potentially fronto-parietal interactions, likely already become relevant prior to 300ms to inform cross-modal inference, primarily later mismatch processing around the P3 reflects modality general inference based on global sequence properties. Finally, at later stages of mismatch processing, attentional resources are recruited based on the estimated relevance of different features of the sensory environment and deployed for future evidence accumulation. The hierarchical stages of mismatch processing are reflective of iterative

incorporation of new observations in light of mismatching beliefs and according updating of the generative model. In basic perceptual sequence processing without relevance for the system, as reflected in the current work, the brain appears to limit the integration of sensory evidence for inference in order to efficiently capture environmental volatility in terms of leaky integration. Overall, the computational processes identified by our empirical investigations of MMRs across the senses are highly suggestive of signatures of probabilistic inference in a complex, ever-changing and inherently multi-modal world.

Bibliography

Akatsuka, K., Wasaka, T., Nakata, H., Inui, K., Hoshiyama, M., and Kakigi, R. (2005). Mismatch responses related to temporal discrimination of somatosensory stimulation. *Clin Neurophysiol*, 116(8):1930–7.

Akatsuka, K., Wasaka, T., Nakata, H., Kida, T., Hoshiyama, M., Tamura, Y., and Kakigi, R. (2007a). Objective examination for two-point stimulation using a somatosensory oddball paradigm: an meg study. *Clin Neurophysiol*, 118(2):403–11.

Akatsuka, K., Wasaka, T., Nakata, H., Kida, T., and Kakigi, R. (2007b). The effect of stimulus probability on the somatosensory mismatch field. *Exp Brain Res*, 181(4):607–14.

Alain, C. and Woods, D. L. (1997). Attention modulates auditory pattern memory as indexed by event-related brain potentials. *Psychophysiology*, 34(5):534–46.

Alain, C., Woods, D. L., and Ogawa, K. H. (1994). Brain indices of automatic pattern processing. *Neuroreport*, 6(1):140–4.

Alho, K. (1995). Cerebral generators of mismatch negativity (mmn) and its magnetic counterpart (mmnm) elicited by sound changes. *Ear and Hearing*, 16(1):38–51.

Allen, M., Fardo, F., Dietz, M. J., Hillebrandt, H., Friston, K. J., Rees, G., and Roepstorff, A. (2016). Anterior insula coordinates hierarchical processing of tactile mismatch responses. *Neuroimage*, 127:34–43.

Andersen, L. M. and Lundqvist, D. (2019). Somatosensory responses to nothing: An meg study of expectations during omission of tactile stimulations. *Neuroimage*, 184:78–89.

Armstrong, B. C., Frost, R., and Christiansen, M. H. (2017). The long road of statistical learning research: past, present and future. *Philos Trans R Soc Lond B Biol Sci*, 372(1711).

Arnal, L. H. and Giraud, A. L. (2012). Cortical oscillations and sensory predictions. *Trends Cogn Sci*, 16(7):390–8.

Aslin, R. N. and Newport, E. L. (2012). Statistical learning: From acquiring specific items to forming general rules. *Curr Dir Psychol Sci*, 21(3):170–176.

Astikainen, P., Lillstrang, E., and Ruusuvirta, T. (2008). Visual mismatch negativity for changes in orientation–a sensory memory-dependent response. *Eur J Neurosci*, 28(11):2319–24.

Astikainen, P., Ruusuvirta, T., Wikgren, J., and Korhonen, T. (2004). The human brain processes visual changes that are not cued by attended auditory stimulation. *Neurosci Lett*, 368(2):231–4.

Astikainen, P., Ruusuvirta, T., Wikgren, J., and Penttonen, M. (2006). Memory-based detection of rare sound feature combinations in anesthetized rats. *Neuroreport*, 17(14):1561–4.

Atas, A., Faivre, N., Timmermans, B., Cleeremans, A., and Kouider, S. (2014). Nonconscious learning from crowded sequences. *Psychol Sci*, 25(1):113–9.

Atienza, M., Cantero, J. L., and Escera, C. (2001). Auditory information processing during human sleep as revealed by event-related brain potentials. *Clin Neurophysiol*, 112(11):2031–45.

Auksztulewicz, R., Barascud, N., Cooray, G., Nobre, A. C., Chait, M., and Friston, K. (2017). The cumulative effects of predictability on synaptic gain in the auditory processing stream. *J Neurosci*, 37(28):6751–6760.

Auksztulewicz, R. and Friston, K. (2015). Attentional enhancement of auditory mismatch responses: a dcm/meg study. *Cereb Cortex*, 25(11):4273–83.

Auksztulewicz, R. and Friston, K. (2016). Repetition suppression and its contextual determinants in predictive coding. *Cortex*, 80:125–40.

Auksztulewicz, R., Schwiedrzik, C. M., Thesen, T., Doyle, W., Devinsky, O., Nobre, A. C., Schroeder, C. E., Friston, K. J., and Melloni, L. (2018). Not all predictions are equal: "what" and "when" predictions modulate activity in auditory cortex through different mechanisms. *J Neurosci*, 38(40):8680–8693.

Avillac, M., Ben Hamed, S., and Duhamel, J. R. (2007). Multisensory integration in the ventral intraparietal area of the macaque monkey. *J Neurosci*, 27(8):1922–32.

Baldeweg, T. (2006). Repetition effects to sounds: evidence for predictive coding in the auditory system. *Trends Cogn Sci*, 10(3):93–4.

Baldeweg, T., Klugman, A., Gruzelier, J., and Hirsch, S. R. (2004). Mismatch negativity potentials and cognitive impairment in schizophrenia. *Schizophr Res*, 69(2-3):203–17.

Baldi, P. and Itti, L. (2010). Of bits and wows: A bayesian theory of surprise with applications to attention. *Neural Netw*, 23(5):649–66.

Bar, M. (2004). Visual objects in context. *Nat Rev Neurosci*, 5(8):617–29.

Bastos, A. M., Litvak, V., Moran, R., Bosman, C. A., Fries, P., and Friston, K. J. (2015). A dcm study of spectral asymmetries in feedforward and feedback connections between visual areas v1 and v4 in the monkey. *Neuroimage*, 108:460–75.

Baum, L. E., Petrie, T., Soules, G., and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The annals of mathematical statistics*, 41(1):164–171.

Beal, M. J. (2003). *Variational algorithms for approximate Bayesian inference*. University of London, University College London (United Kingdom).

Behrens, T. E., Woolrich, M. W., Walton, M. E., and Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nat Neurosci*, 10(9):1214–21.

Bekinschtein, T. A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., and Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. *Proc Natl Acad Sci U S A*, 106(5):1672–7.

Bell, A. H., Summerfield, C., Morin, E. L., Malecek, N. J., and Ungerleider, L. G. (2016). Encoding of stimulus probability in macaque inferior temporal cortex. *Curr Biol*, 26(17):2280–90.

Bendixen, A., Prinz, W., Horvath, J., Trujillo-Barreto, N. J., and Schroger, E. (2008). Rapid extraction of auditory feature contingencies. *Neuroimage*, 41(3):1111–9.

Bendixen, A. and Schroger, E. (2008). Memory trace formation for abstract auditory features and its consequences in different attentional contexts. *Biol Psychol*, 78(3):231–41.

Bendixen, A., Schroger, E., and Winkler, I. (2009). I heard that coming: event-related potential evidence for stimulus-driven prediction in the auditory system. *J Neurosci*, 29(26):8447–51.

Bertels, J., Franco, A., and Destrebecqz, A. (2012). How implicit is visual statistical learning? *J Exp Psychol Learn Mem Cogn*, 38(5):1425–31.

Berti, S. and Schroger, E. (2004). Distraction effects in vision: behavioral and event-related potential indices. *Neuroreport*, 15(4):665–9.

Berti, S. and Schroger, E. (2006). Visual distraction: a behavioral and event-related brain potential study in humans. *Neuroreport*, 17(2):151–5.

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. MIT press.

Brunia, C. H. M., Gaillard, A. W., and Kok, A. (1993). Psychophysiological brain research. *Psychophysiology*, 30.

Buchel, C., Geuter, S., Sprenger, C., and Eippert, F. (2014). Placebo analgesia: a predictive coding perspective. *Neuron*, 81(6):1223–1239.

Buffalo, E. A., Fries, P., Landman, R., Liang, H., and Desimone, R. (2010). A backward progression of attentional effects in the ventral stream. *Proc Natl Acad Sci U S A*, 107(1):361–5.

Butler, J. S., Foxe, J. J., Fiebelkorn, I. C., Mercier, M. R., and Molholm, S. (2012). Multisensory representation of frequency across audition and touch: high density electrical mapping reveals early sensory-perceptual coupling. *J Neurosci*, 32(44):15338–44.

Butler, J. S., Molholm, S., Fiebelkorn, I. C., Mercier, M. R., Schwartz, T. H., and Foxe, J. J. (2011). Common or redundant neural circuits for duration processing across audition and touch. *J Neurosci*, 31(9):3400–6.

Cao, Y., Summerfield, C., Park, H., Giordano, B. L., and Kayser, C. (2019). Causal inference in the multisensory brain. *Neuron*, 102(5):1076–1087 e8.

Carral, V., Huotilainen, M., Ruusuvirta, T., Fellman, V., Naatanen, R., and Escera, C. (2005). A kind of auditory 'primitive intelligence' already present at birth. *Eur J Neurosci*, 21(11):3201–4.

Cerliani, L., Thomas, R. M., Jbabdi, S., Siero, J. C., Nanetti, L., Crippa, A., Gazzola, V., D'Arceuil, H., and Keysers, C. (2012). Probabilistic tractography recovers a rostrocaudal trajectory of connectivity variability in the human insular cortex. *Hum Brain Mapp*, 33(9):2005–34.

Chait, M. (2020). How the brain discovers structure in sound sequences. *Acoustical Science and Technology*, 41(1):48–53.

Chang, A., Bosnyak, D. J., and Trainor, L. J. (2018). Beta oscillatory power modulation reflects the predictability of pitch change. *Cortex*, 106:248–260.

Chao, Z. C., Takaura, K., Wang, L., Fujii, N., and Dehaene, S. (2018). Large-scale cortical networks for hierarchical prediction and prediction error in the primate brain. *Neuron*, 100(5):1252–1266 e3.

Chen, T. L., Babiloni, C., Ferretti, A., Perrucci, M. G., Romani, G. L., Rossini, P. M., Tartaro, A., and Del Gratta, C. (2008). Human secondary somatosensory cortex is involved in the processing of somatosensory rare stimuli: an fmri study. *Neuroimage*, 40(4):1765–71.

Chennu, S., Noreika, V., Gueorguiev, D., Blenkmann, A., Kochen, S., Ibanez, A., Owen, A. M., and Bekinschtein, T. A. (2013). Expectation and attention in hierarchical auditory prediction. *J Neurosci*, 33(27):11194–205.

Chennu, S., Noreika, V., Gueorguiev, D., Shtyrov, Y., Bekinschtein, T. A., and Henson, R. (2016). Silent expectations: Dynamic causal modeling of cortical prediction and attention to sounds that weren't. *J Neurosci*, 36(32):8305–16.

Chomsky, N. (2002). *On nature and language*. Cambridge University Press.

Clark, A. (2013). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behav Brain Sci*, 36(3):181–204.

Conill, J. (1998). [p300 potentials evoked by visual stimulation]. *Rev Neurol*, 26(151):448–51.

Cooper, R. J., Atkinson, R. J., Clark, R. A., and Michie, P. T. (2013). Event-related potentials reveal modelling of auditory repetition in the brain. *Int J Psychophysiol*, 88(1):74–81.

Costa-Faidella, J., Baldeweg, T., Grimm, S., and Escera, C. (2011a). Interactions between "what" and "when" in the auditory system: temporal predictability enhances repetition suppression. *J Neurosci*, 31(50):18590–7.

Costa-Faidella, J., Grimm, S., Slabu, L., Diaz-Santaella, F., and Escera, C. (2011b). Multiple time scales of adaptation in the auditory system as revealed by human evoked potentials. *Psychophysiology*, 48(6):774–83.

Cowan, N., Winkler, I., Teder, W., and Naatanen, R. (1993). Memory prerequisites of mismatch negativity in the auditory event-related potential (erp). *J Exp Psychol Learn Mem Cogn*, 19(4):909–21.

Czigler, I. (2007). Visual mismatch negativity. *Journal of Psychophysiology*, 21(3-4):224–230.

Czigler, I., Balazs, L., and Pato, L. G. (2004). Visual change detection: event-related potentials are dependent on stimulus location in humans. *Neurosci Lett*, 364(3):149–53.

Czigler, I., Balazs, L., and Winkler, I. (2002). Memory-based detection of task-irrelevant visual changes. *Psychophysiology*, 39(6):869–73.

Czigler, I. and Pato, L. (2009). Unnoticed regularity violation elicits change-related brain activity. *Biol Psychol*, 80(3):339–47.

Czigler, I. and Sulykos, I. (2010). Visual mismatch negativity to irrelevant changes is sensitive to task-relevant changes. *Neuropsychologia*, 48(5):1277–82.

Czigler, I., Weisz, J., and Winkler, I. (2006a). Erps and deviance detection: visual mismatch negativity to repeated visual stimuli. *Neurosci Lett*, 401(1-2):178–82.

Czigler, I., Winkler, I., Pato, L., Varnagy, A., Weisz, J., and Balazs, L. (2006b). Visual temporal window of integration as revealed by the visual mismatch negativity event-related potential to stimulus omissions. *Brain Res*, 1104(1):129–40.

Damasio, A. R. (1989). Time-locked multiregional retroactivation: a systems-level proposal for the neural substrates of recall and recognition. *Cognition*, 33(1-2):25–62.

Dayan, P., Hinton, G. E., Neal, R. M., and Zemel, R. S. (1995). The helmholtz machine. *Neural Comput*, 7(5):889–904.

de Gardelle, V., Waszczuk, M., Egner, T., and Summerfield, C. (2013). Concurrent repetition enhancement and suppression responses in extrastriate visual cortex. *Cereb Cortex*, 23(9):2235–44.

de Lange, F. P., Heilbron, M., and Kok, P. (2018). How do expectations shape perception? *Trends Cogn Sci*, 22(9):764–779.

de Vignemont, F., Majid, A., Jola, C., and Haggard, P. (2009). Segmenting the body into parts: evidence from biases in tactile perception. *Q J Exp Psychol (Hove)*, 62(3):500–12.

Debener, S., Ullsperger, M., Siegel, M., Fiehler, K., von Cramon, D. Y., and Engel, A. K. (2005). Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. *J Neurosci*, 25(50):11730–7.

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., and Pallier, C. (2015). The neural representation of sequences: From transition probabilities to algebraic patterns and linguistic trees. *Neuron*, 88(1):2–19.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society: series B (methodological)*, 39(1):1–22.

Deouell, L. Y. (2007). The frontal generator of the mismatch negativity revisited. *Journal of Psychophysiology*, 21(3-4):188–203.

Dien, J., Spencer, K. M., and Donchin, E. (2004). Parsing the late positive complex: mental chronometry and the erp components that inhabit the neighborhood of the p300. *Psychophysiology*, 41(5):665–78.

Doeller, C. F., Opitz, B., Mecklinger, A., Krick, C., Reith, W., and Schroger, E. (2003). Prefrontal cortex involvement in preattentive auditory deviance detection: neuroimaging and electrophysiological evidence. *Neuroimage*, 20(2):1270–82.

Donchin, E. (1981). Presidential address, 1980. surprise!...surprise? *Psychophysiology*, 18(5):493–513.

Doricchi, F., Lasaponara, S., Pazzaglia, M., and Silvetti, M. (2022). Left and right temporal-parietal junctions (tpjs) as "match/mismatch" hedonic machines: A unifying account of tpj function. *Phys Life Rev*, 42:56–92.

Downar, J., Crawley, A. P., Mikulis, D. J., and Davis, K. D. (2000). A multimodal cortical network for the detection of changes in the sensory environment. *Nat Neurosci*, 3(3):277–83.

Downar, J., Crawley, A. P., Mikulis, D. J., and Davis, K. D. (2001). The effect of task relevance on the cortical response to changes in visual and auditory stimuli: an event-related fmri study. *Neuroimage*, 14(6):1256–67.

Downar, J., Crawley, A. P., Mikulis, D. J., and Davis, K. D. (2002). A cortical network sensitive to stimulus salience in a neutral behavioral context across multiple sensory modalities. *J Neurophysiol*, 87(1):615–20.

Driver, J. and Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. *Neuron*, 57(1):11–23.

Duncan, C. C., Barry, R. J., Connolly, J. F., Fischer, C., Michie, P. T., Naatanen, R., Polich, J., Reinvang, I., and Van Petten, C. (2009). Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, p300, and n400. *Clin Neurophysiol*, 120(11):1883–1908.

Duncan-Johnson, C. C. and Donchin, E. (1982). The p300 component of the event-related brain potential as an index of information processing. *Biol Psychol*, 14(1-2):1–52.

Durschmid, S., Edwards, E., Reichert, C., Dewar, C., Hinrichs, H., Heinze, H. J., Kirsch, H. E., Dalal, S. S., Deouell, L. Y., and Knight, R. T. (2016). Hierarchy of prediction errors for auditory events in human temporal and frontal cortex. *Proc Natl Acad Sci U S A*, 113(24):6755–60.

Edwards, E., Soltani, M., Deouell, L. Y., Berger, M. S., and Knight, R. T. (2005). High gamma activity in response to deviant auditory stimuli recorded directly from human cortex. *J Neurophysiol*, 94(6):4269–80.

Egner, T., Monti, J. M., and Summerfield, C. (2010). Expectation and surprise determine neural population responses in the ventral visual stream. *J Neurosci*, 30(49):16601–8.

Egner, T. and Summerfield, C. (2013). Grounding predictive coding models in empirical neuroscience research. *Behav Brain Sci*, 36(3):210–1.

Ekman, M., Kok, P., and de Lange, F. P. (2017). Time-compressed preplay of anticipated events in human primary visual cortex. *Nat Commun*, 8:15276.

El Karoui, I., King, J. R., Sitt, J., Meyniel, F., Van Gaal, S., Hasboun, D., Adam, C., Navarro, V., Baulac, M., Dehaene, S., Cohen, L., and Naccache, L. (2015). Event-related potential, time-frequency, and functional connectivity facets of local and global auditory novelty processing: An intracranial study in humans. *Cereb Cortex*, 25(11):4203–12.

English, G., Ghasemi Nejad, N., Sommerfelt, M., Yanik, M. F., and von der Behrens, W. (2023). Bayesian surprise shapes neural responses in somatosensory cortical circuits. *Cell Reports*, 42(2).

Escera, C., Alho, K., Schroger, E., and Winkler, I. (2000). Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiol Neurootol*, 5(3-4):151–66.

Escera, C. and Corral, M. (2007). Role of mismatch negativity and novelty-p3 in involuntary auditory attention. *Journal of psychophysiology*, 21(3-4):251–264.

Faraji, M., Preuschoff, K., and Gerstner, W. (2018). Balancing new against old information: The role of puzzlement surprise in learning. *Neural Comput*, 30(1):34–83.

Farashahi, S., Donahue, C. H., Khorsand, P., Seo, H., Lee, D., and Soltani, A. (2017). Metaplasticity as a neural substrate for adaptive learning and choice under uncertainty. *Neuron*, 94(2):401–414 e6.

Fardo, F., Auksztulewicz, R., Allen, M., Dietz, M. J., Roepstorff, A., and Friston, K. J. (2017). Expectation violation and attention to pain jointly modulate neural gain in somatosensory cortex. *Neuroimage*, 153:109–121.

Feldman, H. and Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Front Hum Neurosci*, 4:215.

Felleman, D. J. and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex*, 1(1):1–47.

Ferrari, A., Richter, D., and de Lange, F. P. (2022). Updating contextual sensory expectations for adaptive behavior. *J Neurosci*, 42(47):8855–8869.

Fischer, C., Morlet, D., Bouchet, P., Luaute, J., Jourdan, C., and Salord, F. (1999). Mismatch negativity and late auditory evoked potentials in comatose patients. *Clin Neurophysiol*, 110(9):1601–10.

Fitch, W. T. and Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science*, 303(5656):377–80.

Flandin, G. and Penny, W. D. (2007). Bayesian fmri data analysis with sparse spatial basis function priors. *Neuroimage*, 34(3):1108–25.

Friedman, D., Cycowicz, Y. M., and Gaeta, H. (2001). The novelty p3: an event-related brain potential (erp) sign of the brain's evaluation of novelty. *Neurosci Biobehav Rev*, 25(4):355–73.

Friston, K. (2005). A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci*, 360(1456):815–36.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat Rev Neurosci*, 11(2):127–38.

Friston, K. and Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philos Trans R Soc Lond B Biol Sci*, 364(1521):1211–21.

Friston, K. J. (2019). Waves of prediction. *PLoS Biol*, 17(10):e3000426.

Frost, R., Armstrong, B. C., Siegelman, N., and Christiansen, M. H. (2015). Domain generality versus modality specificity: the paradox of statistical learning. *Trends Cogn Sci*, 19(3):117–25.

Fu, S., Fan, S., and Chen, L. (2003). Event-related potentials reveal involuntary processing of orientation changes in the visual modality. *Psychophysiology*, 40(5):770–5.

Garrido, M. I., Friston, K. J., Kiebel, S. J., Stephan, K. E., Baldeweg, T., and Kilner, J. M. (2008). The functional anatomy of the mmn: a dcm study of the roving paradigm. *Neuroimage*, 42(2):936–44.

Garrido, M. I., Kilner, J. M., Kiebel, S. J., and Friston, K. J. (2007a). Evoked brain responses are generated by feedback loops. *Proc Natl Acad Sci U S A*, 104(52):20961–6.

Garrido, M. I., Kilner, J. M., Kiebel, S. J., and Friston, K. J. (2009a). Dynamic causal modeling of the response to frequency deviants. *J Neurophysiol*, 101(5):2620–31.

Garrido, M. I., Kilner, J. M., Kiebel, S. J., Stephan, K. E., and Friston, K. J. (2007b). Dynamic causal modelling of evoked potentials: a reproducibility study. *Neuroimage*, 36(3):571–80.

Garrido, M. I., Kilner, J. M., Stephan, K. E., and Friston, K. J. (2009b). The mismatch negativity: a review of underlying mechanisms. *Clin Neurophysiol*, 120(3):453–63.

Ghazanfar, A. A. and Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends Cogn Sci*, 10(6):278–85.

Giard, M. H., Lavikahen, J., Reinikainen, K., Perrin, F., Bertrand, O., Pernier, J., and Naatanen, R. (1995). Separate representation of stimulus frequency, intensity, and duration in auditory sensory memory: an event-related potential and dipole-model analysis. *J Cogn Neurosci*, 7(2):133–43.

Giard, M. H., Perrin, F., Pernier, J., and Bouchet, P. (1990). Brain generators implicated in the processing of auditory stimulus deviance: a topographic event-related potential study. *Psychophysiology*, 27(6):627–40.

Glaze, C. M., Kable, J. W., and Gold, J. I. (2015). Normative evidence accumulation in unpredictable environments. *Elife*, 4.

Gomes, H., Bernstein, R., Ritter, W., Vaughan, H. G., J., and Miller, J. (1997). Storage of feature conjunctions in transient auditory memory. *Psychophysiology*, 34(6):712–6.

Gregory, R. L. (1980). Perceptions as hypotheses. *Philos Trans R Soc Lond B Biol Sci*, 290(1038):181–97.

Grill-Spector, K., Henson, R., and Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn Sci*, 10(1):14–23.

Grimm, S., Bendixen, A., Deouell, L. Y., and Schroger, E. (2009). Distraction in a visual multi-deviant paradigm: behavioral and event-related potential effects. *Int J Psychophysiol*, 72(3):260–6.

Grimm, S. and Schroger, E. (2007). The processing of frequency deviations within sounds: evidence for the predictive nature of the mismatch negativity (mmn) system. *Restor Neurol Neurosci*, 25(3-4):241–9.

Haenschel, C., Vernon, D. J., Dwivedi, P., Gruzelier, J. H., and Baldeweg, T. (2005). Event-related brain potential correlates of human auditory sensory memory-trace formation. *J Neurosci*, 25(45):10494–501.

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., and Sporns, O. (2008). Mapping the structural core of human cerebral cortex. *PLoS Biol*, 6(7):e159.

Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*, 298(5598):1569–79.

Hautasaari, P., Kujala, U. M., and Tarkka, I. M. (2019). Detecting differences with magnetoencephalography of somatosensory processing after tactile and electrical stimuli. *J Neurosci Methods*, 311:331–337.

He, C., Hotson, L., and Trainor, L. J. (2009). Development of infant mismatch responses to auditory pattern changes between 2 and 4 months old. *Eur J Neurosci*, 29(4):861–7.

He, X., Zhang, J., Zhang, Z., Go, R., Wu, J., Li, C., Gan, K., and Chen, D. (2020). Effects of visual attentional load on the tactile sensory memory indexed by somatosensory mismatch negativity. *Front Neuroinform*, 14:575078.

Hedge, C., Stothart, G., Todd Jones, J., Rojas Frias, P., Magee, K. L., and Brooks, J. C. (2015). A frontal attention mechanism in the visual mismatch negativity. *Behav Brain Res*, 293:173–81.

Heilbron, M. and Chait, M. (2018). Great expectations: Is there evidence for predictive coding in auditory cortex? *Neuroscience*, 389:54–73.

Heilbron, M. and Meyniel, F. (2019). Confidence resets reveal hierarchical adaptive learning in humans. *PLoS Comput Biol*, 15(4):e1006972.

Heslenfeld, D. J. (2003). Visual mismatch negativity. *Detection of change: Event-related potential and fMRI findings*, pages 41–59.

Horton, J. C. and Hoyt, W. F. (1991). The representation of the visual field in human striate cortex. a revision of the classic holmes map. *Arch Ophthalmol*, 109(6):816–24.

Horvath, J. and Winkler, I. (2004). How the human auditory system treats repetition amongst change. *Neurosci Lett*, 368(2):157–61.

Horvath, J., Winkler, I., and Bendixen, A. (2008). Do n1/mmn, p3a, and ron form a strongly coupled chain reflecting the three stages of auditory distraction? *Biol Psychol*, 79(2):139–47.

Hu, L., Zhao, C., Li, H., and Valentini, E. (2013). Mismatch responses evoked by nociceptive stimuli. *Psychophysiology*, 50(2):158–73.

Huang, M. X., Lee, R. R., Miller, G. A., Thoma, R. J., Hanlon, F. M., Paulson, K. M., Martin, K., Harrington, D. L., Weisend, M. P., Edgar, J. C., and Canive, J. M. (2005). A parietal-frontal network studied by somatosensory oddball meg responses, and its cross-modal consistency. *Neuroimage*, 28(1):99–114.

Hughes, H. C., Darcey, T. M., Barkan, H. I., Williamson, P. D., Roberts, D. W., and Aslin, C. H. (2001). Responses of human auditory association cortex to the omission of an expected acoustic event. *Neuroimage*, 13(6 Pt 1):1073–89.

Igelstrom, K. M. and Graziano, M. S. A. (2017). The inferior parietal lobule and temporoparietal junction: A network perspective. *Neuropsychologia*, 105:70–83.

Iglesias, S., Mathys, C., Brodersen, K. H., Kasper, L., Piccirelli, M., den Ouden, H. E., and Stephan, K. E. (2013). Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. *Neuron*, 80(2):519–30.

Imada, T., Hari, R., Loveless, N., McEvoy, L., and Sams, M. (1993). Determinants of the auditory mismatch response. *Electroencephalogr Clin Neurophysiol*, 87(3):144–53.

Itti, L. and Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Res*, 49(10):1295–306.

Jaaskelainen, I. P., Ahveninen, J., Bonmassar, G., Dale, A. M., Ilmoniemi, R. J., Levanen, S., Lin, F. H., May, P., Melcher, J., Stufflebeam, S., Tiitinen, H., and Belliveau, J. W. (2004). Human posterior auditory cortex gates novel sounds to consciousness. *Proc Natl Acad Sci U S A*, 101(17):6809–14.

Jalewa, J., Todd, J., Michie, P. T., Hodgson, D. M., and Harms, L. (2021). Do rat auditory event related potentials exhibit human mismatch negativity attributes related to predictive coding? *Hear Res*, 399:107992.

Javitt, D. C., Grochowski, S., Shelley, A. M., and Ritter, W. (1998). Impaired mismatch negativity (mmn) generation in schizophrenia as a function of stimulus deviance, probability, and interstimulus/interdeviant interval. *Electroencephalogr Clin Neurophysiol*, 108(2):143–53.

Johnson, M. R., Mitchell, K. J., Raye, C. L., D'Esposito, M., and Johnson, M. K. (2007). A brief thought can modulate activity in extrastriate visual areas: Top-down effects of refreshing just-seen visual stimuli. *Neuroimage*, 37(1):290–9.

Kastner, S., De Weerd, P., Desimone, R., and Ungerleider, L. G. (1998). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional mri. *Science*, 282(5386):108–11.

Kekoni, J., Hamalainen, H., Saarinen, M., Grohn, J., Reinikainen, K., Lehtokoski, A., and Naatanen, R. (1997). Rate effect and mismatch responses in the somatosensory system: Erp-recordings in humans. *Biol Psychol*, 46(2):125–42.

Kenemans, J. L., Hebly, W., van den Heuvel, E. H., and Grent, T. J. T. (2010). Moderate alcohol disrupts a mechanism for detection of rare events in human visual cortex. *J Psychopharmacol*, 24(6):839–45.

Kenemans, J. L., Jong, T. G., and Verbaten, M. N. (2003). Detection of visual change: mismatch or rareness? *Neuroreport*, 14(9):1239–42.

Kiebel, S. J., Daunizeau, J., and Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS Comput Biol*, 4(11):e1000209.

Kim, H. (2014). Involvement of the dorsal and ventral attention networks in oddball stimulus processing: a meta-analysis. *Hum Brain Mapp*, 35(5):2265–84.

Kim, R., Seitz, A., Feenstra, H., and Shams, L. (2009). Testing assumptions of statistical learning: is it long-term and implicit? *Neurosci Lett*, 461(2):145–9.

Kimura, M., Katayama, J., and Murohashi, H. (2005a). Neural correlates of pre-attentive and attentive processing of visual changes. *Neuroreport*, 16(18):2061–4.

Kimura, M., Katayama, J., and Murohashi, H. (2005b). Positive difference in erps reflects independent processing of visual changes. *Psychophysiology*, 42(4):369–79.

Kimura, M., Katayama, J., and Murohashi, H. (2006). Probability-independent and -dependent erps reflecting visual change detection. *Psychophysiology*, 43(2):180–9.

Kimura, M., Katayama, J., Ohira, H., and Schroger, E. (2009). Visual mismatch negativity: new evidence from the equiprobable paradigm. *Psychophysiology*, 46(2):402–9.

Kimura, M., Kondo, H., Ohira, H., and Schroger, E. (2012). Unintentional temporal context-based prediction of emotional faces: an electrophysiological study. *Cereb Cortex*, 22(8):1774–85.

Kimura, M., Ohira, H., and Schroger, E. (2010a). Localizing sensory and cognitive systems for pre-attentive visual deviance detection: an sloreta analysis of the data of kimura et al. (2009). *Neurosci Lett*, 485(3):198–203.

Kimura, M., Schroger, E., and Czigler, I. (2011). Visual mismatch negativity and its importance in visual cognitive sciences. *Neuroreport*, 22(14):669–73.

Kimura, M., Widmann, A., and Schroger, E. (2010b). Human visual system automatically represents large-scale sequential regularities. *Brain Res*, 1317:165–79.

King, J. R., Gramfort, A., Schurger, A., Naccache, L., and Dehaene, S. (2014). Two distinct dynamic modes subtend the detection of unexpected sounds. *PLoS One*, 9(1):e85791.

Knight, R. T. and Scabini, D. (1998). Anatomic bases of event-related potentials and their relationship to novelty detection in humans. *J Clin Neurophysiol*, 15(1):3–13.

Knill, D. C. and Pouget, A. (2004). The bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci*, 27(12):712–9.

Koelsch, S., Busch, T., Jentschke, S., and Rohrmeier, M. (2016). Under the hood of statistical learning: A statistical mmn reflects the magnitude of transitional probabilities in auditory sequences. *Sci Rep*, 6:19741.

Kogo, N. and Trengove, C. (2015). Is predictive coding theory articulated enough to be testable? *Front Comput Neurosci*, 9:111.

Kolossa, A., Fingscheidt, T., Wessel, K., and Kopp, B. (2012). A model-based approach to trial-by-trial p300 amplitude fluctuations. *Front Hum Neurosci*, 6:359.

Kolossa, A., Kopp, B., and Fingscheidt, T. (2015). A computational analysis of the neural bases of bayesian inference. *Neuroimage*, 106:222–37.

Kopp, B., Seer, C., Lange, F., Kluytmans, A., Kolossa, A., Fingscheidt, T., and Hoijtink, H. (2016). P300 amplitude variations, prior probabilities, and likelihoods: A bayesian erp study. *Cogn Affect Behav Neurosci*, 16(5):911–28.

Kording, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., and Shams, L. (2007). Causal inference in multisensory perception. *PLoS One*, 2(9):e943.

Kremlacek, J., Kuba, M., Kubova, Z., and Langrova, J. (2006). Visual mismatch negativity elicited by magnocellular system activation. *Vision Res*, 46(4):485–90.

Lashley, K. S. (1951). *The problem of serial order in behavior*, volume 21. Bobbs-Merrill Oxford.

Lawson, R. P., Mathys, C., and Rees, G. (2017). Adults with autism overestimate the volatility of the sensory environment. *Nat Neurosci*, 20(9):1293–1299.

Lecaignard, F., Bertrand, O., Caclin, A., and Mattout, J. (2022). Neurocomputational underpinnings of expected surprise. *J Neurosci*, 42(3):474–486.

Lee, T. S. and Mumford, D. (2003). Hierarchical bayesian inference in the visual cortex. *J Opt Soc Am A Opt Image Sci Vis*, 20(7):1434–48.

Leonard, M. K., Bouchard, K. E., Tang, C., and Chang, E. F. (2015). Dynamic encoding of speech sequence probability in human temporal cortex. *J Neurosci*, 35(18):7203–14.

Leong, Y. C., Radulescu, A., Daniel, R., DeWoskin, V., and Niv, Y. (2017). Dynamic interaction between reinforcement learning and attention in multidimensional environments. *Neuron*, 93(2):451–463.

Liasis, A., Towell, A., Alho, K., and Boyd, S. (2001). Intracranial identification of an electric frontal-cortex response to auditory stimulus change: a case study. *Brain Res Cogn Brain Res*, 11(2):227–33.

Lieder, F., Daunizeau, J., Garrido, M. I., Friston, K. J., and Stephan, K. E. (2013). Modelling trial-by-trial changes in the mismatch negativity. *PLoS Comput Biol*, 9(2):e1002911.

Linden, D. E. (2005). The p300: where in the brain is it produced and what does it tell us? *Neuroscientist*, 11(6):563–76.

Lorenzo-Lopez, L., Amenedo, E., Pazo-Alvarez, P., and Cadaveira, F. (2004). Pre-attentive detection of motion direction changes in normal aging. *Neuroreport*, 15(17):2633–6.

Lovero, K. L., Simmons, A. N., Aron, J. L., and Paulus, M. P. (2009). Anterior insular cortex anticipates impending stimulus significance. *Neuroimage*, 45(3):976–83.

Luo, D., Liu, J., Auksztulewicz, R., Wing Yip, T. K., Kanold, P. O., and Schnupp, J. W. (2023). Hierarchical deviant processing in auditory cortex of awake mice. *bioRxiv*.

Macdonald, M. and Campbell, K. (2011). Effects of a violation of an expected increase or decrease in intensity on detection of change within an auditory pattern. *Brain Cogn*, 77(3):438–45.

Maekawa, T., Goto, Y., Kinukawa, N., Taniwaki, T., Kanba, S., and Tobimatsu, S. (2005). Functional characterization of mismatch negativity to a visual stimulus. *Clin Neurophysiol*, 116(10):2392–402.

Maekawa, T., Tobimatsu, S., Ogata, K., Onitsuka, T., and Kanba, S. (2009). Preattentive visual change detection as reflected by the mismatch negativity (mmn)–evidence for a memory-based process. *Neurosci Res*, 65(1):107–12.

Maheu, M., Dehaene, S., and Meyniel, F. (2019). Brain signatures of a multiscale process of sequence learning in humans. *Elife*, 8.

Male, A. G., O'Shea, R. P., Schroger, E., Muller, D., Roeber, U., and Widmann, A. (2020). The quest for the genuine visual mismatch negativity (vmmn): Event-related potential indications of deviance detection for low-level visual features. *Psychophysiology*, 57(6):e13576.

Mars, R. B., Debener, S., Gladwin, T. E., Harrison, L. M., Haggard, P., Rothwell, J. C., and Bestmann, S. (2008). Trial-by-trial fluctuations in the event-related electroencephalogram reflect dynamic changes in the degree of surprise. *J Neurosci*, 28(47):12539–45.

Mathys, C., Daunizeau, J., Friston, K. J., and Stephan, K. E. (2011). A bayesian foundation for individual learning under uncertainty. *Front Hum Neurosci*, 5:39.

Mathys, C. D., Lomakina, E. I., Daunizeau, J., Iglesias, S., Brodersen, K. H., Friston, K. J., and Stephan, K. E. (2014). Uncertainty in perception and the hierarchical gaussian filter. *Front Hum Neurosci*, 8:825.

Max, C., Widmann, A., Schroger, E., and Sussman, E. (2015). Effects of explicit knowledge and predictability on auditory distraction and target performance. *Int J Psychophysiol*, 98(2 Pt 1):174–81.

May, P., Tiitinen, H., Ilmoniemi, R. J., Nyman, G., Taylor, J. G., and Naatanen, R. (1999). Frequency change detection in human auditory cortex. *J Comput Neurosci*, 6(2):99–120.

May, P. J. and Tiitinen, H. (2010). Mismatch negativity (mmn), the deviance-elicited auditory deflection, explained. *Psychophysiology*, 47(1):66–122.

McKenzie, D. N. and Barry, R. J. (2006). The independence of memory traces of attended and unattended stimuli. *Cereb Cortex*, 16(11):1566–70.

Melloni, L., Schwiedrzik, C. M., Muller, N., Rodriguez, E., and Singer, W. (2011). Expectations change the signatures and timing of electrophysiological correlates of perceptual awareness. *J Neurosci*, 31(4):1386–96.

Mesulam, M. M. (1998). From sensation to cognition. *Brain*, 121 (Pt 6):1013–52.

Meyer, T. and Olson, C. R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. *Proc Natl Acad Sci U S A*, 108(48):19401–6.

Meyer, T., Ramachandran, S., and Olson, C. R. (2014). Statistical learning of serial visual transitions by neurons in monkey inferotemporal cortex. *J Neurosci*, 34(28):9332–7.

Meyniel, F. (2020). Brain dynamics for confidence-weighted learning. *PLoS Comput Biol*, 16(6):e1007935.

Meyniel, F. and Dehaene, S. (2017). Brain networks for confidence weighting and hierarchical inference during probabilistic learning. *Proc Natl Acad Sci U S A*, 114(19):E3859–E3868.

Meyniel, F., Maheu, M., and Dehaene, S. (2016). Human inferences about sequences: A minimal transition probability model. *PLoS Comput Biol*, 12(12):e1005260.

Meyniel, F., Schlunegger, D., and Dehaene, S. (2015). The sense of confidence during probabilistic learning: A normative account. *PLoS Comput Biol*, 11(6):e1004305.

Millidge, B., Seth, A., and Buckley, C. L. (2021). Predictive coding: a theoretical and experimental review. *arXiv preprint arXiv:2107.12979*.

Mittag, M., Takegata, R., and Winkler, I. (2016). Transitional probabilities are prioritized over stimulus/pattern probabilities in auditory deviance detection: Memory basis for predictive sound processing. *J Neurosci*, 36(37):9572–9.

Mo, L., Xu, G., Kay, P., and Tan, L. H. (2011). Electrophysiological evidence for the left-lateralized effect of language on preattentive categorical perception of color. *Proc Natl Acad Sci U S A*, 108(34):14026–30.

Modirshanechi, A., Brea, J., and Gerstner, W. (2022). A taxonomy of surprise definitions. *Journal of Mathematical Psychology*, 110:102712.

Modirshanechi, A., Kiani, M. M., and Aghajan, H. (2019). Trial-by-trial surprise-decoding model for visual and auditory binary oddball tasks. *Neuroimage*, 196:302–317.

Mohr, H. M., Linder, N. S., Linden, D. E., Kaiser, J., and Sireteanu, R. (2009). Orientation-specific adaptation to mentally generated lines in human visual cortex. *Neuroimage*, 47(1):384–91.

Molholm, S., Martinez, A., Ritter, W., Javitt, D. C., and Foxe, J. J. (2005). The neural circuitry of preattentive auditory change-detection: an fmri study of pitch and duration mismatch negativity generators. *Cereb Cortex*, 15(5):545–51.

Mulders, D., Seymour, B., Mouraux, A., and Mancini, F. (2023). Confidence of probabilistic predictions modulates the cortical response to pain. *Proc Natl Acad Sci U S A*, 120(4):e2212252120.

Mumford, D. (1992). On the computational architecture of the neocortex. ii. the role of cortico-cortical loops. *Biol Cybern*, 66(3):241–51.

Naatanen, R. (2000). Mismatch negativity (mmn): perspectives for application. *Int J Psychophysiol*, 37(1):3–10.

Naatanen, R. (2009). Somatosensory mismatch negativity: a new clinical tool for developmental neurological research? *Dev Med Child Neurol*, 51(12):930–1.

Naatanen, R., Astikainen, P., Ruusuvirta, T., and Huotilainen, M. (2010). Automatic auditory intelligence: an expression of the sensory-cognitive core of cognitive processes. *Brain Res Rev*, 64(1):123–36.

Naatanen, R., Gaillard, A. W., and Mantysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol (Amst)*, 42(4):313–29.

Naatanen, R., Jacobsen, T., and Winkler, I. (2005). Memory-based or afferent processes in mismatch negativity (mmn): a review of the evidence. *Psychophysiology*, 42(1):25–32.

Näätänen, R., Kujala, T., and Light, G. (2019). *Mismatch negativity: a window to the brain*. Oxford University Press.

Naatanen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). The mismatch negativity (mmn) in basic research of central auditory processing: a review. *Clin Neurophysiol*, 118(12):2544–90.

Naatanen, R., Paavilainen, P., Tiitinen, H., Jiang, D., and Alho, K. (1993). Attention and mismatch negativity. *Psychophysiology*, 30(5):436–50.

Naatanen, R. and Picton, T. (1987). The n1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology*, 24(4):375–425.

Naatanen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., and Winkler, I. (2001). "primitive intelligence" in the auditory cortex. *Trends Neurosci*, 24(5):283–8.

Naatanen, R. and Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychol Bull*, 125(6):826–59.

Naeije, G., Vaulet, T., Wens, V., Marty, B., Goldman, S., and De Tiege, X. (2016). Multilevel cortical processing of somatosensory novelty: A magnetoencephalography study. *Front Hum Neurosci*, 10:259.

Naeije, G., Vaulet, T., Wens, V., Marty, B., Goldman, S., and De Tiege, X. (2018). Neural basis of early somatosensory change detection: A magnetoencephalography study. *Brain Topogr*, 31(2):242–256.

Nager, W., Teder-Sälejärvi, W., Kunze, S., and Münte, T. F. (2003). Preattentive evaluation of multiple perceptual streams in human audition. *Neuroreport*, 14(6):871–874.

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., and Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nat Neurosci*, 15(7):1040–6.

Nassar, M. R., Wilson, R. C., Heasly, B., and Gold, J. I. (2010). An approximately bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *J Neurosci*, 30(37):12366–78.

Nelken, I., Rotman, Y., and Bar Yosef, O. (1999). Responses of auditory-cortex neurons to structural features of natural sounds. *Nature*, 397(6715):154–7.

Neumann, O., van der Heijden, A. H., and Allport, D. A. (1986). Visual selective attention: introductory remarks. *Psychol Res*, 48(4):185–8.

Niedernhuber, M., Raimondo, F., Sitt, J. D., and Bekinschtein, T. A. (2022). Sensory target detection at local and global timescales reveals a hierarchy of supramodal dynamics in the human cortex. *J Neurosci*.

Niv, Y., Daniel, R., Geana, A., Gershman, S. J., Leong, Y. C., Radulescu, A., and Wilson, R. C. (2015). Reinforcement learning in multidimensional environments relies on attention mechanisms. *J Neurosci*, 35(21):8145–57.

Noppeney, U. (2021). Perceptual inference, learning, and attention in a multisensory world. *Annu Rev Neurosci*, 44:449–473.

Nordby, H., Roth, W. T., and Pfefferbaum, A. (1988). Event-related potentials to breaks in sequences of alternating pitches or interstimulus intervals. *Psychophysiology*, 25(3):262–8.

Näätänen, R. (1992). *Attention and brain function*. Psychology Press.

Näätänen, R. and Gaillard, A. (1983). *5 The orienting reflex and the N2 deflection of the event-related potential (ERP)*, volume 10, pages 119–141. Elsevier.

Opitz, B., Rinne, T., Mecklinger, A., von Cramon, D. Y., and Schroger, E. (2002). Differential contribution of frontal and temporal cortices to auditory change detection: fmri and erp results. *Neuroimage*, 15(1):167–74.

O'Reilly, J. X., Schuffelgen, U., Cuell, S. F., Behrens, T. E., Mars, R. B., and Rushworth, M. F. (2013). Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proc Natl Acad Sci U S A*, 110(38):E3660–9.

Ossmy, O., Moran, R., Pfeffer, T., Tsetsos, K., Usher, M., and Donner, T. H. (2013). The timescale of perceptual evidence integration can be adapted to the environment. *Curr Biol*, 23(11):981–6.

Ostwald, D., Spitzer, B., Guggenmos, M., Schmidt, T. T., Kiebel, S. J., and Blankenburg, F. (2012). Evidence for neural encoding of bayesian surprise in human somatosensation. *Neuroimage*, 62(1):177–88.

Paavilainen, P. (2013). The mismatch-negativity (mmn) component of the auditory event-related potential to violations of abstract regularities: a review. *Int J Psychophysiol*, 88(2):109–23.

Paavilainen, P., Arajarvi, P., and Takegata, R. (2007). Preattentive detection of nonsalient contingencies between auditory features. *Neuroreport*, 18(2):159–63.

Paavilainen, P., Jaramillo, M., Naatanen, R., and Winkler, I. (1999). Neuronal populations in the human brain extracting invariant relationships from acoustic variance. *Neurosci Lett*, 265(3):179–82.

Paavilainen, P., Saarinen, J., Tervaniemi, M., and Näätänen, R. (1995). Mismatch negativity to changes in abstract sound features during dichotic listening. *Journal of Psychophysiology*.

Paavilainen, P., Simola, J., Jaramillo, M., Näätänen, R., and Winkler, I. (2001). Preattentive extraction of abstract feature conjunctions from auditory stimulation as reflected by the mismatch negativity (mmn). *Psychophysiology*, 38(2):359–365.

Parmentier, F. B., Elsley, J. V., Andres, P., and Barcelo, F. (2011). Why are auditory novels distracting? contrasting the roles of novelty, violation of expectation and stimulus change. *Cognition*, 119(3):374–80.

Parr, T., Kilner, J., and Friston, K. (2023). Functional asymmetry and the consequences of action: Comment on: Left and right temporal-parietal junctions (tpjs) as "match/mismatch" hedonic machines: A unifying account of tpj function by fabrizio doricchi et al. *Phys Life Rev*, 44:145–147.

Parras, G. G., Casado-Roman, L., Schroger, E., and Malmierca, M. S. (2021). The posterior auditory field is the chief generator of prediction error signals in the auditory cortex. *Neuroimage*, 242:118446.

Parras, G. G., Nieto-Diego, J., Carbajal, G. V., Valdes-Baizabal, C., Escera, C., and Malmierca, M. S. (2017). Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. *Nat Commun*, 8(1):2148.

Pazo-Alvarez, P., Amenedo, E., and Cadaveira, F. (2004a). Automatic detection of motion direction changes in the human brain. *Eur J Neurosci*, 19(7):1978–86.

Pazo-Alvarez, P., Amenedo, E., Lorenzo-Lopez, L., and Cadaveira, F. (2004b). Effects of stimulus location on automatic detection of changes in motion direction in the human brain. *Neurosci Lett*, 371(2-3):111–6.

Pazo-Alvarez, P., Cadaveira, F., and Amenedo, E. (2003). Mmn in the visual modality: a review. *Biological Psychology*, 63(3):199–236.

Penny, W., Kiebel, S., and Friston, K. (2003). Variational bayesian inference for fmri time series. *Neuroimage*, 19(3):727–41.

Penny, W. D. (2012). Comparing dynamic causal models using aic, bic and free energy. *Neuroimage*, 59(1):319–30.

Penny, W. D., Stephan, K. E., Daunizeau, J., Rosa, M. J., Friston, K. J., Schofield, T. M., and Leff, A. P. (2010). Comparing families of dynamic causal models. *PLoS Comput Biol*, 6(3):e1000709.

Penny, W. D., Trujillo-Barreto, N. J., and Friston, K. J. (2005). Bayesian fmri time series analysis with spatial priors. *Neuroimage*, 24(2):350–62.

Perruchet, P. and Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *Trends Cogn Sci*, 10(5):233–8.

Phillips, H. N., Blenkmann, A., Hughes, L. E., Bekinschtein, T. A., and Rowe, J. B. (2015). Hierarchical organization of frontotemporal networks for the prediction of stimuli across multiple dimensions. *J Neurosci*, 35(25):9255–64.

Phillips, H. N., Blenkmann, A., Hughes, L. E., Kochen, S., Bekinschtein, T. A., Cam, C., and Rowe, J. B. (2016). Convergent evidence for hierarchical prediction networks from human electrocorticography and magnetoencephalography. *Cortex*, 82:192–205.

Picton, T. W. (1992). The p300 wave of the human event-related potential. *J Clin Neurophysiol*, 9(4):456–79.

Picton, T. W., Alain, C., Otten, L., Ritter, W., and Achim, A. (2000). Mismatch negativity: different water in the same river. *Audiol Neurootol*, 5(3-4):111–39.

Planton, S. and Dehaene, S. (2021). Cerebral representation of sequence patterns across multiple presentation formats. *Cortex*, 145:13–36.

Polich, J. (2007). Updating p300: an integrative theory of p3a and p3b. *Clin Neurophysiol*, 118(10):2128–48.

Prete, D. A., Heikoop, D., McGillivray, J. E., Reilly, J. P., and Trainor, L. J. (2022). The sound of silence: Predictive error responses to unexpected sound omission in adults. *European Journal of Neuroscience*, 55(8):1972–1985.

Purves, D. E., Augustine, G. J., Fitzpatrick, D. E., Hall, W. C., LaMantia, A.-S. E., McNamara, J. O., and White, L. E. (2008). Neuroscience.

Rao, R. P. and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci*, 2(1):79–87.

Reber, A. S. and Lewis, S. (1977). Implicit learning: An analysis of the form and structure of a body of tacit knowledge. *Cognition*, 5(4):333–361.

Restuccia, D., Della Marca, G., Valeriani, M., Leggio, M. G., and Molinari, M. (2007). Cerebellar damage impairs detection of somatosensory input changes. a somatosensory mismatch-negativity study. *Brain*, 130(Pt 1):276–87.

Restuccia, D., Zanini, S., Cazzagon, M., Del Piero, I., Martucci, L., and Della Marca, G. (2009). Somatosensory mismatch negativity in healthy children. *Dev Med Child Neurol*, 51(12):991–8.

Rigoux, L., Stephan, K. E., Friston, K. J., and Daunizeau, J. (2014). Bayesian model selection for group studies - revisited. *Neuroimage*, 84:971–85.

Rinne, T., Degerman, A., and Alho, K. (2005). Superior temporal and inferior frontal cortices are activated by infrequent sound duration decrements: an fmri study. *Neuroimage*, 26(1):66–72.

Rinne, T., Sarkka, A., Degerman, A., Schroger, E., and Alho, K. (2006). Two separate mechanisms underlie auditory change detection and involuntary control of attention. *Brain Res*, 1077(1):135–43.

Ritter, W., Paavilainen, P., Lavikainen, J., Reinikainen, K., Alho, K., Sams, M., and Naatanen, R. (1992). Event-related potentials to repetition and change of auditory stimuli. *Electroencephalogr Clin Neurophysiol*, 83(5):306–21.

Ritter, W., Sussman, E., Deacon, D., Cowan, N., and Vaughan, H. G., J. (1999). Two cognitive systems simultaneously prepared for opposite events. *Psychophysiology*, 36(6):835–8.

Ritter, W., Sussman, E., and Molholm, S. (2000). Evidence that the mismatch negativity system works on the basis of objects. *Neuroreport*, 11(1):61–3.

Rohe, T. and Noppeney, U. (2015). Cortical hierarchies perform bayesian causal inference in multisensory perception. *PLoS Biol*, 13(2):e1002073.

Rosburg, T., Trautner, P., Dietl, T., Korzyukov, O. A., Boutros, N. N., Schaller, C., Elger, C. E., and Kurthen, M. (2005). Subdural recordings of the mismatch negativity (mmn) in patients with focal epilepsy. *Brain*, 128(Pt 4):819–28.

Rosen, R. and Rosen, R. (2012). *Anticipatory systems*. Springer.

Rouhani, N. and Niv, Y. (2021). Signed and unsigned reward prediction errors dynamically enhance learning and memory. *Elife*, 10.

Rubin, J., Ulanovsky, N., Nelken, I., and Tishby, N. (2016). The representation of prediction error in auditory cortex. *PLoS Comput Biol*, 12(8):e1005058.

Ruderman, D. L. and Bialek, W. (1994). Statistics of natural images: Scaling in the woods. *Phys Rev Lett*, 73(6):814–817.

Ruusuvirta, T., Huotilainen, M., Fellman, V., and Naatanen, R. (2003). The newborn human brain binds sound features together. *Neuroreport*, 14(16):2117–9.

Ruusuvirta, T., Huotilainen, M., Fellman, V., and Naatanen, R. (2004). Newborn human brain identifies repeated auditory feature conjunctions of low sequential probability. *Eur J Neurosci*, 20(10):2819–21.

Ruusuvirta, T., Huotilainen, M., and Naatanen, R. (2007a). Preperceptual human number sense for sequential sounds, as revealed by mismatch negativity brain response? *Cereb Cortex*, 17(12):2777–9.

Ruusuvirta, T., Koivisto, K., Wikgren, J., and Astikainen, P. (2007b). Processing of melodic contours in urethane-anaesthetized rats. *Eur J Neurosci*, 26(3):701–3.

Ryali, C. K., Reddy, G., and Yu, A. J. (2018). Demystifying excessively volatile human learning: A bayesian persistent prior and a neural approximation. *Adv Neural Inf Process Syst*, 31:2781–2790.

Saarinen, J., Paavilainen, P., Schoger, E., Tervaniemi, M., and Naatanen, R. (1992). Representation of abstract attributes of auditory stimuli in the human brain. *Neuroreport*, 3(12):1149–51.

Sabri, M., Kareken, D. A., Dzemidzic, M., Lowe, M. J., and Melara, R. D. (2004). Neural correlates of auditory sensory memory and automatic change detection. *Neuroimage*, 21(1):69–74.

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294):1926–8.

Saffran, J. R. and Thiessen, E. D. (2007). Domain-general learning capacities.

Salisbury, D. F. (2012). Finding the missing stimulus mismatch negativity (mmn): emitted mmn to violations of an auditory gestalt. *Psychophysiology*, 49(4):544–8.

Sams, M., Alho, K., and Naatanen, R. (1983). Sequential effects on the erp in discriminating two stimuli. *Biol Psychol*, 17(1):41–58.

Sams, M., Paavilainen, P., Alho, K., and Naatanen, R. (1985). Auditory frequency discrimination and event-related potentials. *Electroencephalogr Clin Neurophysiol*, 62(6):437–48.

SanMiguel, I., Saupe, K., and Schroger, E. (2013a). I know what is missing here: electrophysiological prediction error signals elicited by omissions of predicted "what" but not "when". *Front Hum Neurosci*, 7:407.

SanMiguel, I., Widmann, A., Bendixen, A., Trujillo-Barreto, N., and Schroger, E. (2013b). Hearing silences: human auditory processing relies on preactivation of sound-specific brain activity patterns. *J Neurosci*, 33(20):8633–9.

Santolin, C. and Saffran, J. R. (2018). Constraints on statistical learning across species. *Trends Cogn Sci*, 22(1):52–63.

Schlossmacher, I., Lucka, F., Peters, A., Bruchmann, M., and Straube, T. (2022). Effects of awareness and task relevance on neurocomputational models of mismatch negativity generation. *Neuroimage*, 262:119530.

Schroger, E., Bendixen, A., Denham, S. L., Mill, R. W., Bohm, T. M., and Winkler, I. (2014). Predictive regularity representations in violation detection and auditory stream segregation: from conceptual to computational models. *Brain Topogr*, 27(4):565–77.

Schroger, E., Bendixen, A., Trujillo-Barreto, N. J., and Roeber, U. (2007). Processing of abstract rule violations in audition. *PLoS One*, 2(11):e1131.

Seer, C., Lange, F., Boos, M., Dengler, R., and Kopp, B. (2016). Prior probabilities modulate cortical surprise responses: A study of event-related potentials. *Brain Cogn*, 106:78–89.

Seghier, M. L. (2013). The angular gyrus: multiple functions and multiple subdivisions. *Neuroscientist*, 19(1):43–61.

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., Rosen, B. R., and Tootell, R. B. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, 268(5212):889–93.

Sereno, M. I. and Huang, R. S. (2014). Multisensory maps in parietal cortex. *Curr Opin Neurobiol*, 24(1):39–46.

Seth, A. K., Suzuki, K., and Critchley, H. D. (2011). An interoceptive predictive coding model of conscious presence. *Front Psychol*, 2:395.

Shams, L. and Beierholm, U. (2022). Bayesian causal inference: A unifying neuroscience theory. *Neurosci Biobehav Rev*, 137:104619.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423.

Shen, G., Meltzoff, A. N., Weiss, S. M., and Marshall, P. J. (2020). Body representation in infants: Categorical boundaries of body parts as assessed by somatosensory mismatch negativity. *Dev Cogn Neurosci*, 44:100795.

Shen, G., Smyk, N. J., Meltzoff, A. N., and Marshall, P. J. (2018a). Neuropsychology of human body parts: Exploring categorical boundaries of tactile perception using somatosensory mismatch responses. *J Cogn Neurosci*, 30(12):1858–1869.

Shen, G., Smyk, N. J., Meltzoff, A. N., and Marshall, P. J. (2018b). Using somatosensory mismatch responses as a window into somatotopic processing of tactile stimulation. *Psychophysiology*, 55(5):e13030.

Sherman, B. E., Graves, K. N., and Turk-Browne, N. B. (2020). The prevalence and importance of statistical learning in human cognition and behavior. *Curr Opin Behav Sci*, 32:15–20.

Shinozaki, N., Yabe, H., Sato, Y., Sutoh, T., Hiruma, T., Nashida, T., and Kaneko, S. (2000). Mismatch negativity (mmn) reveals sound grouping in the human brain. *Neuroreport*, 11(8):1597–1601.

Shinozaki, N., Yabe, H., Sutoh, T., Hiruma, T., and Kaneko, S. (1998). Somatosensory automatic responses to deviant stimuli. *Brain Res Cogn Brain Res*, 7(2):165–71.

Sinnett, S., Spence, C., and Soto-Faraco, S. (2007). Visual dominance and attention: the colavita effect revisited. *Percept Psychophys*, 69(5):673–86.

Sohoglu, E. and Chait, M. (2016). Detecting and representing predictable structure during auditory scene analysis. *Elife*, 5.

Spackman, L. A., Boyd, S. G., and Towell, A. (2007). Effects of stimulus frequency and duration on somatosensory discrimination responses. *Exp Brain Res*, 177(1):21–30.

Spackman, L. A., Towell, A., and Boyd, S. G. (2010). Somatosensory discrimination: an intracranial event-related potential study of children with refractory epilepsy. *Brain Res*, 1310:68–76.

Squires, N. K., Squires, K. C., and Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalogr Clin Neurophysiol*, 38(4):387–401.

Stagg, C., Hindley, P., Tales, A., and Butler, S. (2004). Visual mismatch negativity: the detection of stimulus change. *Neuroreport*, 15(4):659–63.

Stefanics, G., Heinzle, J., Horvath, A. A., and Stephan, K. E. (2018). Visual mismatch and predictive coding: A computational single-trial erp study. *J Neurosci*, 38(16):4020–4030.

Stefanics, G., Kimura, M., and Czigler, I. (2011). Visual mismatch negativity reveals automatic detection of sequential regularity violation. *Front Hum Neurosci*, 5:46.

Stefanics, G., Kremlacek, J., and Czigler, I. (2014). Visual mismatch negativity: a predictive coding view. *Front Hum Neurosci*, 8:666.

Stefanics, G., Stephan, K. E., and Heinzle, J. (2019). Feature-specific prediction errors for visual mismatch. *Neuroimage*, 196:142–151.

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., and Friston, K. J. (2009). Bayesian model selection for group studies. *Neuroimage*, 46(4):1004–17.

Stephan, K. E., Penny, W. D., Moran, R. J., den Ouden, H. E., Daunizeau, J., and Friston, K. J. (2010). Ten simple rules for dynamic causal modeling. *Neuroimage*, 49(4):3099–109.

Strasburger, H., Rentschler, I., and Juttner, M. (2011). Peripheral vision and pattern recognition: a review. *J Vis*, 11(5):13.

Strommer, J. M., Tarkka, I. M., and Astikainen, P. (2014). Somatosensory mismatch response in young and elderly adults. *Front Aging Neurosci*, 6:293.

Sulykos, I. and Czigler, I. (2011). One plus one is less than two: visual features elicit non-additive mismatch-related brain activity. *Brain Res*, 1398:64–71.

Summerfield, C. and de Lange, F. P. (2014). Expectation in perceptual decision making: neural and computational mechanisms. *Nat Rev Neurosci*, 15(11):745–56.

Summerfield, C. and Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends Cogn Sci*, 13(9):403–9.

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., and Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nat Neurosci*, 11(9):1004–6.

Summerfield, C., Wyart, V., Johnen, V. M., and de Gardelle, V. (2011). Human scalp electroencephalography reveals that repetition suppression varies with expectation. *Front Hum Neurosci*, 5:67.

Sussman, E., Ritter, W., and Vaughan, H. G., J. (1998). Predictability of stimulus deviance and the mismatch negativity. *Neuroreport*, 9(18):4167–70.

Sussman, E., Winkler, I., Huotilainen, M., Ritter, W., and Naatanen, R. (2002). Top-down effects can modify the initially stimulus-driven auditory organization. *Brain Res Cogn Brain Res*, 13(3):393–405.

Sussman, E. S. (2005). Integration and segregation in auditory scene analysis. *J Acoust Soc Am*, 117(3 Pt 1):1285–98.

Sussman, E. S. (2007). A new view on the mmn and attention debate: The role of context in processing auditory events. *Journal of Psychophysiology*, 21(3-4):164–175.

Sussman, E. S. (2013). Attention matters: Pitch vs. pattern processing in adolescence. *Front Psychol*, 4:333.

Sutton, S. (1979). P300–thirteen years later. *Evoked brain potentials and behavior*, pages 107–126.

Sutton, S., Braren, M., Zubin, J., and John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, 150(3700):1187–8.

Sutton, S., Tueting, P., Zubin, J., and John, E. R. (1967). Information delivery and the sensory evoked potential. *Science*, 155(3768):1436–9.

Takegata, R. and Morotomi, T. (1999). Integrated neural representation of sound and temporal features in human auditory sensory memory: an event-related potential study. *Neurosci Lett*, 274(3):207–10.

Tales, A. and Butler, S. (2006). Visual mismatch negativity highlights abnormal preattentive visual processing in alzheimer's disease. *Neuroreport*, 17(9):887–90.

Tales, A., Newton, P., Troscianko, T., and Butler, S. (1999). Mismatch negativity in the visual modality. *Neuroreport*, 10(16):3363–7.

Tales, A., Troscianko, T., Wilcock, G. K., Newton, P., and Butler, S. R. (2002). Age-related changes in the preattentional detection of visual change. *Neuroreport*, 13(7):969–72.

Tervaniemi, M., Maury, S., and Naatanen, R. (1994). Neural representations of abstract stimulus features in the human brain as reflected by the mismatch negativity. *Neuroreport*, 5(7):844–6.

Tiitinen, H., May, P., Reinikainen, K., and Naatanen, R. (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature*, 372(6501):90–2.

Todorovic, A. and de Lange, F. P. (2012). Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. *J Neurosci*, 32(39):13389–95.

Todorovic, A., van Ede, F., Maris, E., and de Lange, F. P. (2011). Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an meg study. *J Neurosci*, 31(25):9118–23.

Tomasi, D. and Volkow, N. D. (2011). Association between functional connectivity hubs and brain networks. *Cereb Cortex*, 21(9):2003–13.

Trejo, L. J., Ryan-Jones, D. L., and Kramer, A. F. (1995). Attentional modulation of the mismatch negativity elicited by frequency differences between binaurally presented tone bursts. *Psychophysiology*, 32(4):319–28.

Tse, C. Y., Shum, Y. H., Xiao, X. Z., and Wang, Y. (2021). Fronto-occipital mismatch responses in pre-attentive detection of visual changes: Implication on a generic brain network underlying mismatch negativity (mmn). *Neuroimage*, 244:118633.

Uhrig, L., Dehaene, S., and Jarraya, B. (2014). A hierarchy of responses to auditory regularities in the macaque brain. *J Neurosci*, 34(4):1127–32.

Ulanovsky, N., Las, L., Farkas, D., and Nelken, I. (2004). Multiple time scales of adaptation in auditory cortex neurons. *J Neurosci*, 24(46):10440–53.

Ulanovsky, N., Las, L., and Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nat Neurosci*, 6(4):391–8.

Urakawa, T., Inui, K., Yamashiro, K., and Kakigi, R. (2010). Cortical dynamics of the visual change detection process. *Psychophysiology*, 47(5):905–12.

van Kerkoerle, T., Self, M. W., Dagnino, B., Gariel-Mathis, M. A., Poort, J., van der Togt, C., and Roelfsema, P. R. (2014). Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. *Proc Natl Acad Sci U S A*, 111(40):14332–41.

van Zuijen, T. L., Simoens, V. L., Paavilainen, P., Naatanen, R., and Tervaniemi, M. (2006). Implicit, intuitive, and explicit knowledge of abstract regularities in a sound sequence: an event-related brain potential study. *J Cogn Neurosci*, 18(8):1292–303.

von Carlowitz-Ghori, K. M., Hohlefeld, F. U., Bayraktaroglu, Z., Curio, G., and Nikulin, V. V. (2011). Effect of complete stimulus predictability on p3 and n2 components: an electroencephalographic study. *Neuroreport*, 22(9):459–63.

von Helmholtz, H. (1867). *Handbuch der physiologischen Optik*, volume 9. Voss.

Wacongne, C., Changeux, J. P., and Dehaene, S. (2012). A neuronal model of predictive coding accounting for the mismatch negativity. *J Neurosci*, 32(11):3665–78.

Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., and Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proc Natl Acad Sci U S A*, 108(51):20754–9.

Wang, L., Amalric, M., Fang, W., Jiang, X., Pallier, C., Figueira, S., Sigman, M., and Dehaene, S. (2019). Representation of spatial sequences using nested rules in human prefrontal cortex. *Neuroimage*, 186:245–255.

Wang, L., Uhrig, L., Jarraya, B., and Dehaene, S. (2015). Representation of numerical and sequential patterns in macaque and human brains. *Curr Biol*, 25(15):1966–74.

Wang, R., Shen, Y., Tino, P., Welchman, A. E., and Kourtzi, Z. (2017). Learning predictive statistics: Strategies and brain mechanisms. *J Neurosci*, 37(35):8412–8427.

Weber, L. A., Diaconescu, A. O., Mathys, C., Schmidt, A., Kometer, M., Vollenweider, F., and Stephan, K. E. (2020). Ketamine affects prediction errors about statistical regularities: A computational single-trial analysis of the mismatch negativity. *J Neurosci*, 40(29):5658–5668.

Weber, L. A., Tomiello, S., Schobi, D., Wellstein, K. V., Mueller, D., Iglesias, S., and Stephan, K. E. (2022). Auditory mismatch responses are differentially sensitive to changes in muscarinic acetylcholine versus dopamine receptor function. *Elife*, 11.

Wessel, J. R. and Huber, D. E. (2019). Frontal cortex tracks surprise separately for different sensory modalities but engages a common inhibitory control mechanism. *PLoS Comput Biol*, 15(7):e1006927.

Wetzel, N. and Schroger, E. (2014). On the development of auditory distraction: A review. *Psych J*, 3(1):72–91.

Wijnen, V. J., van Boxtel, G. J., Eilander, H. J., and de Gelder, B. (2007). Mismatch negativity predicts recovery from the vegetative state. *Clin Neurophysiol*, 118(3):597–605.

Winkler, I. (2007). Interpreting the mismatch negativity. *Journal of psychophysiology*, 21(3-4):147–163.

Winkler, I. and Czigler, I. (2012). Evidence from auditory and visual event-related potential (erp) studies of deviance detection (mmn and vmmn) linking predictive coding theories and perceptual object representations. *Int J Psychophysiol*, 83(2):132–43.

Winkler, I., Denham, S. L., and Nelken, I. (2009). Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends Cogn Sci*, 13(12):532–40.

Winkler, I., Kushnerenko, E., Horvath, J., Ceponiene, R., Fellman, V., Huotilainen, M., Naatanen, R., and Sussman, E. (2003). Newborn infants can organize the auditory world. *Proc Natl Acad Sci U S A*, 100(20):11812–5.

Winkler, I., Paavilainen, P., and Naatanen, R. (1992). Can echoic memory store two traces simultaneously? a study of event-related brain potentials. *Psychophysiology*, 29(3):337–49.

Winkler, I., Takegata, R., and Sussman, E. (2005). Event-related brain potentials reveal multiple stages in the perceptual organization of sound. *Brain Res Cogn Brain Res*, 25(1):291–9.

Woldorff, M. G., Hackley, S. A., and Hillyard, S. A. (1991). The effects of channel-selective attention on the mismatch negativity wave elicited by deviant tones. *Psychophysiology*, 28(1):30–42.

Woldorff, M. G., Hillyard, S. A., Gallen, C. C., Hampson, S. R., and Bloom, F. E. (1998). Magnetoencephalographic recordings demonstrate attentional modulation of mismatch-related neural activity in human auditory cortex. *Psychophysiology*, 35(3):283–92.

Xu, H. A., Modirshanechi, A., Lehmann, M. P., Gerstner, W., and Herzog, M. H. (2021a). Novelty is not surprise: Human exploratory and adaptive behavior in sequential decision-making. *PLoS Comput Biol*, 17(6):e1009070.

Xu, Q., Ye, C., Hamalainen, J. A., Ruohonen, E. M., Li, X., and Astikainen, P. (2021b). Magnetoencephalography responses to unpredictable and predictable rare somatosensory stimuli in healthy adult humans. *Front Hum Neurosci*, 15:641273.

Yabe, H., Tervaniemi, M., Reinikainen, K., and Naatanen, R. (1997). Temporal window of integration revealed by mmn to sound omission. *Neuroreport*, 8(8):1971–4.

Yamaguchi, S. and Knight, R. T. (1991). Age effects on the p300 to novel somatosensory stimuli. *Electroencephalogr Clin Neurophysiol*, 78(4):297–301.

Yamaguchi, S. and Knight, R. T. (1992). Effects of temporal-parietal lesions on the somatosensory p3 to lower limb stimulation. *Electroencephalogr Clin Neurophysiol*, 84(2):139–48.

Yamashiro, K., Inui, K., Otsuru, N., Kida, T., Akatsuka, K., and Kakigi, R. (2008). Somatosensory off-response in humans: an erp study. *Exp Brain Res*, 190(2):207–13.

Yaron, A., Hershenhoren, I., and Nelken, I. (2012). Sensitivity to complex statistical regularities in rat auditory cortex. *Neuron*, 76(3):603–15.

Yu, A. J. and Cohen, J. D. (2008). Sequential effects: Superstition or rational behavior? *Adv Neural Inf Process Syst*, 21:1873–1880.

Yu, A. J. and Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, 46(4):681–92.

Yucel, G., McCarthy, G., and Belger, A. (2007). fmri reveals that involuntary visual deviance processing is resource limited. *Neuroimage*, 34(3):1245–52.

Yucel, G., Petty, C., McCarthy, G., and Belger, A. (2005). Graded visual attention modulates brain responses evoked by task-irrelevant auditory pitch changes. *J Cogn Neurosci*, 17(12):1819–28.

Zachlod, D., Ruttgers, B., Bludau, S., Mohlberg, H., Langner, R., Zilles, K., and Amunts, K. (2020). Four new cytoarchitectonic areas surrounding the primary and early auditory cortex in human brains. *Cortex*, 128:1–21.

Zhang, Z., Guo, G., Zhang, J., Li, C., Huang, Q., Fukuyama, H., Funahashi, S., Yan, T., and Wu, J. (2019). Do theta oscillations explain the somatosensory change detection mechanism? *Biol Psychol*, 143:103–112.

Zhao, J., Al-Aidroos, N., and Turk-Browne, N. B. (2013). Attention is spontaneously biased toward regularities. *Psychol Sci*, 24(5):667–77.

Zvyagintsev, M., Zweerings, J., Sarkheil, P., Bergert, S., Baqapuri, H., Neuner, I., Gaebler, A. J., and Mathiak, K. (2020). Auditory mismatch processing: Role of paradigm and stimulus characteristics as detected by fmri. *Biol Psychol*, 154:107887.

Appendix

Original Publication of Study 1

Gijsen*, S., **Grundei*, M.**, Lange, R. T., Ostwald, D. and Blankenburg, F. (2021). Neural surprise in somatosensory Bayesian learning. *PLoS computational biology, 17(2)*, e1008068.
*equal contribution

PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Neural surprise in somatosensory Bayesian learning

Sam Gijsen[1,4☯]*, **Miro Grundei**[1,4☯]*, **Robert T. Lange**[2,5], **Dirk Ostwald**[3], **Felix Blankenburg**[1]

1 Neurocomputation and Neuroimaging Unit, Freie Universität Berlin, Germany, **2** Berlin Institute of Technology, Berlin, Germany, **3** Computational Cognitive Neuroscience, Freie Universität Berlin, Germany, **4** Humboldt-Universität zu Berlin, Faculty of Philosophy, Berlin School of Mind and Brain, Berlin, Germany, **5** Einstein Center for Neurosciences, Berlin, Germany

☯ These authors contributed equally to this work.
* sam.gijsen@fu-berlin.de (SG); m.grundei@fu-berlin.de (MG)

Citation: Gijsen S, Grundei M, Lange RT, Ostwald D, Blankenburg F (2021) Neural surprise in somatosensory Bayesian learning. PLoS Comput Biol 17(2): e1008068. https://doi.org/10.1371/journal.pcbi.1008068

Editor: Philipp Schwartenbeck, UCL, UNITED KINGDOM

Received: June 12, 2020

Accepted: December 18, 2020

Published: February 2, 2021

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pcbi.1008068

Data Availability Statement: The full, raw dataset can be found at: https://osf.io/83pgq/ with DOI 10.17605/OSF.IO/83PGQ The analysis and modeling code can be found at: https://github.com/SamGijsen/SurpriseInSomesthesis.

Abstract

Tracking statistical regularities of the environment is important for shaping human behavior and perception. Evidence suggests that the brain learns environmental dependencies using Bayesian principles. However, much remains unknown about the employed algorithms, for somesthesis in particular. Here, we describe the cortical dynamics of the somatosensory learning system to investigate both the form of the generative model as well as its neural surprise signatures. Specifically, we recorded EEG data from 40 participants subjected to a somatosensory roving-stimulus paradigm and performed single-trial modeling across peristimulus time in both sensor and source space. Our Bayesian model selection procedure indicates that evoked potentials are best described by a non-hierarchical learning model that tracks transitions between observations using leaky integration. From around 70ms post-stimulus onset, secondary somatosensory cortices are found to represent confidence-corrected surprise as a measure of model inadequacy. Indications of Bayesian surprise encoding, reflecting model updating, are found in primary somatosensory cortex from around 140ms. This dissociation is compatible with the idea that early surprise signals may control subsequent model update rates. In sum, our findings support the hypothesis that early somatosensory processing reflects Bayesian perceptual learning and contribute to an understanding of its underlying mechanisms.

Author summary

Our environment features statistical regularities, such as a drop of rain predicting imminent rainfall. Despite the importance for behavior and survival, much remains unknown about how these dependencies are learned, particularly for somatosensation. As surprise signalling about novel observations indicates a mismatch between one's beliefs and the world, it has been hypothesized that surprise computation plays an important role in perceptual learning. By analyzing EEG data from human participants receiving sequences of tactile stimulation, we compare different formulations of surprise and investigate the

Funding: This work was supported by Deutscher Akademischer Austauschdienst (SG, https://www.daad.de/en/), Humboldt-Universität zu Berlin, Faculty of Philosophy, Berlin School of Mind and Brain (SG & MG, http://www.mind-and-brain.de/home/), and Einstein Center for Neurosciences Berlin (RTL, https://www.ecn-berlin.de/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

employed underlying learning model. Our results indicate that the brain estimates transitions between observations. Furthermore, we identified different signatures of surprise computation and thereby provide a dissociation of the neural correlates of belief inadequacy and belief updating. Specifically, early surprise responses from around 70ms were found to signal the need for changes to the model, with encoding of its subsequent updating occurring from around 140ms. These results provide insights into how somatosensory surprise signals may contribute to the learning of environmental statistics.

Introduction

The world is governed by statistical regularities, such that a single drop of rain on the skin might predict further tactile sensations through imminent rainfall. The learning of such probabilistic dependencies facilitates adaptive behaviour and ultimately survival. Building on ideas tracing back to Helmholtz [1], it has been suggested that the brain employs an internal generative model of the environment which generates predictions of future sensory input. More recent accounts of perception and perceptual learning, including predictive coding [2, 3] and the free energy principle [4], propose that these models are continuously updated in light of new sensory evidence using Bayesian inference. Under such a view, the generative model is composed of a likelihood function of sensory input given external causes and a prior probability distribution over causes [4, 5]. Perception is interpreted as the computation of a posterior distribution over causes of sensory input and model parameters, while perceptual learning is seen as the updating of the prior distribution based on the computed posterior [6]. Such a description of Bayesian perceptual learning has been successfully used to explain aspects of learning in the auditory [7, 8, 9], visual [10, 11, 12], as well as somatosensory domain [13].

To investigate the underlying neuronal dynamics of perceptual inference, predictions formed by the brain can be probed by violating statistical regularities. Widely researched neurobiological markers of regularity violation include EEG components such as the auditory mismatch negativity (aMMN) and the P300 in response to deviant stimuli following regularity inducing standard stimuli. As an alternative to the oddball paradigm typically used to elicit such mismatch responses (MMR's) [14], the roving-stimulus paradigm features stimulus sequences that alternate between different trains of repeated identical stimuli [15]. Expectations are built up across a train of stimuli of variable length and are subsequently violated by alternating to a different stimulus train. The paradigm thereby allows for the study of MMR's based on the sequence history and independently of the physical stimulus properties. Analogues to the aMMN have also been reported for vision [16] and somatosensation (sMMN). The sMMN was first reported by Kekoni et al. [17] and has since been shown in response to deviant stimuli with different properties, including spatial location [18, 19, 20, 21, 22, 23, 24, 25, 26], vibrotactile frequency [17, 27, 28, 29], and stimulus duration [30, 31]. Increasing evidence has been reported for an account of the MMN as a reflection of Bayesian perceptual learning processes for the auditory [8, 32, 33], visual [12, 16], and to a lesser extent the somatosensory domain [13]. However, the precise mechanisms remain unknown, as it is unclear whether the MMN reflects the signaling of the inadequacy of the current beliefs or their adjustment, due to the lack of direct comparisons between these competing accounts.

In the context of probabilistic inference, the signalling of a mismatch between predicted and observed sensory input may be formally described using computational quantities of surprise [6, 34]. By adopting the vocabulary introduced by Faraji et al. [35] surprise can be grouped into two classes: puzzlement and enlightenment surprise. Puzzlement surprise refers

to the initial realization of a mismatch between the world and an internal model. Predictive surprise (PS) captures this concept based on the measure of information as introduced by Shannon [36]. Specifically, PS considers the belief about the probability of an event such that the occurrence of a rare event (i.e. an event estimated to have low probability of occurrence) is more informative and results in greater surprise. Confidence-corrected surprise (CS), as introduced by Faraji et al. [35] extends the concept of puzzlement surprise by additionally considering belief commitment. It quantifies the idea that surprise elicited by events depends on both the estimated probability of occurrence as well as the confidence in this estimate, with greater confidence leading to higher surprise. For example, in order for the percept of a drop of rain on the skin to be surprising, commitment to a belief about a clear sky may be necessary. The concept of enlightenment surprise, on the other hand, directly relates to the size of the update of the world model that may follow initial puzzlement. Bayesian surprise (BS) captures this notion by quantifying the degree to which an observer adapts their internal generative model in order to accommodate novel observations [37, 38].

Both predictive surprise [9] and Bayesian surprise [13] have been successfully applied to the full time-window of peri-stimulus EEG data to model neural surprise signals. However, the majority of studies have focused on P300 amplitudes, with applications of both predictive surprise [39, 40, 41, 42] and Bayesian surprise [40, 43, 44]. Earlier EEG signals have received less attention, although the MMN was reported to reflect PS [42]. Furthermore, due to the close relationship between model updating and prediction violation, only few studies have attempted to dissociate their signals. Although the use of different surprise functions in principle allows for a direct comparison of the computations potentially underlying EEG mismatch responses, such studies remain scarce. Previous research either focused on their spatial identification using fMRI [11, 45, 46, 47] or temporally specific, late EEG components [40]. Finally, to the best of our knowledge, only one recent pre-print study compared all three prominent surprise functions in a reanalysis of existing data, reporting PS to be better decoded across the entire post stimulus time-window [48].

Despite the successful account of perceptual learning using Bayesian approaches, the framework is broad and much remains unclear about the nature of MMR's, their description as surprise signals, and the underlying generative models that give rise to them. This is especially the case for the somatosensory modality, though evidence has been reported for the encoding of Bayesian surprise using the roving paradigm [13]. The current study expands on this work by recording EEG responses to a roving paradigm formulated as a generative model with discrete hidden states. We explore different mismatch responses, including the somatosensory analogue to the MMN, independent of the physical properties of stimuli. Using single-trial modeling, we systematically investigate the structure of the generative model employed by the brain. Having established the most likely probabilistic model, we provide a spatiotemporal description of its different surprise signatures in electrode and source space. As direct comparisons are scarce, we thus contribute by dissecting the dynamics of multiple aspects of Bayesian computation utilized for somatosensory learning across peri-stimulus time by incorporating them into one hierarchical analysis.

Materials and methods

Ethics statement

The study was approved by the local ethics committee of the Freie Universität Berlin (internal reference number: 51/2013) and written informed consent was obtained from all subjects prior to the experiment.

Experimental design

Participants. 44 healthy volunteers (18-38 years old, mean age: 26, 28 females, all right-handed) participated for monetary compensation of 10 Euro per hour or an equivalent in course credit.

Experimental procedure. In order to study somatosensory mismatch responses and model them as single-trial surprise signals, we used a roving-stimulus paradigm [15]. Stimuli were applied in consecutive trains of alternating stimuli based on a probabilistic model (see below) with an inter-stimulus interval of 750ms (see Fig 1). Trains of stimuli consisted of two possible stimulation intensities. The first and last stimulus in a train were labeled as a deviant and standard, respectively. Thus, as opposed to a classic oddball design, the roving paradigm allows for both stimulus types to function as a standard or deviant.

Adhesive electrodes (GVB-geliMED GmbH, Bad Segeberg, Germany) were attached to the wrist through which the electrical stimuli with a 0.2ms duration were administered. In order to account for interpersonal differences in sensory thresholds, the two intensity levels were determined on a subject basis. The low intensity level (mean $5.05mA \pm 1.88$) was set in proximity to the detection threshold yet so that stimuli were clearly perceivable. The high intensity level (mean $7.16mA \pm 1.73$) was determined for each subject to be easily distinguishable from the low intensity level, yet remaining non-painful and below the motor threshold. The catch stimulus (described below) featured a threefold repetition of the 0.2ms stimulus at an interval of 50ms and was presented at either the low or high intensity level with equal probability.

Following familiarization with the electrical stimulation, 800 stimuli were administered in each of 5 experimental runs à 10 minutes. To ensure the subjects maintained attention on the electrical stimulation, they were instructed to count the number of catch trials (targets). In order to make the task non-trivial, the probability of the occurrence of a catch stimulus was set to either 0.01, 0.015, 0.02, 0.025, or 0.03, corresponding to a range of 3-32 trials per run. A subject received a stimulus sequence corresponding to each catch trial probability only once, with the order randomized between subjects. Following an experimental run, subjects indicated their counted number of catch trials and received feedback in the form of the correct amount.

EEG data collection and preprocessing. Data were collected using a 64-channel active electrode system (ActiveTwo, BioSemi, Amsterdam, Netherlands) at a sampling rate of 2048Hz, with head electrodes placed in accordance to the extended 10-20 system. Individual electrode positions were digitalized and recorded using an electrode positioning system (zebris Medical GmbH, Isny, Germany) with respect to three fiducial markers placed on the subject's face; left and right preauricular points and the nasion. This approach aided subsequent source reconstruction analyses.

Preprocessing was performed using SPM12 (Wellcome Trust Centre for Neuroimaging, Institute for Neurology, University College London, London, UK) and in-house scripts. First, the data were referenced against the average reference, high-pass filtered (0.01Hz), and down-sampled to 512Hz. Consequently, eye-blinks were corrected using a topological confound approach [49] and epoched using a peri-stimulus time interval of -100 to 600ms. All trials were then visually inspected and removed in case any significant artefacts were deemed to be present. The EEG data of four subjects were found to contain excessive noise due to hardware issues, resulting in their omission from further analyses and leaving 40 subjects. Finally, a low-pass filter was applied (45Hz). Grand mean somatosensory evoked potentials (SEPs) were calculated for deviant stimuli ('deviants') and for the standard stimuli directly preceding a deviant to balance the number of trials ('standards'). The preproccesed EEG data was baseline corrected with respect to the pre-stimulus interval of -100 to -5 ms. For the GLM analyses, each trial of the electrode data was subsequently linearly interpolated into a 32x32 plane for each

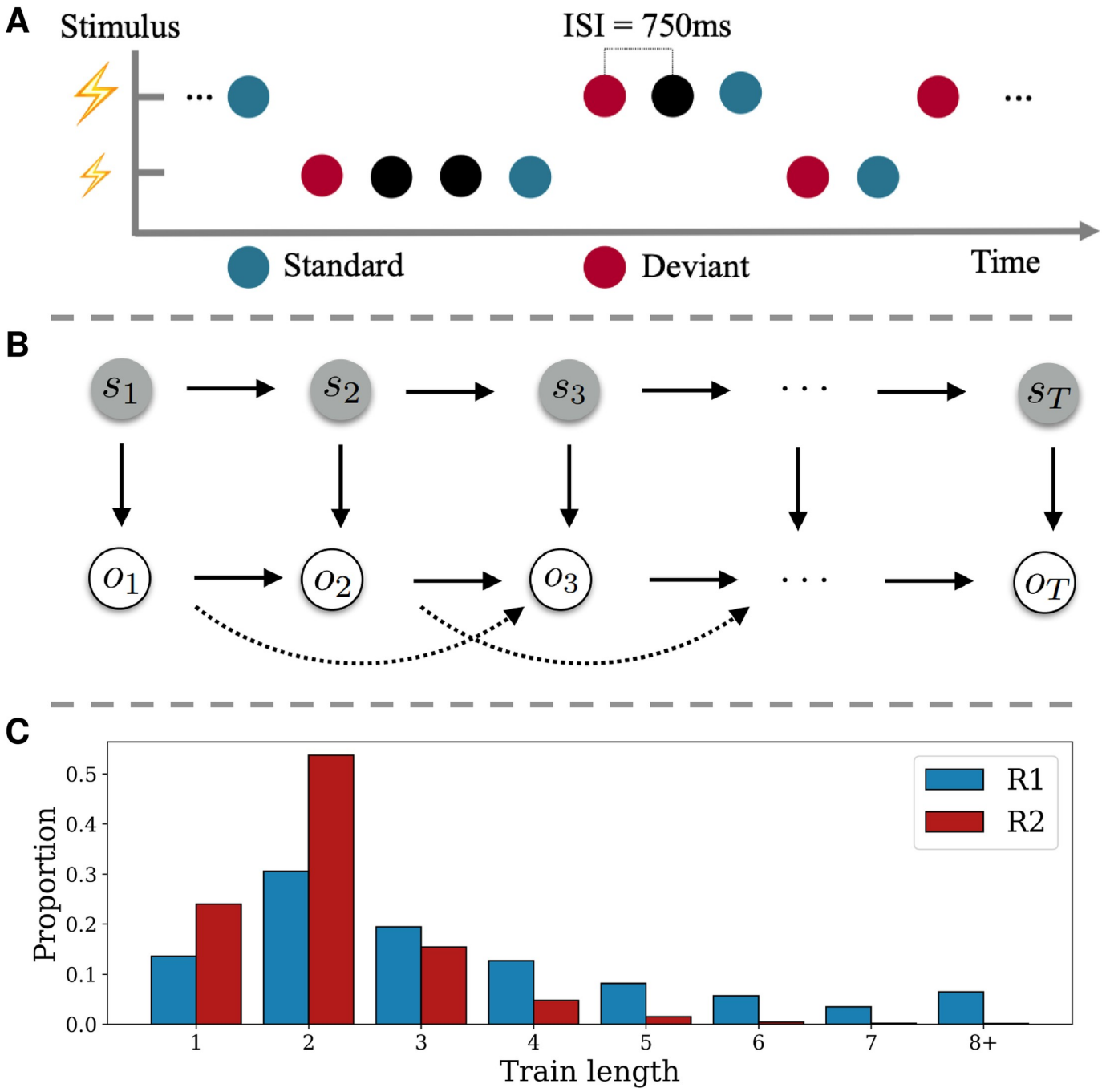


Fig 1. Experimental design and stimulus generation. A) Presentation of experimental stimuli using a roving-stimulus paradigm. Stimuli with two different intensities are presented. Their role as standard or deviant depends on their respective position within the presentation sequence. B) Graphical model of data-generating process. Upper row depicts the evolution of states s_t over time according to a Markov chain. The states emit observations o_t (lower row), which themselves feature second order dependencies on the observation level. C) Average proportion of resulting stimuli train lengths. Higher proportion of shorter trains for the fast switching regime (R_2; red) and more distributed proportion across higher train lengths for the slow switching regime (R_1; blue).

https://doi.org/10.1371/journal.pcbi.1008068.g001

timepoint, resulting in a 32x32x308 image per trial. To allow for the use of random field theory to control for family-wise errors, the images were smoothed with a 12 by 12 mm full-width half-maximum (FWHM) Gaussian kernel. Catch trials were omitted for both the ERP and single-trial analyses.

Generation of stimuli sequences

A property of generative models that is highly relevant for learning in dynamic environments is the manner by which they may adapt their estimated statistics in the face of environmental changes. By incorporating occasional switches between sets of sequence statistics, we aimed to compare generative models that embody different mechanisms of adapting to such change-points. Specifically, the sequential presentation of the stimuli originated from a partially observable probabilistic model for which the hidden state evolved according to a Markov chain (Fig 1) with 3 states s. The state transition ($p(s_t|s_{t-1})$) and emission probabilities $p(o_t|o_{t-1}, o_{t-2}, s_t)$ of the observations o are listed in Table 1. One of the states was observable as it was guaranteed to emit a catch trial, while the other two states were latent, resembling fast and slow switching regimes. As the latter was specified with higher transition probabilities associated with repeating observations ($p(0|00)$ and $p(0|01)$) it thus produced longer stimulus trains on average. For every run, the sequence was initialized by starting either in the slow or fast switching regime with equal probability ($p(s_1) = \{0.5, 0.5, 0\}$, with catch probability being 0) and likewise producing a high or low stimulus with equal probability ($p(o_1|s_1) = \{0.5, 0.5\}$).

Event-related potentials

To investigate the event-related response to the experimental conditions on the EEG data, the statistical design was implemented with the general linear model using SPM12. On the first level, the single-trial data of each participant was subject to a multiple regression approach with several regressors each coding for a level of an experimental variable: stimulus type (levels: standard and deviant), train length (levels: 2, 3, 4, 5, >6 stimuli) and a factor of experimental block as nuisance regressors (levels: block 1-5). An additional GLM with a balanced number of standard and deviant trials for the regimes (levels: fast and slow switching regime) showed no effect of regime or interaction of regime and stimulus type. The restricted maximum likelihood estimation implemented in SPM12 yielded β-parameter estimates for each model regressor over (scalp-)space and time which were further analysed at the group level. The second level consisted of a mass-univariate multiple regression analysis of the individual β scalp-time images with a design matrix specifying regressors for stimulus type and regime as well as parametric regressors for train length and block and an additional subject factor. The

Table 1. Data-generating process.

State transition matrix				Sampling distribution
	R_1	R_2	R_3	$p(o_t\|o_{t-1}, o_{t-2}, s_t)$
R_1	$0.99 - \frac{1}{2}p(c)$	$0.01 - \frac{1}{2}p(c)$	$p(c)$	$p(0\|00) = 0.65$, $p(0\|01) = 0.85$, $p(0\|10) = 0.15$, $p(0\|11) = 0.35$
R_2	$0.01 - \frac{1}{2}p(c)$	$0.99 - \frac{1}{2}p(c)$	$p(c)$	$p(0\|00) = 0.3$, $p(0\|01) = 0.75$, $p(0\|10) = 0.25$, $p(0\|11) = 0.7$
R_3	$0.5 - \frac{1}{2}p(c)$	$0.5 - \frac{1}{2}p(c)$	$p(c)$	$p(2) = 1$

Left: The state transition matrix. Right: Sampling distribution of the slow switching (R_1), fast switching (R_2), and catch-trial regime (R_3), emitting low intensity ($o_t = 0$), high intensity ($o_t = 1$), and catch stimuli ($o_t = 2$, with $p(c) = p(o_t = 2)$). Complementary probabilities are omitted (e.g. $p(1|00) = 1 - p(0|00)$).

https://doi.org/10.1371/journal.pcbi.1008068.t001

condition contrasts were then computed by weighted summation of the group level regressors' β estimates. To control for multiple comparisons, the scalp-time images were corrected with SPM's random field theory-based family wise error correction (FWE) [50]. The significant peaks of the GLM were further inspected by looking at their effect of train length and the corresponding β-parameter estimates of each train length were subjected to a linear fit for visualization purposes.

Distributed source localization

In order to establish the somatosensory system as the driving dipolar generator of the EEG signals prior to 200ms, we followed a two-stage source reconstruction analysis consisting of a distributed and an equivalent current dipole (ECD) approach. While we report and model later EEG components in sensor-space, we refrained from source localizing these, as they most likely originate from a more distributed network of multiple sources [51, 52]. Furthermore, the somatosensory system has been shown to be involved in mismatch processing in the time window prior to 200ms [18, 19, 23, 26, 30, 53].

The distributed source reconstruction algorithm as implemented in SPM12 was used to determine the sources of the ERP's on a subject level. Specifically, subject-specific forward models were created using a 8196 vertex template cortical mesh which was co-registered with the electrode positions using the three aforementioned fiducial markers. SPM12's BEM EEG head model was used to construct the forward model's lead field. The multiple sparse priors under group constraints were implemented for the subject-specific source estimates [54, 55]. These were subsequently analyzed at the group level using one-sample t-tests. The yielded statistical parametric maps were thresholded at the peak level with $p < 0.05$ after FWE correction. The anatomical correspondence of the MNI coordinates of the cluster peaks were verified via cytoarchitectonic references using the SPM Anatomy toolbox. Details of the distributed source reconstruction can be reviewed in the results section.

Equivalent current dipole fitting & source projection

The results of the distributed source reconstruction were subsequently used to fit ECDs to the grand average ERP data using the variational Bayes ECD fitting algorithm implemented in SPM12. The MNI coordinates resulting from the distributed source reconstruction served as informed location priors with variance of $10mm^2$ to optimize the location and orientation of the dipoles for a time-window around the peak of each component of interest (shown in the results section). For the primary somatosensory cortex (S1), two individual dipoles were fit to the time windows of the N20 and P50 components, respectively, to differentiate two sources of early somatosensory processing. Furthermore, a symmetrical dipolar source was fit to the peak of the N140 component of the evoked response with an informed prior around the secondary somatosensory cortex. Subsequently, the single trial EEG data of each subject was projected with the ECD lead fields onto the 4 sources using SPM12, which enabled model selection analyses in source-space.

Trial-by-trial modeling of sensor- and source-space EEG data

Sequential Bayesian learner models for categorical data. To compare Bayesian learners in terms of their generative models and surprise signals, we specified various probabilistic models which generate the regressors ultimately fitted to the EEG data. Capitalizing on the occasional changes to the sequence statistics included in the experimental stimulus generating model, we assess two approaches to latent state inference. Specifically, a conjugate Dirichlet-Categorical (DC) model as well as a Hidden Markov Model (HMM) [56] were used for

modeling categorical data. The DC model is non-hierarchical and does not feature any explicit detection of the regime-switches. However, it is able to adapt its estimated statistics to account for sequence change-points by favoring recent observations over those in the past, akin to a progressive "forgetting" or leaky integration. The model assumes a real-valued, static hidden state s_t that is shared across time for each observation emission.

In contrast, the HMM is a hierarchical model for which s_t is a discrete variable and assumed to follow a first order Markov Chain, mimicking the data generation process. As such, it contains additional assumptions about the task structure, which allows for flexible adaptation following a regime-switch by performing inference over a set of discrete hidden states K ($s_t \in \{1, \ldots, K\}$). The transition dynamics are given by the row-stochastic matrix $\mathbf{A} \in \mathbb{R}^{K\times K}$ with $a_{ij} \geq 0$ and $\sum_{j=1}^{K} a_{ij} = 1$:

$$p(s_t|s_{t-1}) = \mathbf{A} \Leftrightarrow p(s_t^j|s_{t-1}^i) = a_{ij} \text{ for } t = 1, \ldots, T. \tag{1}$$

Within our two model classes, we differentiate between four probabilistic models. Here, the aim is to investigate which sequence statistics are estimated by the generative model. In the case of Stimulus Probability (SP) inference, the model does not capture any Markov dependence: o_t solely depends on s_t. Alternation Probability (AP) inference captures a limited form of first-order Markov dependency, by estimating the probability of the event of altering observations d_t given the hidden state s_t and the previous observation o_{t-1}, where $d_t = \mathbf{1}_{o_t \neq o_{t-1}}$ takes on the value 1 if the current observation o_t differs from o_{t-1}. With Transition Probability (TP_1) inference, the model accounts for full first-order Markov dependence and estimates separate alternation probabilities depending on o_{t-1} and s_t, i.e. $p(o_t|o_{t-1}, s_t)$. Finally, TP_1 inference may be extended (TP_2) to also depend on o_{t-2}, and by estimating $p(o_t|s_t, o_{t-1}, o_{t-2})$ it most closely resembles the structure underlying the data generation.

Dirichlet-Categorical model. The Dirichlet-Categorical model is a simple Bayesian observer that counts the observations of each unique type to determine its best guess of their probability (Eq 5). Its exponential forgetting parameter implements a gradual discounting of observations the further in the past they occurred (Eq 8). It is part of the Bayesian conjugate pairs and models the likelihood of the observations using the Categorical distribution with {1, . . ., M} different possible realizations per sample y_t. Given the probability vector $\mathbf{s} = \{s_1, \ldots, s_M\}$ defined on the $M - 1$ dimensional simplex $\mathcal{S}_{M-1}$ with $s_i > 0$ and $\sum_{j=1}^{M} s_j = 1$, the probability mass function of an event is given by

$$p(y_t = j|s_1, \ldots, s_M) = s_j \tag{2}$$

Furthermore, the prior distribution over the hidden state $\mathbf{s}$ is given by the Dirichlet distribution which is parametrized by the probability vector $\alpha = \{\alpha_1, \ldots, \alpha_M\}$:

$$p(s_1, \ldots, s_M|\alpha_1, \ldots \alpha_M) = \frac{\Gamma(\sum_{j=1}^{M} \alpha_j)}{\prod_{j=1}^{M} \Gamma(\alpha_j)} \prod_{j=1}^{M} s_j^{\alpha_j - 1}. \tag{3}$$

Hence, we have a Dirichlet prior with $s_1, \ldots, s_M \sim Dir(\alpha_1, \ldots, \alpha_M)$ and a Categorical likelihood with $y \sim Cat(s_1, \ldots, s_M)$. Given a sequence of observations $y_1, \ldots, y_t$ the model then combines the likelihood evidence with prior beliefs in order to refine posterior estimates over the latent variable space (derivations of enumerated formulas may be found in the

supplementary material S1 Appendix):

$$\begin{aligned} p(s_1,\ldots,s_M|y_1,\ldots,y_t) &\propto p(s_1,\ldots,s_M|\alpha_1,\ldots,\alpha_M)\prod_{i=1}^{t}p(y_i|s_1,\ldots,s_M) \\ &= \prod_{j=1}^{M} s_j^{\alpha_j-1+\sum_{i=1}^{t}\mathbf{1}\{y_i=j\}} \end{aligned} \tag{4}$$

Since the Dirichlet prior and Categorical likelihood pair follow the concept of conjugacy, given an initial $\alpha^0 = \{\alpha_1^0,\ldots,\alpha_M^0\}$ (set as a hyperparameter) the filtering distribution can be computed:

$$p(\mathbf{s}_t|y_1,\ldots,y_t) = p(s_1,\ldots,s_M|y_1,\ldots,y_t) = Dir(\alpha^t) \quad \text{with} \quad \alpha_j^t = \alpha_j^0 + \sum_{i=1}^{t}\mathbf{1}\{y_i=j\}. \tag{5}$$

Likewise, one can easily obtain the posterior predictive distribution (needed to compute the predictive surprise readout) by integrating over the space of latent states:

$$\begin{aligned} p(y_t=x|y_1,\ldots,y_{t-1}) &= \int p(y_t=x|s_1,\ldots,s_M)p(s_1,\ldots,s_M|y_1,\ldots,y_{t-1})d\mathcal{S}_M \\ &= \frac{\alpha_x^t}{\sum_{j=1}^{M}\alpha_j^t} \end{aligned} \tag{6}$$

We can evaluate the likelihood of a specific sequence of events which can be used to iteratively compute the posterior:

$$p(y_1,\ldots,y_t) = p(y_1)\prod_{i=2}^{t}p(y_i|y_{1:i}) = \frac{1}{M}\prod_{i=2}^{t}\prod_{j=1}^{M}\frac{\alpha_j^i}{\sum_{k=1}^{M}\alpha_k^i} \tag{7}$$

For the evaluation of the posterior distributions, we differentiate between three inference types which track different statistics of the incoming sequence as described above (for a graphical model see Fig 2):

1. The stimulus probability (SP) model: $y_t = o_t$ for $t = 1, \ldots, T$
2. The alternation probability (AP) model: $y_t = d_t$ for $t = 2, \ldots, T$
3. The transition probability model (TP_1 & TP_2): $y_t = o_t$ for $t = 1, \ldots, T$ with a set of hidden parameters $\mathbf{s}_1^{(i)}$ for each transition from $o_{t-1} = i$ and $\mathbf{s}_2^{(j)}$ for each transition from $o_{t-2} = j$ respectively

Despite a static latent state representation, the DC model may account for hidden dynamics by incorporating an exponential memory-decay parameter $\tau \in [0, 1]$ which discounts observations the further in the past they occurred. Functioning as an exponential forgetting mechanism, it allows for the specification of different timescales of observation integration.

$$\begin{aligned} p(\mathbf{s}_t|y_1,\ldots,y_t) &= p(s_1,\ldots,s_M|y_1,\ldots,y_t) = Dir(\alpha^t) \\ \text{with} \quad \alpha_j^t &= \alpha_j^0 + \sum_{i=1}^{t}e^{-\tau(t-i)}\mathbf{1}\{y_i=j\}. \end{aligned} \tag{8}$$

Hidden Markov model. While the Dirichlet-Categorical model provides a simple yet expressive conjugate Bayesian model for which analytical posterior expressions exist, it is limited in the functionality of the latent state s due to its interpretation as the discrete distribution

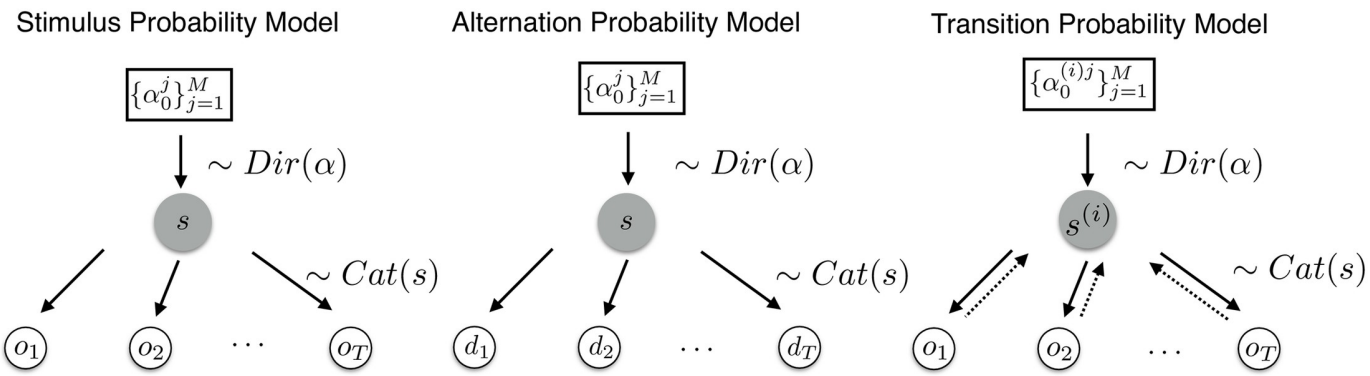


Fig 2. Dirichlet-Categorical model as a graphical model. Left: The stimulus probability model which tracks the hidden state vector determining the sampling process of the raw observations. Middle: The alternation probability model which infers the hidden state distribution based on alternations of the observations. Right: The transition probability model which assumes a different data-generating process based on the previous observations. Hence, it infers M sets of probability vectors α^i.

https://doi.org/10.1371/journal.pcbi.1008068.g002

over categories. Hidden Markov Models (HMMs), on the other hand, are able to capture the dynamics of the hidden state with the transition probabilities of a Markov Chain (MC). Given the hidden state at time t, the categorical observation o_t is sampled according to the stochastic matrix $\mathbf{B} \in \mathbb{R}^{M\times K}$, containing the emission probabilities, $p(o_t|s_t)$. The evolution of the discrete hidden state according to a MC, $p(s_t|s_{t-1})$, is described by the stochastic matrix $\mathbf{A} \in \mathbb{R}^{K\times K}$. The initial hidden state $p(s_1)$ is sampled according to the distribution vector $\pi \in \mathbb{R}^K$. $\mathbf{A}$, $\mathbf{B}$ are both row stochastic, hence $A_{ij}, B_{ij} \geq 0$, $\sum_{j=1}^K A_{ij} = 1$ and $\sum_{j=1}^M B_{ij} = 1$. The graphical model described by the HMM setup is thereby specified as depicted in Fig 3.

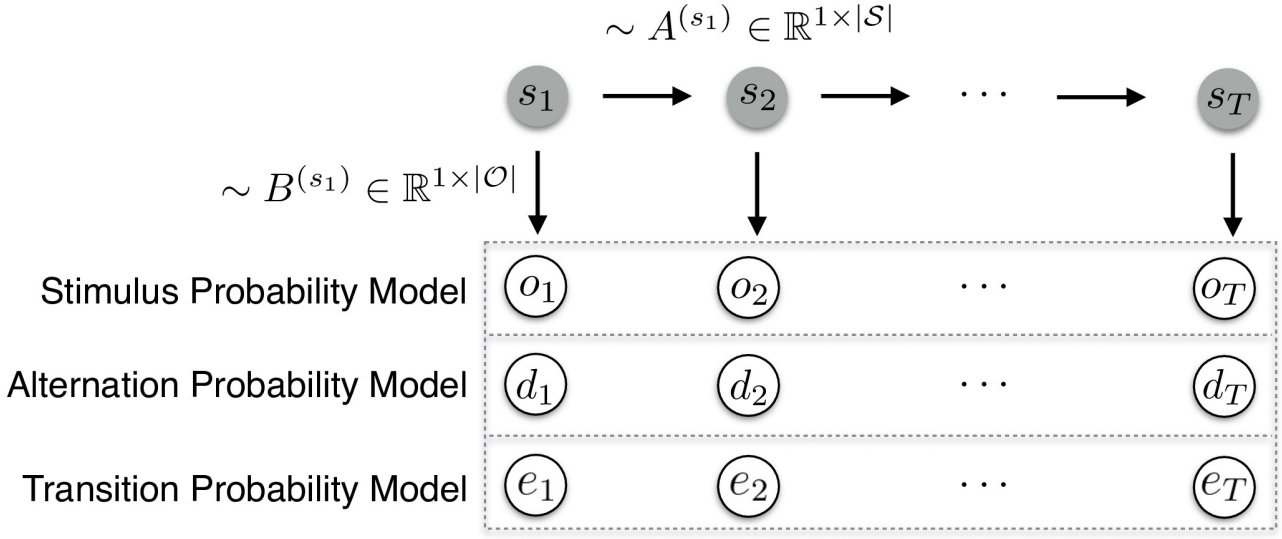


Fig 3. Hidden Markov model as a graphical model. Upper row depicts the evolution of states s_t according to the transition matrix $A^{(s_t)}$. The states emit observational data (dotted rectangle) according to the probabilities specified in stochastic matrix $B^{(s_t)}$ which depends on the type of inference. The stimulus probability model infers the emission probabilities associated with the raw observations o_t. The alternation probability model tracks the alternations of observations with $d_t = \mathbf{1}_{o_t \neq o_{t-1}}$. The transition probability model assumes a data-generating process based on previous observations, with e_t coding for the transitions between observations.

https://doi.org/10.1371/journal.pcbi.1008068.g003

Classically, the parameters of this latent variable are inferred using the Expectation-Maximisation (EM) algorithm. Therefore, and in order to derive the factorisation of the joint likelihood $p(o_{1:t}, s_{1:t})$, the backward and forward probabilities are used in conjunction with the Baum-Welch algorithm in order to perform the inference procedure (see S1 Appendix).

HMM Implementation. The aim of the HMM was to approximate the data generation process more closely by using a model capable of learning the regimes over time and performing latent state inference at each timestep. To this end, prior knowledge was used in its specification by fixing the state transition matrix close to its true values ($p(s_t = s_{t-1}) = 0.99$). The rare catch trials were removed from the data prior to fitting the HMM and thus their accompanying third regime was omitted, resulting in a two-state HMM. Given that an HMM estimates emission probabilities of the form $p(o_t|s_t)$ and thus does not capture any additional explicit dependency on previous observations, the input vector of observations was transformed prior to fitting the models. For AP and TP inference this equated to re-coding the observation o_t to reflect the specific event that occurred. Specifically, for the AP model the input sequence was $d_t = 1_{o_t \neq o_{t-1}}$, while for TP_1 and TP_2 a vector of events was used corresponding to the four possible transitions from o_{t-1} or eight transitions from o_{t-2} respectively. Thus, the HMM estimates two sets (reflecting the two latent states) of emission probabilities which correspond to these events (y_t). Despite this deviation of the fitted models from the underlying data generation process, the AP and TP models reliably captured R_1 and R_2 to their capability, with TP_2 retrieving the true, but unknown underlying emission probabilities (see S1 Fig). As expected, SP inference was agnostic to the regimes, while AP and TP inference allowed for the tracking of the latent state over time (S1 Fig). An example of the filtering posterior may be found in Fig 4.

Surprise readouts. For each of the probabilistic models described above, three different surprise functions were implemented, forming the predictors for the EEG data: predictive surprise $PS(y_t)$, Bayesian surprise $BS(y_t)$, and confidence-corrected surprise $CS(y_t)$. These may be

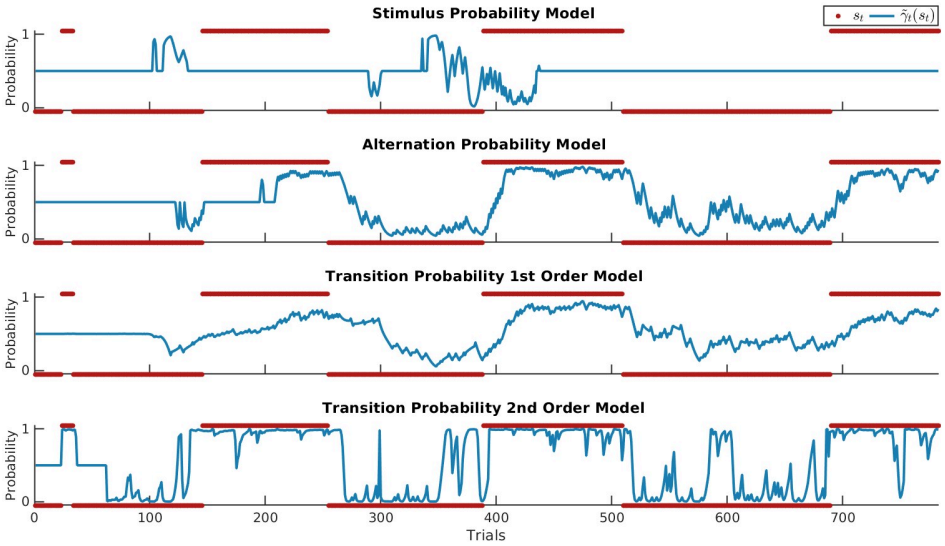


Fig 4. Posterior probabilities of the HMM. Comparison of the filtering posterior $\hat{\gamma}_t(s_t) = p(s_t|o_1, \ldots, o_t)$ of the different HMM inference models for an example sequence. The true, but unknown regimes of the data generation process are plotted in red. Note that, as the regimes were balanced in terms of stimulus probabilities, SP inference is not able to capture the underlying regimes and instead attempts to dissociate two states based on empirical differences in observed stimulus probabilities.

https://doi.org/10.1371/journal.pcbi.1008068.g004

interpreted as read-out functions of the generative model, signalling a mismatch between the world and the internal model.

The predictive surprise is defined as the negative logarithm of the posterior predictive distribution $p(y_t|s_t)$:

$$PS(y_t) \coloneqq -\ln p(y_t|s_t) = -\ln p(y_t|y_1, \ldots, y_{t-1}). \tag{9}$$

A posterior that assigns little probability to an event y_t will cause high (unit-less) predictive surprise and as such is a measure of puzzlement surprise. The Bayesian surprise, on the other hand, quantifies enlightenment surprise and is defined as the Kullback-Leibler (KL) divergence between the posterior pre- and post-update:

$$BS(y_t) \coloneqq KL(p(s_{t-1}|y_{t-1}, \ldots y_1) \| p(s_t|y_t, \ldots, y_1)) \tag{10}$$

Confidence-corrected surprise is an extended definition of puzzlement surprise which additionally considers the commitment of the generative model as it is scaled by the negative entropy of the prior distribution. It is defined as the KL divergence between the informed prior and posterior distribution of a naive observer, corresponding to an agent with a flat prior $\hat{p}(s_t)$ (i.e. all outcomes are equally likely) which observed y_t:

$$CS(y_t) \coloneqq KL(p(s_t) \| \hat{p}(s_t|y_t)), \tag{11}$$

For the DC model, the flat prior $\hat{p}(s_t)$ can be written as $Dir(\alpha_1, \ldots, \alpha_m)$ with $\alpha_m = 1$ for $m = 1, \ldots, M$. The naive observer posterior $\hat{p}(s_t|y_t)$ simply updates the flat prior based on only the most recent observation y_t. Hence, we have $\hat{p}(s_t|y_t) = Dir(\alpha'_1, \ldots, \alpha'_m)$ with $\hat{\alpha}_m = 1 + \mathbf{1}_{y_t=m}$. A detailed account of the readout definitions can be found in S1 Appendix.

For the HMM, the surprise readouts are obtained by iteratively computing the posterior distribution via the Baum-Welch algorithm using the *hmmlearn* Python package [57]. For timestep t this entails fitting the HMM for a stimulus sequence $o_1, \ldots, o_t$ which gives a set of parameter estimates, $\hat{\pi}_t, \hat{A}_t, \hat{B}_t$ and the filtering posterior $\hat{\gamma}_t(s_t) = p(s_t|o_1, \ldots, o_t)$. Predictive, Bayesian, and confidence-corrected surprise may then be expressed as follows (see S1 Appendix).

$$PS(o_{t+1}) \approx -\ln(\hat{B}_t^T \hat{A}_t^T \hat{\gamma}_t(s_t)) \tag{12}$$

$$BS(o_{t+1}) \approx \sum_k^K \hat{\gamma}_t(s_t = k) \ln \frac{\hat{\gamma}_t(s_t = k)}{\hat{\gamma}_{t+1}(s_{t+1} = k)} \tag{13}$$

Following Faraji et al. [35], confidence-corrected surprise may be expressed as a linear combination of predictive surprise, Bayesian surprise, a model commitment term (negative entropy) $C(p(s_t))$, and a data-dependent constant scaling the state space $O(t)$. Here we make use of this alternative expression of CS in order to facilitate the HMM implementation:

$$CS(o_t) = BS(o_t) + PS(o_t) + C(p(s_t)) + \ln O(t) \tag{14}$$

Fig 5 shows the regressors for an example sequence of the HMM TP_1 and DC TP_1 models with an observation half-life of 95. The PS regressors of both models show greater variability in the slow switching regime as compared to the fast-switching regime, where repetitions are more common (and consequently elicit less predictive surprise) while alterations are less common (and thus elicit greater surprise). As such, the PS regressors differ between regimes as a function of the estimated transition probabilities. The speed at which models adapt to the changed statistics depends on the forgetting parameter for the DC model while for the HMM it is

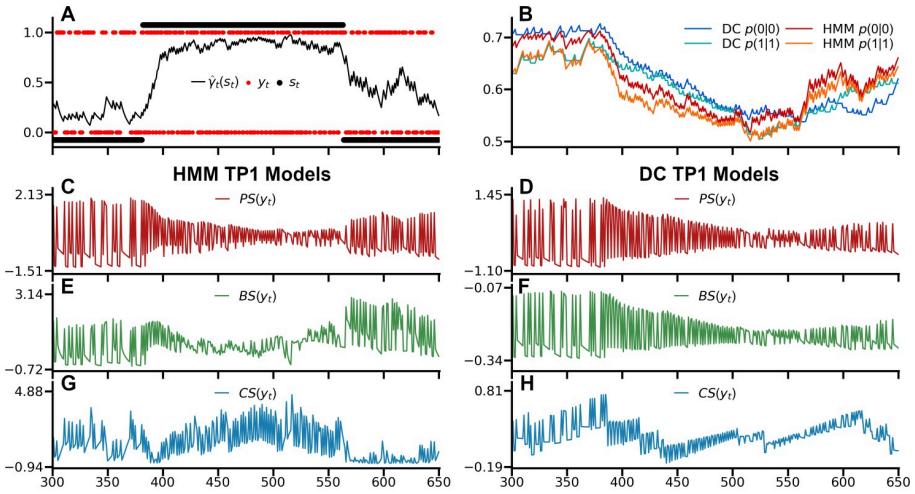


Fig 5. Surprise readouts. A) Example sequence with o_t in red, s_t in black with $s_t = 0$ for the slow-switching regime and $s_t = 1$ for the fast switching regime, and the HMM filtering posterior $\hat{\gamma}_t(s_t)$ in between. The rare catch-trials are not plotted to facilitate a direct comparison between the HMM and DC models. B) The normalized probability estimates of the HMM TP_1 and DC TP_1 model with an observation half-life of 95, displaying differences in estimates arising from different adaptations to regime switches. C,E,G) The z-scored surprise readouts of the HMM TP_1 models: predictive surprise (PS), Bayesian surprise (BS), and confidence-corrected surprise (CS). D,F,H) The z-scored surprise readouts of the DC TP_1 models.

https://doi.org/10.1371/journal.pcbi.1008068.g005

dependent on the degree to which the regimes have been learned. BS is markedly distinct for the two models due to the differently modeled hidden state. DC BS features many small updates during the fast-switching regime, with more irregular, larger updates during the slow-switching regime, while HMM BS expresses the degree to which an observation produces changes in the latent state posterior. Finally, HMM CS is scaled by the confidence in the latent state posterior, tending to greater surprise the more committed the model is to one particular latent state, and lower surprise otherwise, such as at the end of the example sequence. Meanwhile, due to its static latent state, confidence for DC CS results only from commitment to beliefs about the estimated transition probabilities between observations themselves, with rare events causing drops in confidence. Taken together, the HMM regressors ultimately depend on its posterior over latent states, and while this is absent for the DC, its regressors display differences between the two regimes as a function of its integration timescale which in turn allows it to accommodate its probability estimates to the currently active regime.

In an exploratory analysis, the trial-definitions of the GLM analysis of the individual electrode-time point data were applied to the surprise readout regressors. This allowed for the derivation of model-based predictions for the observed beta-weight dynamics of the ERP GLM. First, we generated an additional 25000 sequences of 800 observations using the same generative model used for the subject-specific sequences. The averaged surprise readouts of these simulated sequences yielded model-derived predictions, which allowed for a visual verification of the presence of these predictions in the (200) experimental sequences. As each study subject was exposed to 5 sequences, these sequences were grouped into sets of 5 (yielding 5000 simulated subjects) to mirror the EEG analysis. Besides the HMM, we used the Dirichlet-Categorical models with different values for the forgetting-parameter ('no forgetting', long, medium-length and very short stimulus half-lives) (S2 Fig). To reduce the model-space, only TP_1 models were used for this analysis.

Model fitting via free-form variational inference algorithm. Each combination of model class (DC and HMM), inference type (SP, AP, TP_1, TP_2), and surprise readout function (PS, BS, CS) yields a stimulus sequence-specific regressor. The same models were used across subjects and as such the regressors did not include any subject specific parameters. These regressors, as well as those of a constant null-model, were fitted to the single-trial, event-related electrode and source activation data. Using a free-form variational inference algorithm for multiple linear regression [58, 59, 60], we obtained the model evidences allowing for Bayesian model selection procedures [61], which accounts for the accuracy-complexity trade-off in a formal and well-established manner [62]. In short, the single-subject, single peri-stimulus time bin data $y \in \mathbb{R}^{n\times 1}$ for $n \in N$ trials was modeled in the following form:

$$p(y, \beta, \lambda) = p(y|\beta, \lambda)p(\beta)p(\lambda) \tag{15}$$

with $\beta \in \mathbb{R}^p$ and $\lambda > 0$ denoting regression weights and observation noise precisions, respectively. The parameter-conditional distribution of y, $p(y|\beta, \lambda)$, is specified in terms of a multivariate Gaussian density with expectation parameter $X\beta$ and spherical covariance matrix. The design matrix X consisted of a constant offset (null-model: $X \in \mathbb{R}^{n\times 1}$) and an additional surprise-model specific regressor in case of the non-null models ($X \in \mathbb{R}^{n\times 2}$). Both a detailed description of the algorithm and the test procedure performed on simulated data used to select the prior parameters for the variational distributions of β and λ may be found in the supplementary material S2 Appendix.

Bayesian model selection. Before modeling single subject, single peri-stimulus time bin data (y) as described above, the single-trial regressors of all non-null models as well as the data underwent z-score normalization to allow for the use of the same model estimation procedure for both sensor and source data. For single subjects, data and regressors corresponding to the five experimental runs were concatenated prior to fitting. To allow for the possibility that the brain estimates statistics computed across multiple timescales of integration [9, 63, 64], the forgetting-parameter τ of the DC model was optimized for each subject, model, and peri-stimulus time-bin. To this end, DC model regressors were fitted for a logarithmically spaced vector of 101 τ-values on the interval of 0 to 1 and the value of τ that resulted in the highest model evidence was chosen. To penalize the DC model for having one of its parameters optimized, the degree to which τ optimization on average inflated model evidences was subtracted prior to the BMS procedure. Specifically, the difference in model evidence between its average for all parameter-values and the optimized value was computed and subsequently averaged across post-stimulus timebins, sensors, and subjects. It should be noted that the applied procedure constitutes a heuristic for the penalization of model complexity while no explicit parameter fitting procedure was implemented within model estimation.

The furnished model evidences were subsequently used for a random-effects analysis as implemented in SPM12 [61] to determine the models' relative performance in explaining the EEG data. In order to combat the phenomenon of model-dilution [65], a hierarchical approach to family model comparison was applied (for a graphical overview see S3 Fig). This amounts to a step-wise procedure that leads to data-reduction at subsequent levels. Note that this procedure is performed for each peri-stimulus time bin and electrode independently (resulting in 22976 model comparisons per subject). In a first step, the two model classes DC and HMM were compared against each other and the null-model in a family-wise BMS. A threshold of exceedance probabilities $\varphi > 0.99$ in favour of either the DC or HMM was applied, so that only whenever there was strong evidence in favour of one of the model classes over both the alternative and the null-model the following analyses were applied. As the current analyses are not statistical tests per se, the thresholding of the data by certain exceedance

probabilities ultimately constituted an arbitrary choice to reduce data in order to visualize (and draw conclusions on) effects with certain minimum probabilities within a large model space. For timepoints with exceedance probabilities above this threshold, a family-wise comparison of TP_1 and TP_2 was performed in order to determine which order of transition probabilities would be used for the second level. Subsequently, either the TP_1 or TP_2 models were compared to the SP and AP models. Wherever $\varphi > 0.95$ for one of the inference type families, the third analysis level was called upon. On this final level, surprise read-out functions were compared for the winning model class and corresponding inference type. The direct comparison of read-out models within the winning family allows for the use of protected exceedance probabilities (which are currently not available for family comparisons), which provide a robust alternative to inflated exceedance probabilities [66]. The step-wise procedure allows for spatio-temporal inference on particular read-out functions for which there is evidence for a belonging model class and inference type, facilitating the interpretation of the results. The hierarchical ordering thus moves from general to specific principles: the model class and inference type determine the probability estimates of the model, which are finally read out through surprise computation. While this procedure provides a plausible and interpretable approach to our model comparison, it should be noted that it constitutes an arbitrary choice in order to reduce data and model space and must be interpreted with caution. As a supplementary analysis, we performed non-hierarchical (factorial) family comparison analysis (S4 Fig) which groups the entire model space into the respective families for each family comparison without step-wise data reduction. The same procedure was used for the EEG sensor and source data.

To inspect the values of the forgetting-parameter τ that best fit the dipole data, subject specific free energy values were averaged across the timebins with surprise readout effects of interest for the corresponding dipoles. These were summed across subjects to yield the group log model evidence for each tested value of τ, which were subsequently compared against each other.

Model recovery study. A simulation model recovery study was performed to investigate the ability to recover the models given the sequence data, model fitting procedure, and model comparison scheme. To this end, data was generated for n = 4000 (corresponding to the five concatenated experimental runs) by sampling from a GLM $y \sim N(X\beta, \sigma^2 I_n)$, after which model selection was performed. For the null-model, the design-matrix only comprised a column of ones. For all non-null models, an additional column of the z-normalized regressor was added. We set the true, but unknown β_2 parameter to 1, while varying σ^2, which function as the signal and noise of the data respectively. Given the z-scoring of the data, the β_1 parameter responsible for the offset is largely inconsequential and thus not further discussed. The model fitting procedure was identical to the procedure described in the supplementary material used for the EEG analyses (S2 Appendix).

For each noise level, we generated 40 data sets (corresponding to the number of subjects) to apply our random-effects analyses. This process was repeated 100 times for each of the different comparisons: null model vs DC model vs HMM (C1), DC TP_1 vs TP_2 (C2), DC SP vs AP vs TP_1 (C3), and DC TP_1 PS vs BS vs CS (C4). Family and model retrieval using exceedance probabilities worked well across all levels (S5 Fig), with a bias to the null model as signal-to-noise decreases. By inspecting the posterior expected values of β_2 and λ^{-1} which resulted from fitting the model regressors to the EEG data, an estimate of the signal-to-noise ratio that is representative of the experimental work can be obtained. By applying the thresholds of $\varphi > 0.99$, $\varphi > 0.95$, $\varphi > 0.95$, and $\tilde{\varphi} > 0.95$ across the four comparisons respectively and subsequently inspecting the winning families and models at $\sigma^2 = 750$ (i.e., an SNR of 1/750), no false positives were observed. For C1 and C4, recovery was successful for all true, but unknown models in all of the 100 instances. While for C2 and to a lesser extent C3, concerning the families of

estimated sequence statistics, false negatives were observed only when confidence-corrected surprise was used to generate data. For C2, this led to false negatives in 67 (TP_1 CS) and 55 (TP_2 CS) percent of cases, while for C3 28 (SP CS), 0 (AP CS), and 33 (TP_1 CS) percent false negatives were observed. Each set of 40 data sets was generated with the same true, but unknown model. Due to the limited cognitive flexibility afforded by the distractor task, we did not expect large variability in the models used across subjects. Nevertheless, if this assumption is incorrect these simulations potentially overestimate the recoverability of the different models.

Results

Behavioural results and event-related potentials

Participants showed consistent performance in counting the amount of catch trials during each experimental run, indicating their ability to maintain their attention on the stimuli (robust linear regression of presented with reported targets: slope = 0.96, $p < 0.001$, $R^2 = 0.93$). Upon questioning during the debriefing, no subjects reported explicit awareness of switching regimes during the experiment.

An initial analysis was performed to confirm our paradigm elicited the typical somatosensory responses. Fig 6B shows the average SEP waveforms for contralateral (C4, C6, CP4, CP6) somatosensory electrodes with the expected evoked potentials, i.e. N20, P50, N140 and P300 resulting from stimulation of the left wrist. The corresponding topographic maps (Fig 6C) confirm the right lateralized voltage distribution of the somatosensory EEG components on the scalp. The EEG responses to stimulus mismatch were identified by subtracting the deviant from the standard trials (deviants-standards), thereby obtaining a difference wave for each electrode (see Fig 6D). The scalp topography of the peak differences between standards and deviants within predefined windows of interest indicates mismatch responses over somatosensory electrodes (Fig 6E).

To test for statistical differences in the EEG signatures of mismatch processing we contrasted standard and deviant trials with the general linear model. Three main clusters reached significance after performing family-wise error correction for multiple comparisons. The topographies of resulting F-values are depicted in Fig 7. The earliest significant difference between standard and deviant trials can be observed around 60ms post-stimulus (peak at 57ms, closest electrode CP4, $p_{FWE} = 0.002$, F = 27.21, Z = 5.07), followed by a stronger effect of the hypothesized N140 component around 120ms which will be referred to as the N140 mismatch response (N140 MMR, peak at 119ms, closest electrode: FC4, $p_{FWE} = 0.001$, F = 29.56, Z = 5.29). A third time window of a very strong and elongated difference effect starting around 250ms to 400ms post-stimulus which corresponds to the hypothesized P300 MMR (peak at 361ms, closest electrode: Cz, $p_{FWE} < 0.001$, F = 72.25).

The inspection of the β-parameter estimates at the reported GLM cluster peaks (illustrated in Fig 7) indicates that stimulus train length, i.e. the number of standard stimuli that precede a deviant stimulus, has differentiable effects on the size of EEG responses to standard and deviant stimuli. Both the N140 and P300 MMR effects are found to be parametrically modulated by train length as indicated by a significant linear relationship between β-estimates and train length. Specifically, the N140 MMR effect is reciprocally modulated by stimulus type, such that responses to standards are more positive for higher train lengths (F-statistic vs. constant model: 5.45, $p = 0.021$) while deviant responses become more negative (F-statistic vs. constant model: 5.07, $p = 0.026$). The parametric effect on the P300 MMR is entirely driven by the effect on deviant stimuli (F-statistic vs. constant model: 20.7, $p < 0.001$), with no effect of train

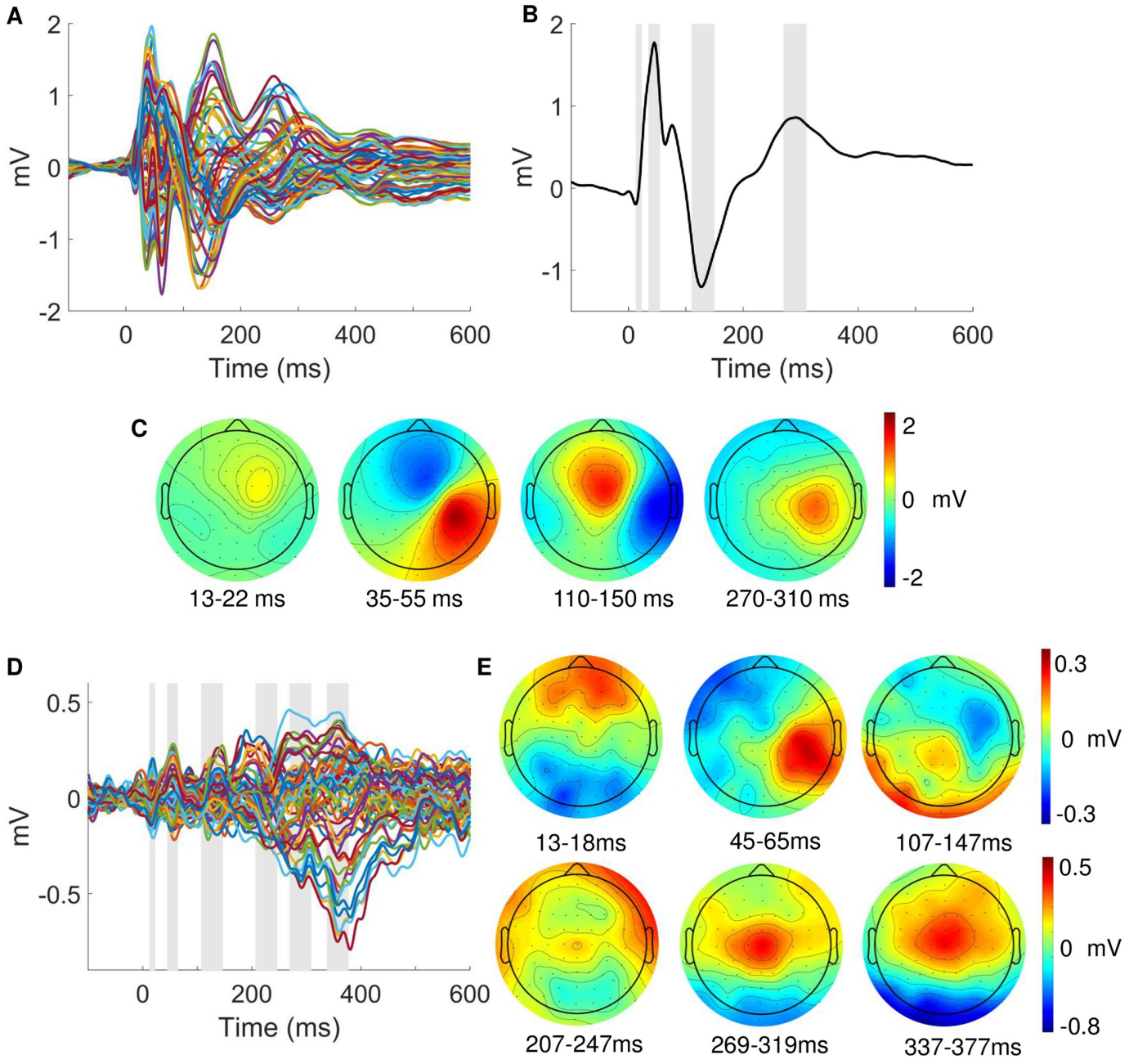


Fig 6. Event-related potentials. (A) Grand average SEP of all 64 electrodes. (B) Average SEP across electrodes C4, C6, CP4, CP6 (contralateral to stimulation). Grey bars indicate time windows around the standard somatosensory ERP components (13-23ms; 35-55ms; 110-150ms; 270-310ms). (C) ERP scalp topographies corresponding to the time windows in B. (D) Grand average ERP of the mismatch response obtained by subtraction of standard from deviant trials of 64 electrodes. Grey bars indicate windows around peaks which were identified within pre-specified time windows of interest around somatosensory ERP or expected mismatch response components (13-18ms; 45-65ms; 107-147ms; 207-247ms 269-319ms; 337-377ms). (E) ERP scalp topographies corresponding to the time windows in D).

https://doi.org/10.1371/journal.pcbi.1008068.g006

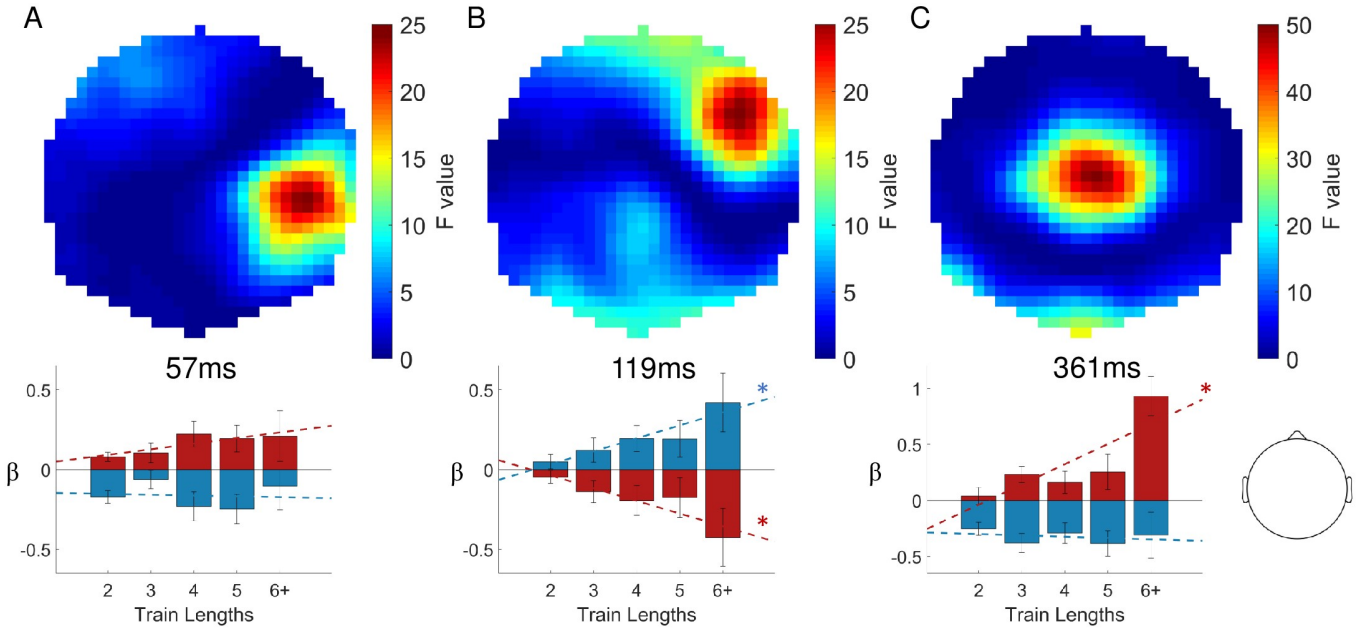


Fig 7. Statistical parametric maps of mismatch responses. Top row: Topographical F maps resulting from contrasting standard and deviant conditions averaged across the times of significant clusters: 57ms (A), 119ms (B) and 361ms (C). Bottom row: Corresponding beta parameter estimates of the significant peaks with deviants in red and standards in blue. Asterisks indicate significant linear fits ($p < 0.05$). Head depiction on the bottom right shows the orientation of the topographic maps.

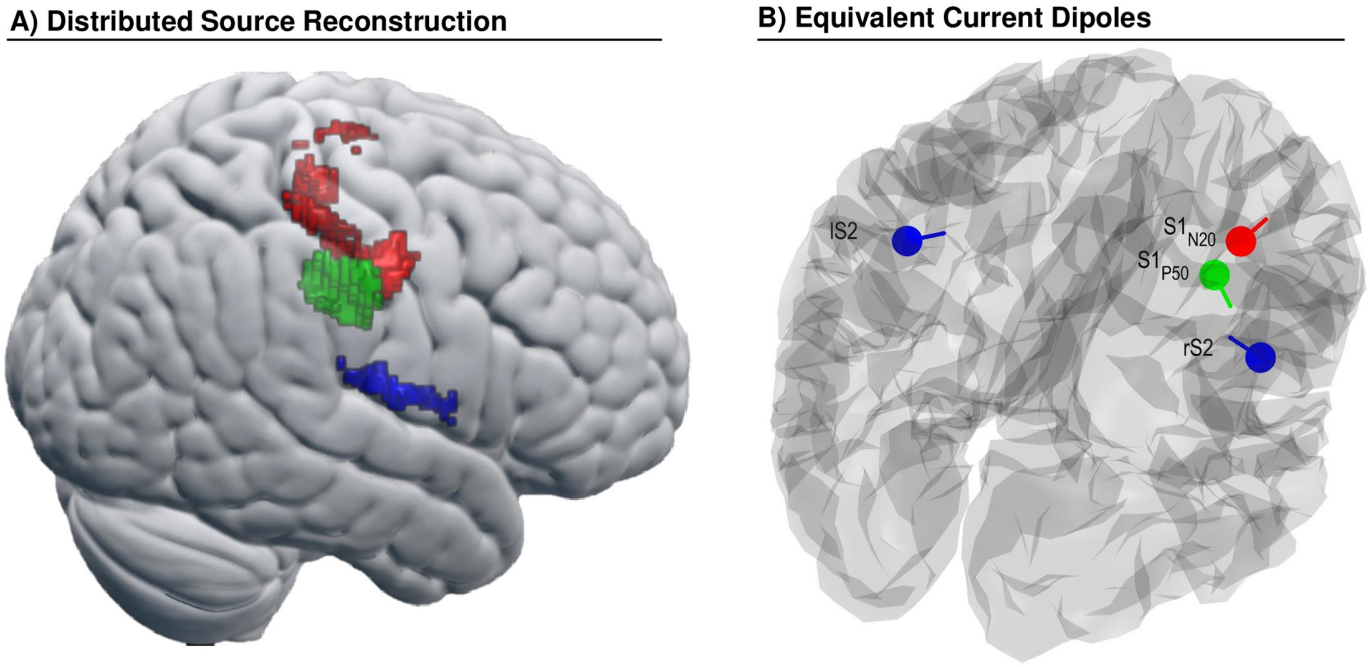
https://doi.org/10.1371/journal.pcbi.1008068.g007

length on the response to standard stimuli ($p > 0.05$). For the early 60ms cluster no effect was found on either standard or deviant stimuli.

Source reconstruction

The distributed source reconstruction resulted in significant clusters at the locations of primary and secondary somatosensory cortex (Fig 8A, with details specified in the corresponding table). The resulting anatomical locations were subsequently used as priors to fit four equivalent current dipoles (Fig 8B, with details specified in corresponding table). Two dipoles were used to model S1 activity at time points around the N20 and the P50 components while an additional symmetric pair captured bilateral S2 activity around the N140 component. The moment posteriors of the S2 dipoles end up not strictly symmetric due to the soft symmetry constraints used by the SPM procedure [67].

To establish the plausibility of the somatosensory dipole model the EEG data was projected onto the four ECD's and the grand average source ERP was computed across subjects for standard and deviant trials. The resulting waveforms, shown in Fig 9, show a neurobiologically plausible spatiotemporal evolution: the two S1 dipoles reflect the early activity of the respective N20 and P50 components while the S2 dipoles become subsequently active and show strongest activity in right (i.e. contralateral) S2. The average response to standards and deviants within time windows around the significant MMR's in sensor space (around 57ms and 119ms; see Fig 7) were compared with simple paired t-tests. The $S1_{P50}$ dipole shows a significant difference at both time windows (at 57ms $p = 0.006$, $t = 2.94$; at 119ms $p = 0.009$, $t = 2.75$; bonferroni corrected) and can be suspected to be the origin of the effect at 57ms as well as contribute to



Label	Time (ms)	p-Peak (FWE)	F value	MNI	Cytoarchitect. Reference
S1 (N20)	18-25	0.019	30.23	44 -16 48	Area 3b: 30%
S1 (P50)	35-45	<0.001	76.24	42 -20 38	Area 3b: 37% Area 3a: 19%
S2 (right)	110-160	0.019	29.85	62 -12 14	Area OP4: 36% Area OP1: 23%

Label	Time (ms)	Location Prior (MNI)	Location Posterior (MNI)	Moment Posterior
S1 (N20)	20	44 -16 48	41 -12 53	2.92 1.67 3.89
S1 (P50)	45	42 -20 38	36 -17 44	17.31 -14.25 -6.49
rS2	130	62 -12 14	46 -12 16	-28.31 5.86 4.24
lS2	130	-62 -12 14	-46 -12 16	12.69 5.86 4.24

Fig 8. EEG source model. (A) Statistical results of distributed source reconstruction. Red: 18-25ms, Green: 35-45ms, Blue: 110-160ms. Below: Table with corresponding detailed data of the clusters. (B) Location and orientation of fitted equivalent current dipoles. Red: S1 (N20), Green: S1 (P50), Blue: bilateral S2. Below: Table with their corresponding values.

https://doi.org/10.1371/journal.pcbi.1008068.g008

the 119ms MMR while the right S2 dipole is mainly driving the strong 119ms effect ($p = 0.001$, $t = 3.44$; bonferroni corrected).

Single trial modeling

We previously established the presence of mismatch responses in sensor space and confirmed their origin in the somatosensory system by modeling the early EEG components in source space. Subsequently, we investigated the temporal and spatial surprise signatures with trial-by-trial modeling of electrode and source data.

Modeling in sensor space. For large time windows at almost all electrodes there is strong evidence in favor of the DC model class ($\varphi > 0.99$), while the HMM model class does not exceed thresholding anywhere, therefore excluding HMM models from further analyses (Fig 10A). The corresponding threshold of expected posterior probabilities to arrive at comparable results lies around $\langle r \rangle > 0.75$ (see S6 Fig). To verify that this result was not merely due to an insufficient penalization of the DC models, the analysis was repeated with $\tau = 0$. Thus,

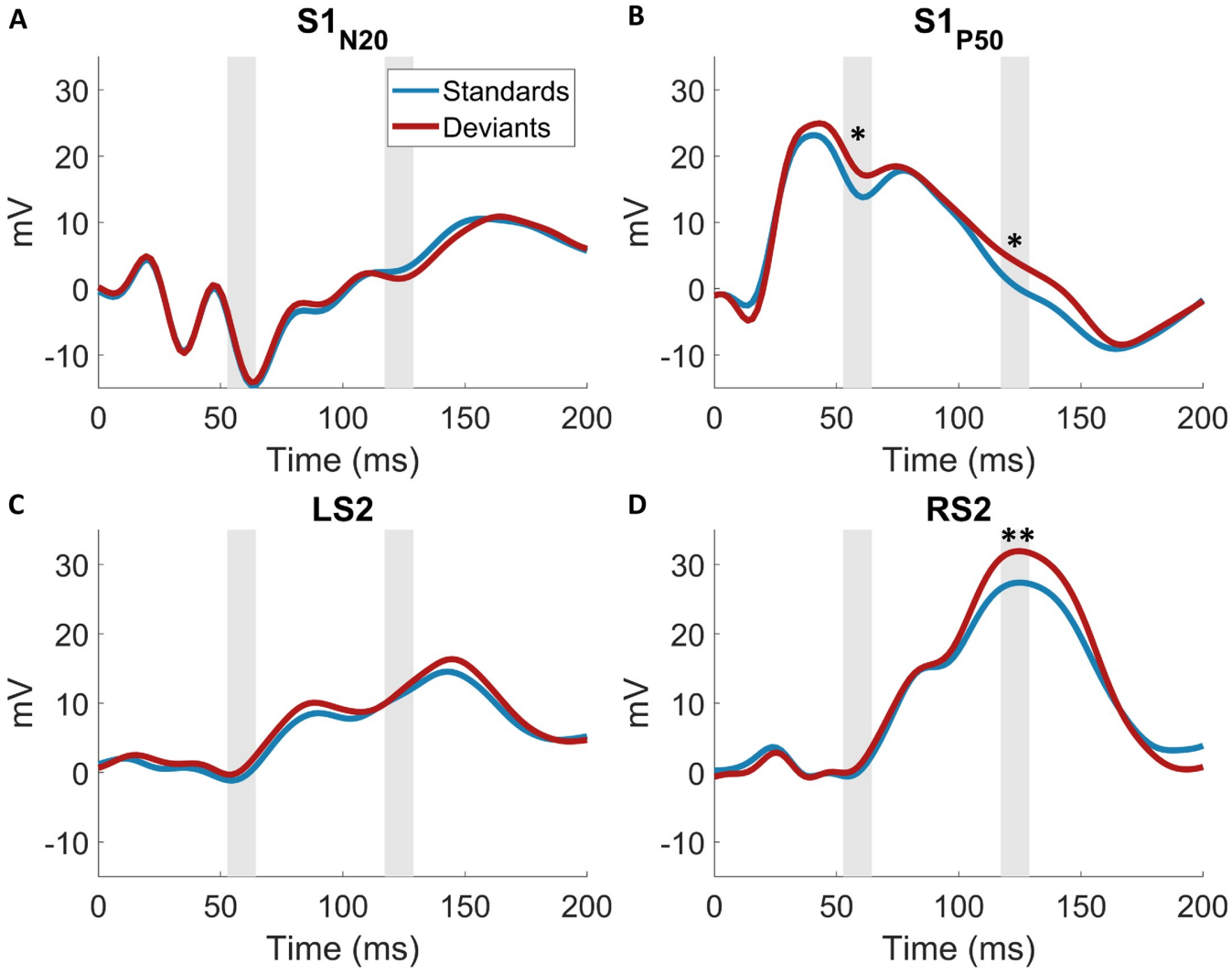


Fig 9. Grand average waveforms of EEG dipole projections. Standards and deviants were contrasted within time windows of interest informed by the GLM in the results section. $^{*}p < 0.05$; $^{**}p < 0.01$; Bonferroni corrected.

https://doi.org/10.1371/journal.pcbi.1008068.g009

under this setting, all instances of the DC model had perfect, global integration similar to the HMM models. Likewise, no results above the threshold were found for the HMM model class (S7 Fig). Next, to ensure that the superiority of the DC model did not solely result from the additionally modeled catch trials, the HMM was compared with a DC model which did not capture these trials. This DC model still consistently outperformed the HMM, though it should be noted that the evidence for such a reduced DC model over the HMM is less pronounced (S6B Fig). For the DC model, TP_1 is found to outperform TP_2 ($\varphi > 0.95$, roughly corresponding to $\langle r \rangle > 0.7$), excluding TP_2 for the second and third level analyses. In the following step, TP_1 clearly performed better than SP and AP at almost all electrodes and time points (see Fig 10B and 10C; $\varphi > 0.95$, roughly corresponding to $\langle r \rangle > 0.7$). Thus, the following section presents the random-effects Bayesian model selection results of the readout functions of the Dirichlet-Categorical TP_1 model (shown in Fig 10D).

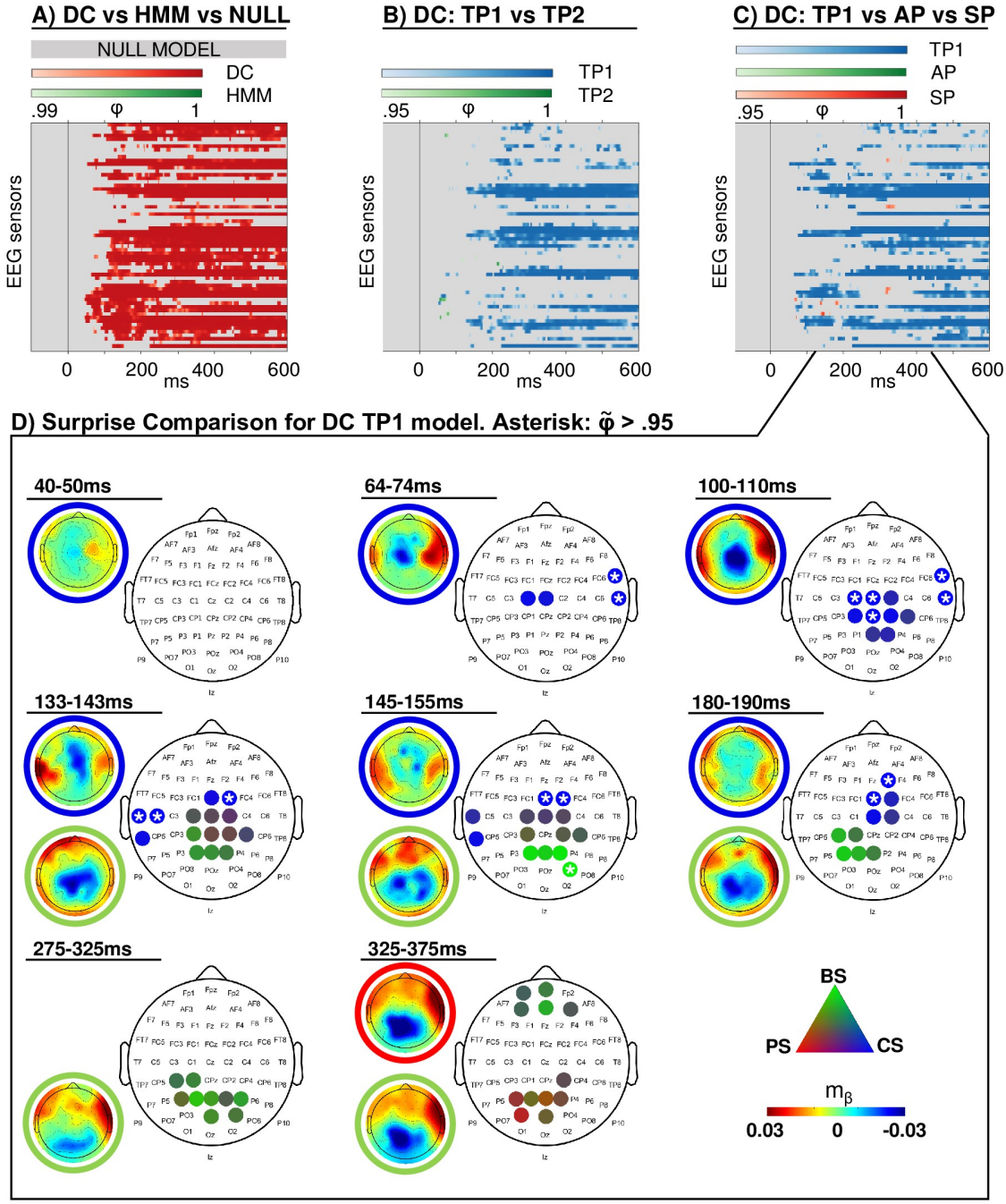


Fig 10. Modeling results. Exceedance probabilities (φ) resulting from the random-effects family-wise model comparison. (A) Dirichlet-Categorical (DC) model, Hidden Markov Model (HMM) and null model family comparison, thresholded at $\varphi > 0.99$ and applied for data reduction at all further levels. (B) Family comparison within the winning DC family, thresholded at $\varphi > 0.95$: first and second order transition probability models (TP_1, TP_2). (C) Family comparison within the winning DC family, thresholded at $\varphi > 0.95$: first order transition probability (TP_1), alternation probability (AP) and stimulus probability (SP) models and applied at the final level. (D) Unthresholded protected exceedance probabilities ($\tilde{\varphi}$) resulting from model comparison of surprise models within the winning DC TP1 family: Large discrete topographies show the electrode clusters of predictive surprise (PS) in red, Bayesian surprise (BS) in green and confidence-corrected surprise (CS) in blue. White asterisks indicate $\tilde{\varphi} > 0.95$ of single electrodes. Small continuous topographies display the converged variational expectation parameter m_β. This parameter may be interpreted as a β weight in regression, indicating the strength and directionality of the weight on the model regressor that maximizes the regressor's fit to the EEG data (see S2 Appendix).

https://doi.org/10.1371/journal.pcbi.1008068.g010

The scalp topographies depict the winning readout functions of the DC TP_1 model at different time windows. Given the difference in temporal dynamics of faster, early (<200 ms) and slower, late (e.g. P300) EEG components, different time windows were applied for averaging. Early clusters were identified by averaging protected exceedance probabilities over 10ms windows and using a minimum cluster size of two electrodes, while 50ms time windows were applied for averaging across later time windows with a minimum cluster size of four. The resulting clusters indicate that from around 70ms on, early surprise effects represented by confidence corrected surprise (CS) best explain the EEG data on contralateral and subsequently ipsilateral electrodes up to around 200ms. As demarcated in the plot, the early CS clusters include electrodes with $\tilde{\varphi} > 0.95$, which is indicative of a strong effect. A weaker cluster of Bayesian surprise (BS) is apparent at centro-posterior electrodes between 140-200ms of which the peak electrodes around 150ms show $\tilde{\varphi}$ between 0.8 and 0.95. As such, the mid-latency BS effect is less strong than the earlier CS clusters and can only provide indications. At the time windows of the P300 around 300 and 350ms, similar centro-posterior electrodes represent weak Bayesian surprise (peak $\tilde{\varphi}$ around 0.75) and predictive surprise (PS) clusters (peak $\tilde{\varphi}$ around 0.72), respectively. The mid-latency BS cluster is temporally in accordance with the putative N140 MMR while the late two clusters of BS and PS might be interpreted as indicative of a P300 MMR. However, especially the weak late clusters do not provide clear evidence in favour of a specific surprise readout function.

We note that the DC TP_1 vs TP_2 comparison in Fig 10B has few results prior to 200ms. This appears to fit with the model recovery study indicating that the least recoverable families are DC TP_1 and TP_2 in case of CS and the observation that CS is a winning surprise model for early time bins. In response, we conducted an additional family comparison between SP, AP, and TP encompassing both TP_1 and TP_2 (see S7 Fig). Clearly, more electrodes and time points with $\varphi > 0.95$ can be observed in the early time window, suggesting that early effects are driven by TP inference but that for empirical data, we are unable to convincingly resolve TP_1 and TP_2 for CS computation. Furthermore, it should be noted that our step-wise model comparison approach constitutes a reasonable, yet arbitrary choice to create summary statistics of our data set and a large model space. In an additional analysis, we performed a non-hierarchical model comparison which grouped the entire model space in the respective families of interest without step-wise data reduction. These results (S4 Fig) broadly replicate the findings from the hierarchical approach across the levels and likewise indicate that the order of transition probability (TP1 and TP2) can not be resolved in early time windows.

Modeling in source space. The topographic distribution of the effects of confidence-corrected surprise seem to indicate an early contribution of secondary somatosensory cortex from around 70ms on that starts contra- and extends ipsilaterally while the weaker BS cluster emerges around the time of the N140 MMR. In order to further investigate this observation and examine the spatial origins of the surprise clusters, we fit our models to the single trial dipole data and used the same hierarchical Bayesian model selection approach as for the sensor-space analysis described in the Materials and Methods section. Results for the source activity were highly similar, with clear results in favour of the DC and TP_1 model families at thresholds of $\varphi > 0.99$ and $\varphi > 0.95$, respectively. Consequently, the surprise readout functions of the DC TP_1 model were subjected to BMS. The results depicted in Fig 11 support the interpretation of an early onset of CS in secondary somatosensory cortex ($\tilde{\varphi} > 0.95$) and allocate the later onset BS cluster in electrode space to primary somatosensory cortex ($\tilde{\varphi}$ ranging from around 0.7 to 0.9). However, as is also apparent in electrode space, this mid-latency BS effect is weak and can only provide an indication.

Leaky integration. We inspected the τ-parameter values that resulted in the highest group log model evidence for the reported dipole effects (Fig 11). All three considered clusters

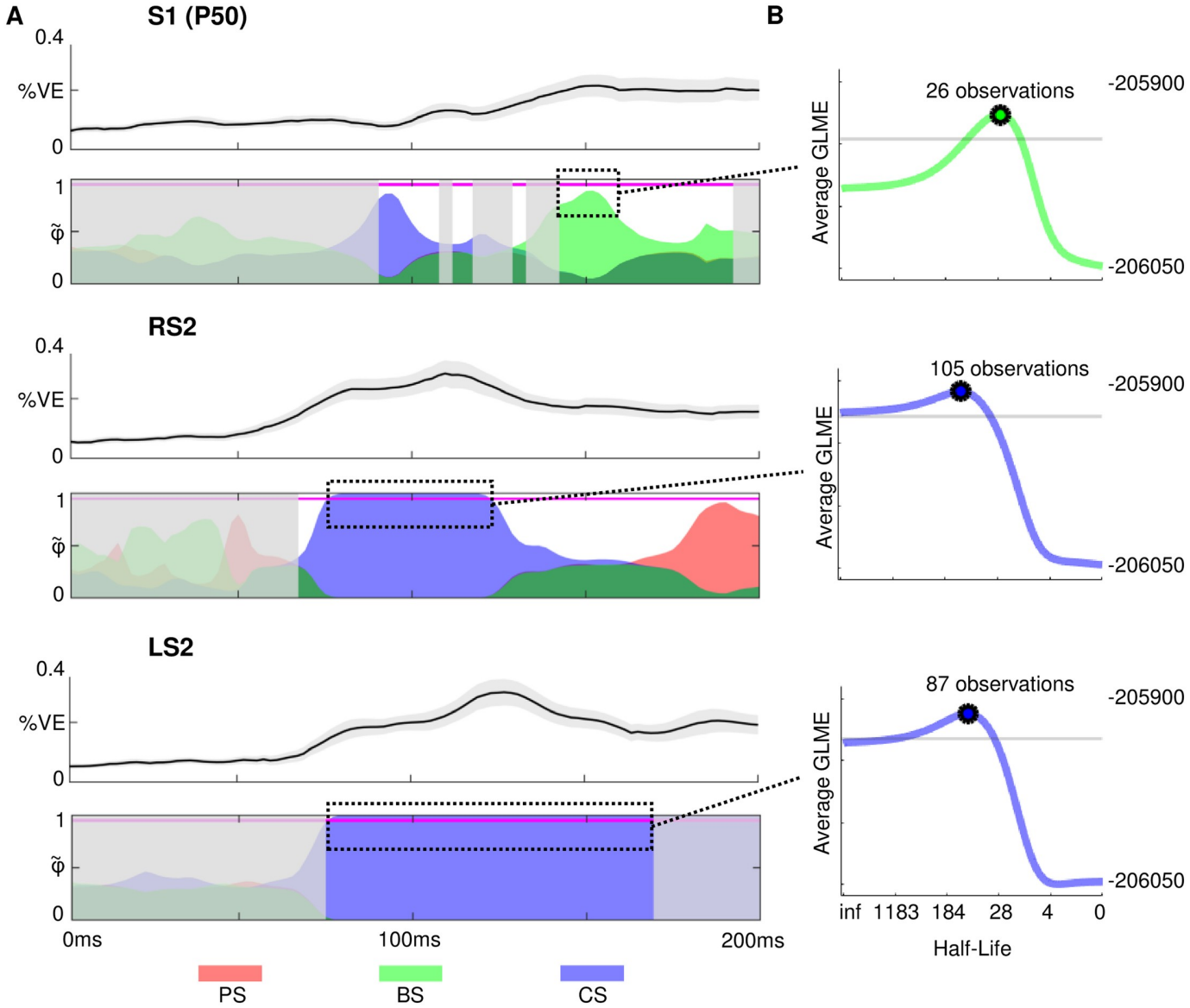


Fig 11. Modeling results in source space with best fitting forgetting-parameter values. Red: Predictive surprise (PS), Green: Bayesian surprise (BS), Blue: Confidence-corrected surprise (CS) A) Colored areas depict protected exceedance probabilities ($\tilde{\varphi}$) of the surprise readout functions of the Dirichlet-Categorical TP_1 model within the dipoles S1P50, right S2 (RS2) and left S2 (LS2) using alpha blending. In grey shaded areas the DC model family shows $\varphi < 0.99$ or the TP_1 model family $\varphi < 0.95$. The S1N20 dipole was omitted in the visualization as no model is observed above this threshold. Magenta horizontal lines indicate $\tilde{\varphi} = 0.95$. Line plots above each dipole plot show the respective mean percent variance explained of the models in dotted rectangles ± standard error. B) The group log model evidence (GLME) values corresponding to the stimulus half-lives for forgetting-parameter τ, after averaging the timebins inside the dotted-rectangles (S1P50: 145-191ms; RS2: 68-143ms; LS2: 76-168ms). The grey lines indicate a difference of 20 GLME from the peak, indicating very strong evidence in favour of the peak half-life value compared to values below this threshold.

https://doi.org/10.1371/journal.pcbi.1008068.g011

indicate a local timescale of integration, with the best-fitting parameter values resulting in a stimulus half-life of $\sim$ 105 and $\sim$ 87 for the confidence-corrected surprise effects at 75-120ms and 75-166ms respectively, and a half-life of $\sim$ 26 observations for the Bayesian surprise effect at 143-157ms. Using the single-subject peaks, τ was found to significantly differ from 0 (i.e., no

forgetting) for the BS effect in S1 ($p < 0.001$) and CS in RS2 ($p < 0.05$), but not in LS2 (p = 0.06). Paired t-tests revealed no significant difference in τ underlying the three effects ($p > 0.05$).

Discussion

In this study, we used a roving paradigm to identify EEG mismatch responses independent of stimulus identity. The early MMR effects were source localized to the somatosensory system and the N140 and P300 MMR's show differential linear dependence on stimulus train lengths for standard and deviant stimuli. Using computational modelling, EEG signals were best described using a non-hierarchical Bayesian learner performing transition probability inference. Furthermore, we provide evidence for an early representation of confidence-corrected surprise localized to bilateral S2 and weak indications for subsequent Bayesian surprise encoding in S1. These computations were shown to use a local, rather than global, timescale of integration.

We report a significant somatosensory mismatch response around three distinct post-stimulus time-points: 57ms, 119ms, and 361ms. These will be referred to as sMMR's as opposed to MMN since the effects at 57ms and 361ms are not negativities and our experimental protocol included an explicit attentional focus on the stimulation. The MMN was originally defined to be a pre-attentive effect and while attention to the stimulus does not seem to influence the MMN in the visual domain [68], we don't address a potential independence of attention here. Nevertheless, the reported sMMR effects integrate well with previous findings on the somatosensory MMN (sMMN). Our 119ms effect is in line with the timing of the most commonly reported sMMN as a modulation of the N140 component between 100-250ms [17, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31]. However, some studies additionally describe a modulation of multiple somatosensory components [17 18, 19, 24], similar to our three distinct sMMR effects. The electrode positions reported in sMMN studies show a large variability of frontocentral and parietal electrodes. These discrepancies might be driven by the differences in stimulation sites (different fingers and hand) and deviant definitions (vibrotactile frequencies, stimulation site, stimulation duration). Here, we present significant effects around C4 and FC4 electrodes for the 57 and 119ms time-points, respectively, indicating EEG generators within the contralateral somatosensory system. This implication is in line with intracranial EEG recordings of the somatosensory cortex during oddball tasks [24, 30]. In accordance with previous MEG studies using source localization [21, 22], our source space analysis suggests the early MMR effects to originate from contralateral primary and secondary somatosensory cortex (cS1 and cS2, respectively), with the earliest MMR (at 57ms) localized to cS1 followed by a combined response of S1 and S2. While evidence exists for a role of S2 in the early phase of mismatch processing [26], the evolution from an initial MMR generated by S1 to an additional involvement of S2 in the mid-latency MMR, as indicated by our findings, is consistent with the sequential activation of the somatosensory hierarchy in general tactile stimulus processing [69, 70, 71]. Finally, the third sMMR effect at 361ms is in accordance with a large body of evidence showing a modulation of the P300 component by mismatch processing [72, 73, 74]. The P300 in response to oddball tasks likely reflects a modality unspecific effect, dependent on task-related resource allocation [75, 76, 77, 78, 79] and contingent on attentional engagement [29].

In addition to three spatiotemporally distinct sMMR effects, we further show their differential modulation by the length of standard stimulus trains preceding the deviant stimulus. This finding supports the interpretation that distinct mechanisms underlie the generation of the different sMMR's. The earliest effect around 57ms is not affected by train length, possibly

reflecting a basic change detection mechanism that signals a deviation from previous input regardless of statistical regularities. The mid-latency MMR around 119ms, on the other hand, shows a significant linear dependence on stimulus train length for both deviant and standard stimuli. Longer train lengths result in parametrically stronger negative responses to deviant stimuli while responses to standard stimuli are increasingly reduced. This effect is in accordance with repetition suppression effects reported for the MMN [80, 81] which have been shown to be dependent on sequence statistics and are interpreted to reflect perceptual learning [82, 83]. While it has been indicated that the number of preceding standards can also enhance the sMMN [26], no previous studies show comparable effects to our parametric modulation of the mid-latency sMMR. The reciprocal effect of repetition for standard and deviant stimuli shown here indicate early perceptual learning mechanics in the somatosensory system, likely originating from S2 in interaction with S1. In contrast, later mismatch processing reflected by the sMMR at 361ms only shows a linear dependence of deviant stimuli on train length, while the response to standards remains constant. This is in line with the interpretation that perceptual learning in the P300 reflects a recruitment of attention in response to environmental changes, possibly accompanied by updates to this attentional-control system [41].

In addition to average-based ERP analyses, single-trial brain potentials in response to sequential input can provide a unique window into the mechanisms underlying probabilistic inference in the brain. Here, we investigated the learning of statistical regularities using different Bayesian learner models with single-trial surprise regressors. Partitioning the model space allowed us to infer on distinguishing features between the model families using Bayesian model selection (BMS). The first comparison concerned the form of hidden state representation: In order for a learner to adequately adapt one's beliefs in the face of changes to environmental statistics, more recent observations may be favored over past ones without modeling hidden state dynamics (Dirichlet-Categorical model; DC), or different sets of statistics may be estimated for a discretized latent state (Hidden Markov Model; HMM). Our comparison of these two learning approaches provides strong evidence for the DC model class over the HMM for the large majority of electrodes and post-stimulus time. The superiority of the DC model was found to be irrespective of the inclusion of leaky integration to the DC model, indicating the advantage of a non-hierarchical model in explaining the EEG data. It is noteworthy that part of the strength of the DC model depended on the modelling of the catch trial, although a reduced DC model still outperformed the HMM. Participants were neither aware of the existence of the hidden states in the data generation process, nor was their dissociation or any tracking of sequence statistics required to perform the behavioural task. As such, the early EEG signals studied here are likely to reflect a form of non-conscious, implicit learning of environmental statistics [84, 85, 86]. However, it is possible that the brain implements different learning algorithms in different environments, resorting to more complex ones only when the situation demands it. As the discrete hidden states produced relatively similar observation sequences, more noticeable transitions between hidden states may provide an environment with greater incentive to implement a more complex model to track these states, which might have yielded different results. Indeed, humans seem to assume different generative models in different contexts, possibly depending on task instructions [87]. This may in part explain why evidence has been provided for the use of both hierarchical [88, 89] and non-hierarchical models [90, 91]. Nevertheless, it has been suggested that the brain displays a sensitivity to latent causes in low-level learning contexts [92], which might indicate the relevance of other factors. For example, it is possible the currently tested HMM may be too constrained and a simpler, more general changepoint-detection model [89] may have performed better. By omitting instructions to learn the task-irrelevant statistics, our study potentially avoids the issue of invoking a certain generative model. We might therefore report on a 'default' model of the

brain used to non-consciously infer environmental statistics. The sort of computations (relating to surprise and belief updating) and learning models we consider might be viewed in light of theories such as predictive coding and the free energy principle for which preliminary work suggests implementational plausibility (e.g. [93]). The computation models tested in the current study do not provide a biophysically plausible manner by which the brain acquires the estimated transition probabilities and subsequent surprise quantities. Rather, the models serve to identify qualities that a future successful biophysically plausible algorithm should exhibit.

In order to investigate which statistics are estimated by the brain during the learning of categorical sequential inputs, we compared three models within the DC model family that use different sequence properties to perform inference on future observations: stimulus probability (SP), alternation probability (AP), and transition probability (TP) inference. The TP model subsumes SP and AP models and is thus more general by maintaining a larger hypothesis space. Our results show that the TP model family clearly outperformed the SP and AP families, suggesting that the brain captures sequence dependencies by tracking transitions between types of observations for future inference. We thereby provide further evidence for an implementation of a minimal transition probability model in the brain as recently concluded from the analysis of several perceptual learning studies [94], extending it to include somesthesis. Additionally, we expand upon previous studies by comparing a first order TP model (TP_1), capturing transitions between stimuli conditional only on the previous observation, with a second order TP model (TP_2), which tracks transitions conditional on the past two observations. Our results suggest that the additional complexity of the second order dependencies contained in our stimulus sequence were not captured by the brain, although we were not able to convincingly show this for early CS computation. Nevertheless, the brain may resort to alternative, more compressed representations [95].

The BMS analyses of the partitioned model space suggests that the brain's processing of the stimulus sequences is best described by a Bayesian learner with a static hidden state (akin to the DC model) which estimates first-order transition probabilities (TP_1). Within the DC-TP_1 model family, we compared the surprise quantifications themselves as the readout functions for the estimated statistics of the Bayesian learner: predictive surprise (PS), Bayesian surprise (BS), and confidence-corrected surprise (CS). The results indicate that the first surprise effect is represented by CS from around 70ms over contralateral somatosensory electrodes which extends bilaterally and dissipates around 200ms. BS is found as a second, weaker centro-posterior electrode cluster of surprise between 140-180ms. As proposed by Faraji et al. [35], CS is a fast-computable measure of surprise in the sense that it may be computed before model updating occurs. In contrast, as BS describes the degree of belief updating, which requires the posterior belief distribution, it is expected to be represented only during the update step or later. As such, the temporal evolution of the observed CS and BS effects is in accordance with the computational implications of these surprise measures. Specifically, our study provides support for the hypothesis that the representation of CS, as a measure of puzzlement surprise, precedes model updating and may serve to control update rates. While PS is also a fast-computable puzzlement surprise measure and (similarly to CS) is scaled by the subjective probability of an observation, CS additionally depends on the confidence of the learner, read out as the (negative) entropy of the model. Evidence for a sensitivity to confidence of prior knowledge in humans has been reported in a variety of tasks and modalities [96, 97, 98]. This further speaks to the possibility that CS informs belief updating, as confidence has been suggested to modulate belief updating for other modalities in the literature [99, 100] and is explicitly captured in terms of belief precision by other promising Bayesian models [101, 102, 103]. We suspect that, similarly, confidence concerns the influence of new observations on current beliefs in somatosensation. However, as this was not explicitly modelled and investigated in

the current work we were not able to test it directly. Furthermore, as the state transition probability between regimes was fixed in the current study, it is not well suited to address the effects of the volatility of the environment on belief updating. Future work might focus on the interplay of environmental volatility and confidence in their effects on the integration of novel observations. It is important to note that one may also be confident about novel sensory evidence (e.g. due to low noise) which may result in larger model updates [104]. This aspect of confidence, however, lies outside the scope of the current work.

Our source reconstruction analyses attributed the early CS effects to the bilateral S2 dipoles, which is in accordance with the timing of S2 activation reported in the literature [69, 70, 71]. This finding suggests that the secondary somatosensory cortex may be involved in representing confidence about the internal model. The BS effect around 140ms was less pronounced in source space only peaking at $\tilde{\varphi}$ of 0.89 and was localized to S1. Despite the weak evidence for this BS representation around a 140ms somatosensory MMR, its timing matches prior work using modeling of Bayesian surprise signals in the somatosensory system [13]. Generally, our findings are in accordance with previous accounts of perceptual learning in the somatosensory system [105]. In sum, these results suggest that the secondary somatosensory cortex may represent confidence about the internal model and compute early surprise responses, potentially controlling the rate of model updating. Signatures of such belief updating, were found around the time of the N140 somatosensory response and were localized to S1. Together, these effects might be interpreted as a possible interaction between S1 and S2 that could be responsible for both a signaling of the inadequacy of the current beliefs and their subsequent updating.

In an attempt to relate the surprise readouts to the mismatch responses, we averaged surprise regressors to obtain model-based predictions for the standard-deviant contrasts. First, all TP_1 models except HMM CS predict the existence of an MMR, i.e., a difference in the averaged response between standard and deviant trials. Second, for multiple models an increase in train length leads to reduced surprise to standards and increased surprise to deviants. The CS readout is scaled by PS and BS, as well as by belief commitment, which increases for standards and decreases for deviants. This counteracting effect of belief commitment and the surprise terms can lead to independence of CS and train length when responses are averaged, manifesting in the current sequences only for standard trials. As the early MMR was found to be independent of train length, this indicates a possibility for a potential relation between these results. The intermediate MMR roughly temporally co-occurs with a simultaneous representation of BS and CS in S1 and S2. The dependence of the mid-latency MMR on train-length for both standards and deviants and the encoding of belief inadequacy and updating quantities is suggestive of convergent support in favor of a perceptual learning response which involves both somatosensory cortices. DC BS is however not the only model which predicts this dependence, highlighting the reduced ability to distinguish between models by averaging trials. At the P300 MMR it was found that only the response to deviants is dependent on train length. The averaged response of DC CS is most compatible with this ERP, however, this is unlikely to be meaningful as the model was not found to fit the single-trial EEG data well around this time. It is noteworthy that belief updating as described by DC BS, which is best describing the EEG data around that time, does not accurately predict the ERP dynamics of the P300, which matches the relative weakness of the BS effect in the single-trial EEG analysis. While a role of the P300 response in Bayesian updating has previously been reported [13, 40], the currently presented P300 dynamics may better be captured by alternate accounts, such as a reflection of an updating process of the attention allocating mechanism as suggested by Kopp and Lange [106].

Our implementation of the Dirichlet-Categorical model incorporates a memory-decay parameter τ that exponentially discounts observations in the past. The τ-values for the winning models of our BMS analyses that best fit the data for the surprise effects of interest

indicate relatively short integration windows for both CS and BS with stimulus half-lives of approximately 95 and 26 observations, respectively. This suggests that, within our experimental setup, the brain uses local sequence information to infer upcoming observations rather than global integration, for which all previous observations are considered. For a sub-optimal inference model with a static hidden state representation, the incorporation of leaky observation integration on a more local scale can serve as an approximation to optimal inference resorting to dynamic latent state representation and can thereby capture a belief about a changing environment [94].

Given a very large timescale, BS converges to zero as the divergence between prior and posterior distributions decreases over time, imposing an upper bound on the timescale. Meanwhile, for PS and CS it tends to lead to more accurate estimates of $p(y_t|s_t)$ as more observations are considered. However, given the regime switches in our data generation process, a trade-off exists where a timescale that is too large prevents flexible adaptation following such a switch. In the current context, the timescales are local enough where the estimated statistics are able to be adapted in response to regime switches (with a switch occurring every 100 stimuli on average). Especially CS shows a large range of τ-values producing similarly high model evidence due to the high correlation between regressors. In sum, it is possible that the same timescale is used for the computation of both the CS and BS signals, as the differences in optimal τ-values between clusters were not found to be significant. This interpretation is most intuitively compatible with the hypothesis that the early surprise signals may control later belief updating signals. Although the uncertainty regarding the exact half-lives is in line with the large variability found in the literature, local over global integration is consistently reported [9, 13, 39, 48, 94, 95]. Given a fixed inter-stimulus interval of 750ms, a horizon of 95 and 26 observations may be equated to a half-life timescale of approximately 71 to 20 seconds, with regime switches expected to occur every 75 seconds.

Some considerations of the current study deserve mention. First, the behavioural task required participants to make a decision about the identity of the stimulus so as to identify target (catch) trials. Thus, one may wonder to what extent the results contain conscious decision making signals, rather than implicit, non-conscious learning activity. However, decision making-related signals are described to occur relatively late in the trial [107, 108] and we assume to largely avoid them here by focusing on early signals prior to 200ms. Secondly, a large model space of both hierarchical as well as non-hierarchical Bayesian learners exists. As such, it is possible that the brain resorts to some hierarchical representation different from the ones tested here. We chose to use an HMM as it closely resembled the underlying data structure, offers the optimal solution for a discrete state environment, and contributes to the field as it has seen only limited application for probabilistic perceptual learning. Furthermore, some limitations might concern the stepwise model comparison intended to yield interpretable results by allowing inference on the generative model giving rise to surprise signatures. A reduction of both data and model space is not a standard procedure in Bayesian model comparison and we stress that we do not provide a methodological validation of this approach. Nevertheless, we argue that this scheme capitalizes on the hierarchical structure of the model space, provide model recoverability simulations, and present similar results using a standard factorial family comparison to support that the main conclusions are not dependent on the exact model comparison approach. The analyses performed here include a large number of independent Bayesian model comparisons (as is not uncommon in neuroimaging), yet no corrections are applied. While the resulting exceedance probabilities are reported here only above a given threshold, these model comparisons do not constitute statistical tests per se, as they do not provide a mapping from the data to binary outcomes. It follows that the analyses do not suffer from a classical multiple testing problem, which can be addressed using the control of multiple

testing error rates (e.g. the control of the family-wise error rate for fMRI inference based on random field theory). Nevertheless, it would be valuable for methodological advances to consider the possibility of randomly occurring high exceedance probabilities given a large number of independent model comparisons. A multilevel scheme which adjusts priors over models, rather than the current ubiquitous use of flat priors, may be developed as a satisfactory approach [109, 110, 111]. As the current method is agnostic to the large number of model comparisons we need to stress that we only report preliminary evidence.

In conclusion, we show that signals of early somatosensory processing can be accounted for by (surprise) signatures of Bayesian perceptual learning. The system appears to capture a changing environment using a static latent state model that integrates evidence on a local, rather than global, timescale and estimates transition probabilities of observations using first order dependencies. In turn, we provide evidence that the estimated statistics are used to compute a variety of surprise signatures in response to new observations, including both puzzlement surprise scaled by confidence (CS) in secondary somatosensory cortex and weak indications for enlightenment surprise (i.e. model updating; BS) in primary somatosensory cortex.

Supporting information

S1 Appendix. Bayesian learner models. In this supplementary text we provide the derivations for the presented equations of the compared Bayesian learner models.
(PDF)

S2 Appendix. A free-form variational inference algorithm for general linear models with spherical error covariance matrix. In this supplementary text we present the algorithm used to approximate log model evidence for subsequent Bayesian model comparison.
(PDF)

S1 Fig. Estimated emission probabilities and latent regime inference of the hidden Markov model. (A) The average emission probabilities of the stimulus probability (SP), alternation probability (AP), and transition probability (TP) hidden Markov model (HMM) for both states (s) at the final timestep of each sequence. For TP_2, a comparison is provided of the emission probabilities used for data generation and the average, normalized emission probabilities estimated by the HMM. Error bars represent the standard error of the mean. (B) Correlating the true regimes and filtering posterior over time confirms that AP and TP inference allow for the tracking of the fast and slow-switching regimes, while SP inference does not capture the necessary dependencies due to the regimes being balanced in terms of stimulus probabilities.
(TIF)

S2 Fig. Model-derived predictions for standard and deviant stimuli. Averaged surprise readouts using either the (left) 25000 total sequences or (right) 200 sequences administered to the participants elicited for standard and deviant stimuli following a certain amount of repeating stimuli (train length). The model-derived predictions are relatively well-preserved in the smaller data-set. Only first-order transition probability models are plotted. Error bars indicate standard deviations. The used stimulus half-lives of 95 and 26 are representative of the winning models in the single-trial EEG analysis. DC: Dirichlet-Categorical model; HMM: Hidden Markov Model; PS: Predictive surprise; BS: Bayesian surprise; CS: Confidence-corrected surprise; No F: model without forgetting (i.e. perfect integration); HL: stimulus half-life.
(TIF)

S3 Fig. Schematic of the hierarchical approach to family-wise Bayesian model selection. First level (depicted in the top row): The 12 DC models and the 12 HMM models were

grouped into their corresponding model class family and compared via BMS against each other and an offset Null-Model. Second level (lower row, left rectangle): Within the DC model class, the two transition probability models TP_1 and TP_2 were grouped into families and the winner of the BMS was used for the comparison against the other two inference type models (Stimulus Probability (SP) and Alternation Probability (AP)). Third Level (lower row, middle rectangle): The surprise readouts of the DC TP_1 model were subjected to BMS and the resulting exceedance probabilities are reported in the main results. Thresholding of the model class families and inference types was again applied at successive levels leading to data reduction.
(TIF)

S4 Fig. Non-hierarchical family-wise Bayesian model selection. Exceedance probabilities (φ) resulting from the RFX family model comparison by investigating the full model space in each comparison. A) Family comparison of the first order transition probability (TP_1), second order transition probability (TP_2), alternation probability (AP; no above-threshold results with $\varphi > 0.95$) and stimulus probability (SP) models; thresholded at $\varphi > 0.95$. B) Unthresholded family comparison of surprise models. Large discrete topographies show the electrode clusters of predictive surprise (PS) in red, Bayesian surprise (BS) in green and confidence-corrected surprise (CS) in blue. White asterisks indicate $\varphi > 0.95$. Small continuous topographies display the converged variational expectation parameter (m_β).
(TIF)

S5 Fig. Model recovery study. A model recovery study was performed using simulated data. Subplots (A-D) show the average exceedance probabilities (shading represents standard deviations) of 100 random-effects Bayesian model selection analyses under different signal-to-noise ratios. This was performed for (A) Null Model vs DC Model vs HMM families, (B) DC TP_1 vs TP_2 families, (C) DC SP vs AP vs TP_1 families, and (D) DC TP_1 PS, BS, and CS models. Noteworthy is that the instances of reduced differentiability for (B) and (C) occurred only when the true, but unknown model was confidence-corrected surprise. (E) An estimate of the signal-to-noise of the experimental single-trial EEG analyses by inspecting the ratio of the expected posterior estimates of the model fitting procedure for β^2 and λ^{-1}.
(TIF)

S6 Fig. Expected posterior probabilities of hierarchical Bayesian model-selection. Expected posterior probabilities ($\langle r \rangle$) resulting from family model comparisons. A) Dirichlet-Categorical (DC) model, Hidden Markov Model (HMM) and Null model family comparison, thresholded at $\langle r \rangle > 0.75$. B) Family comparison within the winning DC family, thresholded at $\langle r \rangle > 0.7$: first and second order transition probability models (TP_1, TP_2). C) Family comparison within the winning DC family, thresholded at $\langle r \rangle > 0.7$: first order transition probability (TP_1), alternation probability (AP) and stimulus probability (SP) models.
(TIF)

S7 Fig. Additional random effects family-wise comparisons. (A) Comparison of the model families: Null model, Dirichlet-Categorical model (DC) with tau = 0 (i.e. no forgetting and no penalization) and Hidden Markov Model (HMM). (B) Comparison of the model families: Null model, DC without modelling the catch trials and HMM. (C) Comparison of the model families: Null model, DC with and DC without modelling the catch trials. (D) Comparison of the model families within the DC model: Stimulus probability model (SP), alternation probability model (AP) and transition probability model family (TP) subsuming first and second order TP models in one family. Exceedance probabilities (φ) are plotted for all comparisons.
(TIF)

Acknowledgments

The authors would like to thank the HPC Service of ZEDAT, Freie Universität Berlin, for computing time.

Author Contributions

Conceptualization: Sam Gijsen, Miro Grundei, Robert T. Lange, Dirk Ostwald, Felix Blankenburg.

Data curation: Sam Gijsen, Miro Grundei.

Formal analysis: Sam Gijsen, Miro Grundei.

Funding acquisition: Felix Blankenburg.

Investigation: Sam Gijsen, Miro Grundei.

Methodology: Sam Gijsen, Miro Grundei, Robert T. Lange, Dirk Ostwald.

Project administration: Sam Gijsen, Miro Grundei, Felix Blankenburg.

Resources: Felix Blankenburg.

Software: Sam Gijsen, Miro Grundei, Robert T. Lange, Dirk Ostwald.

Supervision: Dirk Ostwald, Felix Blankenburg.

Validation: Sam Gijsen, Miro Grundei.

Visualization: Sam Gijsen, Miro Grundei, Robert T. Lange.

Writing – original draft: Sam Gijsen, Miro Grundei, Robert T. Lange.

Writing – review & editing: Sam Gijsen, Miro Grundei, Dirk Ostwald, Felix Blankenburg.

References

1. Helmholtz Hv. Treatise of physiological optics: Concerning the perceptions in general. Classics in psychology. 1856; p. 79–127.
2. Rao RP, Ballard DH. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nature neuroscience. 1999; 2(1):79–87. https://doi.org/10.1038/4580 PMID: 10195184
3. Friston K, Kiebel S. Predictive coding under the free-energy principle. Philosophical Transactions of the Royal Society B: Biological Sciences. 2009; 364(1521):1211–1221. https://doi.org/10.1098/rstb.2008.0300 PMID: 19528002
4. Friston K. The free-energy principle: a unified brain theory? Nature reviews neuroscience. 2010; 11 (2):127. https://doi.org/10.1038/nrn2787 PMID: 20068583
5. Knill DC, Pouget A. The Bayesian brain: the role of uncertainty in neural coding and computation. TRENDS in Neurosciences. 2004; 27(12):712–719. https://doi.org/10.1016/j.tins.2004.10.007 PMID: 15541511
6. Friston K. A theory of cortical responses. Philosophical transactions of the Royal Society B: Biological sciences. 2005; 360(1456):815–836. https://doi.org/10.1098/rstb.2005.1622 PMID: 15937014
7. Winkler I, Denham SL, Nelken I. Modeling the auditory scene: predictive regularity representations and perceptual objects. Trends in cognitive sciences. 2009; 13(12):532–540. https://doi.org/10.1016/j.tics.2009.09.003 PMID: 19828357
8. Lieder F, Daunizeau J, Garrido MI, Friston KJ, Stephan KE. Modelling trial-by-trial changes in the mismatch negativity. PLoS computational biology. 2013; 9(2). https://doi.org/10.1371/journal.pcbi.1002911 PMID: 23436989
9. Maheu M, Dehaene S, Meyniel F. Brain signatures of a multiscale process of sequence learning in humans. Elife. 2019; 8:e41541. https://doi.org/10.7554/eLife.41541 PMID: 30714904

10. Turk-Browne NB, Scholl BJ, Johnson MK, Chun MM. Implicit perceptual anticipation triggered by statistical learning. Journal of Neuroscience. 2010; 30(33):11177–11187. https://doi.org/10.1523/JNEUROSCI.0858-10.2010 PMID: 20720125

11. O'Reilly JX, Schüffelgen U, Cuell SF, Behrens TE, Mars RB, Rushworth MF. Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. Proceedings of the National Academy of Sciences. 2013; 110(38):E3660–E3669. https://doi.org/10.1073/pnas.1305373110 PMID: 23986499

12. Stefanics G, Heinzle J, Horváth AA, Stephan KE. Visual mismatch and predictive coding: A computational single-trial ERP study. Journal of Neuroscience. 2018; 38(16):4020–4030. https://doi.org/10.1523/JNEUROSCI.3365-17.2018 PMID: 29581379

13. 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. https://doi.org/10.1016/j.neuroimage.2012.04.050 PMID: 22579866

14. Squires NK, Squires KC, Hillyard SA. Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. Electroencephalography and clinical neurophysiology. 1975; 38 (4):387–401. https://doi.org/10.1016/0013-4694(75)90263-1 PMID: 46819

15. Baldeweg T, Klugman A, Gruzelier J, Hirsch SR. Mismatch negativity potentials and cognitive impairment in schizophrenia. Schizophrenia research. 2004; 69(2-3):203–217. https://doi.org/10.1016/j.schres.2003.09.009 PMID: 15469194

16. Stefanics G, Kremláček J, Czigler I. Visual mismatch negativity: a predictive coding view. Frontiers in human neuroscience. 2014; 8:666. https://doi.org/10.3389/fnhum.2014.00666 PMID: 25278859

17. Kekoni J, Hämäläinen H, Saarinen M, Gröhn J, Reinikainen K, Lehtokoski A, et al. Rate effect and mismatch responses in the somatosensory system: ERP-recordings in humans. Biological psychology. 1997; 46(2):125–142. https://doi.org/10.1016/S0301-0511(97)05249-6 PMID: 9288410

18. Shinozaki N, Yabe H, Sutoh T, Hiruma T, Kaneko S. Somatosensory automatic responses to deviant stimuli. Cognitive Brain Research. 1998; 7(2):165–171. https://doi.org/10.1016/S0926-6410(98)00020-2 PMID: 9774724

19. Akatsuka K, Wasaka T, Nakata H, Inui K, Hoshiyama M, Kakigi R. Mismatch responses related to temporal discrimination of somatosensory stimulation. Clinical neurophysiology. 2005; 116(8):1930–1937. https://doi.org/10.1016/j.clinph.2005.04.021 PMID: 15982927

20. Huang MX, Lee RR, Miller GA, Thoma RJ, Hanlon FM, Paulson KM, et al. A parietal–frontal network studied by somatosensory oddball MEG responses, and its cross-modal consistency. Neuroimage. 2005; 28(1):99–114. https://doi.org/10.1016/j.neuroimage.2005.05.036 PMID: 15979344

21. Akatsuka K, Wasaka T, Nakata H, Kida T, Hoshiyama M, Tamura Y, et al. Objective examination for two-point stimulation using a somatosensory oddball paradigm: an MEG study. Clinical neurophysiology. 2007; 118(2):403–411. https://doi.org/10.1016/j.clinph.2006.09.030 PMID: 17095288

22. Akatsuka K, Wasaka T, Nakata H, Kida T, Kakigi R. The effect of stimulus probability on the somatosensory mismatch field. Experimental brain research. 2007; 181(4):607–614. https://doi.org/10.1007/s00221-007-0958-4 PMID: 17516059

23. Restuccia D, Marca GD, Valeriani M, Leggio MG, Molinari M. Cerebellar damage impairs detection of somatosensory input changes. A somatosensory mismatch-negativity study. Brain. 2007; 130(1):276–287. https://doi.org/10.1093/brain/awl236 PMID: 16982654

24. Spackman L, Towell A, Boyd S. Somatosensory discrimination: an intracranial event-related potential study of children with refractory epilepsy. Brain research. 2010; 1310:68–76. https://doi.org/10.1016/j.brainres.2009.10.072 PMID: 19896930

25. Naeije G, Vaulet T, Wens V, Marty B, Goldman S, De Tiège X. Multilevel cortical processing of somatosensory novelty: a magnetoencephalography study. Frontiers in human neuroscience. 2016; 10:259. https://doi.org/10.3389/fnhum.2016.00259 PMID: 27313523

26. Naeije G, Vaulet T, Wens V, Marty B, Goldman S, De Tiege X. Neural basis of early somatosensory change detection: a magnetoencephalography study. Brain topography. 2018; 31(2):242–256. https://doi.org/10.1007/s10548-017-0591-x PMID: 28913778

27. Spackman L, Boyd S, Towell A. Effects of stimulus frequency and duration on somatosensory discrimination responses. Experimental brain research. 2007; 177(1):21. https://doi.org/10.1007/s00221-006-0650-0 PMID: 16917766

28. Butler JS, Foxe JJ, Fiebelkorn IC, Mercier MR, Molholm S. Multisensory representation of frequency across audition and touch: high density electrical mapping reveals early sensory-perceptual coupling. Journal of Neuroscience. 2012; 32(44):15338–15344. https://doi.org/10.1523/JNEUROSCI.1796-12.2012 PMID: 23115172

29. Chennu S, Noreika V, Gueorguiev D, Blenkmann A, Kochen S, Ibánez A, et al. Expectation and attention in hierarchical auditory prediction. Journal of Neuroscience. 2013; 33(27):11194–11205. https://doi.org/10.1523/JNEUROSCI.0114-13.2013 PMID: 23825422

30. Butler JS, Molholm S, Fiebelkorn IC, Mercier MR, Schwartz TH, Foxe JJ. Common or redundant neural circuits for duration processing across audition and touch. Journal of Neuroscience. 2011; 31 (9):3400–3406. https://doi.org/10.1523/JNEUROSCI.3296-10.2011 PMID: 21368051

31. Hu L, Zhao C, Li H, Valentini E. Mismatch responses evoked by nociceptive stimuli. Psychophysiology. 2013; 50(2):158–173. https://doi.org/10.1111/psyp.12000 PMID: 23256883

32. 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. https://doi.org/10.1016/j.neuroimage.2008.05.018 PMID: 18602841

33. Garrido MI, Kilner JM, Stephan KE, Friston KJ. The mismatch negativity: a review of underlying mechanisms. Clinical neurophysiology. 2009; 120(3):453–463. https://doi.org/10.1016/j.clinph.2008.11.029 PMID: 19181570

34. Friston K. Learning and inference in the brain. Neural Networks. 2003; 16(9):1325–1352. https://doi.org/10.1016/j.neunet.2003.06.005 PMID: 14622888

35. Faraji M, Preuschoff K, Gerstner W. Balancing new against old information: the role of puzzlement surprise in learning. Neural computation. 2018; 30(1):34–83. https://doi.org/10.1162/neco_a_01025 PMID: 29064784

36. Shannon CE. A mathematical theory of communication. Bell system technical journal. 1948; 27 (3):379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

37. Itti L, Baldi P. Bayesian surprise attracts human attention. Vision research. 2009; 49(10):1295–1306. https://doi.org/10.1016/j.visres.2008.09.007 PMID: 18834898

38. Baldi P, Itti L. Of bits and wows: A Bayesian theory of surprise with applications to attention. Neural Networks. 2010; 23(5):649–666. https://doi.org/10.1016/j.neunet.2009.12.007 PMID: 20080025

39. Kolossa A, Fingscheidt T, Wessel K, Kopp B. A model-based approach to trial-by-trial P300 amplitude fluctuations. Frontiers in human neuroscience. 2013; 6:359. https://doi.org/10.3389/fnhum.2012.00359 PMID: 23404628

40. Kolossa A, Kopp B, Fingscheidt T. A computational analysis of the neural bases of Bayesian inference. Neuroimage. 2015; 106:222–237. https://doi.org/10.1016/j.neuroimage.2014.11.007 PMID: 25462794

41. Kopp B, Seer C, Lange F, Kluytmans A, Kolossa A, Fingscheidt T, et al. P300 amplitude variations, prior probabilities, and likelihoods: A Bayesian ERP study. Cognitive, Affective, & Behavioral Neuroscience. 2016; 16(5):911–928. https://doi.org/10.3758/s13415-016-0442-3 PMID: 27406085

42. Modirshanechi A, Kiani MM, Aghajan H. Trial-by-trial surprise-decoding model for visual and auditory binary oddball tasks. Neuroimage. 2019; 196:302–317. https://doi.org/10.1016/j.neuroimage.2019.04.028 PMID: 30980899

43. Mars RB, Debener S, Gladwin TE, Harrison LM, Haggard P, Rothwell JC, et al. Trial-by-trial fluctuations in the event-related electroencephalogram reflect dynamic changes in the degree of surprise. Journal of Neuroscience. 2008; 28(47):12539–12545. https://doi.org/10.1523/JNEUROSCI.2925-08.2008 PMID: 19020046

44. Seer C, Lange F, Boos M, Dengler R, Kopp B. Prior probabilities modulate cortical surprise responses: a study of event-related potentials. Brain and cognition. 2016; 106:78–89. https://doi.org/10.1016/j.bandc.2016.04.011 PMID: 27266394

45. Schwartenbeck P, FitzGerald TH, Dolan R. Neural signals encoding shifts in beliefs. Neuroimage. 2016; 125:578–586. https://doi.org/10.1016/j.neuroimage.2015.10.067 PMID: 26520774

46. Kobayashi K, Hsu M. Neural mechanisms of updating under reducible and irreducible uncertainty. Journal of Neuroscience. 2017; 37(29):6972–6982. https://doi.org/10.1523/JNEUROSCI.0535-17.2017 PMID: 28626019

47. Visalli A, Capizzi M, Ambrosini E, Mazzonetto I, Vallesi A. Bayesian modeling of temporal expectations in the human brain. Neuroimage. 2019; 202:116097. https://doi.org/10.1016/j.neuroimage.2019.116097 PMID: 31415885

48. Mousavi Z, Kiani MM, Aghajan H. Brain signatures of surprise in EEG and MEG data. bioRxiv. 2020.

49. Berg P, Scherg M. A multiple source approach to the correction of eye artifacts. Electroencephalography and clinical neurophysiology. 1994; 90(3):229–241. https://doi.org/10.1016/0013-4694(94)90094-9 PMID: 7511504

50. Kilner JM, Kiebel SJ, Friston KJ. Applications of random field theory to electrophysiology. Neuroscience letters. 2005; 374(3):174–178. https://doi.org/10.1016/j.neulet.2004.10.052 PMID: 15663957

51. Linden DE. The P300: where in the brain is it produced and what does it tell us? The Neuroscientist. 2005; 11(6):563–576. https://doi.org/10.1177/1073858405280524 PMID: 16282597

52. Sabeti M, Katebi S, Rastgar K, Azimifar Z. A multi-resolution approach to localize neural sources of P300 event-related brain potential. Computer methods and programs in biomedicine. 2016; 133:155–168. https://doi.org/10.1016/j.cmpb.2016.05.013 PMID: 27393807

53. Strömmer JM, Tarkka IM, Astikainen P. Somatosensory mismatch response in young and elderly adults. Frontiers in aging neuroscience. 2014; 6:293. https://doi.org/10.3389/fnagi.2014.00293 PMID: 25386140

54. Friston K, Harrison L, Daunizeau J, Kiebel S, Phillips C, Trujillo-Barreto N, et al. Multiple sparse priors for the M/EEG inverse problem. Neuroimage. 2008; 39(3):1104–1120. https://doi.org/10.1016/j.neuroimage.2007.09.048 PMID: 17997111

55. Litvak V, Friston K. Electromagnetic source reconstruction for group studies. Neuroimage. 2008; 42 (4):1490–1498. https://doi.org/10.1016/j.neuroimage.2008.06.022 PMID: 18639641

56. Rabiner L, Juang B. An introduction to hidden Markov models. ieee assp magazine. 1986; 3(1):4–16. https://doi.org/10.1109/MASSP.1986.1165342

57. hmmlearn; 2019. Available from: https://github.com/hmmlearn/hmmlearn.

58. Flandin G, Penny WD. Bayesian fMRI Data Analysis with Sparse Spatial Basis Function Priors. Neuroimage. 2007; 34(3):1108–1125. https://doi.org/10.1016/j.neuroimage.2006.10.005 PMID: 17157034

59. Penny WD, Trujillo-Barreto NJ, Friston KJ. Bayesian fMRI Time Series Analysis with Spatial Priors. Neuroimage. 2005; 24(2):350–362. https://doi.org/10.1016/j.neuroimage.2004.08.034 PMID: 15627578

60. Penny W, Kiebel S, Friston K. Variational Bayesian Inference for fMRI Time Series. Neuroimage. 2003; 19(3):727–741. https://doi.org/10.1016/S1053-8119(03)00071-5 PMID: 12880802

61. Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ. Bayesian model selection for group studies. Neuroimage. 2009; 46(4):1004–1017. https://doi.org/10.1016/j.neuroimage.2009.03.025 PMID: 19306932

62. Woolrich MW. Bayesian inference in FMRI. Neuroimage. 2012; 62(2):801–810. https://doi.org/10.1016/j.neuroimage.2011.10.047 PMID: 22063092

63. Ossmy O, Moran R, Pfeffer T, Tsetsos K, Usher M, Donner TH. The timescale of perceptual evidence integration can be adapted to the environment. Current Biology. 2013; 23(11):981–986. https://doi.org/10.1016/j.cub.2013.04.039 PMID: 23684972

64. Runyan CA, Piasini E, Panzeri S, Harvey CD. Distinct timescales of population coding across cortex. Nature. 2017; 548(7665):92–96. https://doi.org/10.1038/nature23020 PMID: 28723889

65. Penny WD, Stephan KE, Daunizeau J, Rosa MJ, Friston KJ, Schofield TM, et al. Comparing families of dynamic causal models. PLoS computational biology. 2010; 6(3). https://doi.org/10.1371/journal.pcbi.1000709 PMID: 20300649

66. Rigoux L, Stephan KE, Friston KJ, Daunizeau J. Bayesian model selection for group studies—revisited. Neuroimage. 2014; 84:971–985. https://doi.org/10.1016/j.neuroimage.2013.08.065 PMID: 24018303

67. Fastenrath M, Friston KJ, Kiebel SJ. Dynamical causal modelling for M/EEG: spatial and temporal symmetry constraints. Neuroimage. 2009; 44(1):154–163. https://doi.org/10.1016/j.neuroimage.2008.07.041 PMID: 18718870

68. Otten LJ, Alain C, Picton TW. Effects of visual attentional load on auditory processing. Neuroreport. 2000; 11(4):875–880. https://doi.org/10.1097/00001756-200003200-00043 PMID: 10757537

69. Jones SR, Pritchett DL, Stufflebeam SM, Hämäläinen M, Moore CI. Neural correlates of tactile detection: a combined magnetoencephalography and biophysically based computational modeling study. Journal of Neuroscience. 2007; 27(40):10751–10764. https://doi.org/10.1523/JNEUROSCI.0482-07.2007 PMID: 17913909

70. Avanzini P, Abdollahi RO, Sartori I, Caruana F, Pelliccia V, Casaceli G, et al. Four-dimensional maps of the human somatosensory system. Proceedings of the National Academy of Sciences. 2016; 113 (13):E1936–E1943. https://doi.org/10.1073/pnas.1601889113 PMID: 26976579

71. Avanzini P, Pelliccia V, Russo GL, Orban GA, Rizzolatti G. Multiple time courses of somatosensory responses in human cortex. Neuroimage. 2018; 169:212–226. https://doi.org/10.1016/j.neuroimage.2017.12.037 PMID: 29248698

72. Squires KC, Wickens C, Squires NK, Donchin E. The effect of stimulus sequence on the waveform of the cortical event-related potential. Science. 1976; 193(4258):1142–1146. https://doi.org/10.1126/science.959831 PMID: 959831

73. Duncan-Johnson CC, Donchin E. On quantifying surprise: The variation of event-related potentials with subjective probability. Psychophysiology. 1977; 14(5):456–467. https://doi.org/10.1111/j.1469-8986.1977.tb01312.x PMID: 905483

74. Polich J. Updating P300: an integrative theory of P3a and P3b. Clinical neurophysiology. 2007; 118 (10):2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019 PMID: 17573239

75. Isreal JB, Chesney GL, Wickens CD, Donchin E. P300 and tracking difficulty: Evidence for multiple resources in dual-task performance. Psychophysiology. 1980; 17(3):259–273. https://doi.org/10.1111/j.1469-8986.1980.tb00146.x PMID: 7384376

76. Kramer AF, Wickens CD, Donchin E. Processing of stimulus properties: evidence for dual-task integrality. Journal of Experimental Psychology: Human Perception and Performance. 1985; 11(4):393. PMID: 3161983

77. Wickens C, Kramer A, Vanasse L, Donchin E. Performance of concurrent tasks: a psychophysiological analysis of the reciprocity of information-processing resources. Science. 1983; 221(4615):1080–1082. https://doi.org/10.1126/science.6879207 PMID: 6879207

78. Kida T, Nishihira Y, Hatta A, Wasaka T, Tazoe T, Sakajiri Y, et al. Resource allocation and somatosensory P300 amplitude during dual task: effects of tracking speed and predictability of tracking direction. Clinical Neurophysiology. 2004; 115(11):2616–2628. https://doi.org/10.1016/j.clinph.2004.06.013 PMID: 15465451

79. Kok A. On the utility of P3 amplitude as a measure of processing capacity. Psychophysiology. 2001; 38(3):557–577. https://doi.org/10.1017/S0048577201990559 PMID: 11352145

80. Haenschel C, Vernon DJ, Dwivedi P, Gruzelier JH, Baldeweg T. Event-related brain potential correlates of human auditory sensory memory-trace formation. Journal of Neuroscience. 2005; 25 (45):10494–10501. https://doi.org/10.1523/JNEUROSCI.1227-05.2005 PMID: 16280587

81. Baldeweg T. Repetition effects to sounds: evidence for predictive coding in the auditory system. Trends in cognitive sciences. 2006;. https://doi.org/10.1016/j.tics.2006.01.010 PMID: 16460994

82. Summerfield C, Trittschuh EH, Monti JM, Mesulam MM, Egner T. Neural repetition suppression reflects fulfilled perceptual expectations. Nature neuroscience. 2008; 11(9):1004. https://doi.org/10.1038/nn.2163 PMID: 19160497

83. Auksztulewicz R, Friston K. Repetition suppression and its contextual determinants in predictive coding. cortex. 2016; 80:125–140. https://doi.org/10.1016/j.cortex.2015.11.024 PMID: 26861557

84. Van Zuijen TL, Simoens VL, Paavilainen P, Näätänen R, Tervaniemi M. Implicit, intuitive, and explicit knowledge of abstract regularities in a sound sequence: an event-related brain potential study. Journal of Cognitive Neuroscience. 2006; 18(8):1292–1303. https://doi.org/10.1162/jocn.2006.18.8.1292 PMID: 16859415

85. Atas A, Faivre N, Timmermans B, Cleeremans A, Kouider S. Nonconscious learning from crowded sequences. Psychological science. 2014; 25(1):113–119. https://doi.org/10.1177/0956797613499591 PMID: 24186918

86. Koelsch S, Busch T, Jentschke S, Rohrmeier M. Under the hood of statistical learning: A statistical MMN reflects the magnitude of transitional probabilities in auditory sequences. Scientific reports. 2016; 6:19741. https://doi.org/10.1038/srep19741 PMID: 26830652

87. Green C, Benson C, Kersten D, Schrater P. Alterations in choice behavior by manipulations of world model. Proceedings of the National Academy of Sciences. 2010; 107(37):16401–16406. https://doi.org/10.1073/pnas.1001709107 PMID: 20805507

88. Behrens TE, Woolrich MW, Walton ME, Rushworth MF. Learning the value of information in an uncertain world. Nature neuroscience. 2007; 10(9):1214–1221. https://doi.org/10.1038/nn1954 PMID: 17676057

89. Heilbron M, Meyniel F. Confidence resets reveal hierarchical adaptive learning in humans. PLoS computational biology. 2019; 15(4):e1006972. https://doi.org/10.1371/journal.pcbi.1006972 PMID: 30964861

90. Summerfield C, Behrens TE, Koechlin E. Perceptual classification in a rapidly changing environment. Neuron. 2011; 71(4):725–736. https://doi.org/10.1016/j.neuron.2011.06.022 PMID: 21867887

91. Farashahi S, Donahue CH, Khorsand P, Seo H, Lee D, Soltani A. Metaplasticity as a neural substrate for adaptive learning and choice under uncertainty. Neuron. 2017; 94(2):401–414. https://doi.org/10.1016/j.neuron.2017.03.044 PMID: 28426971

92. Gershman SJ, Norman KA, Niv Y. Discovering latent causes in reinforcement learning. Current Opinion in Behavioral Sciences. 2015; 5:43–50. https://doi.org/10.1016/j.cobeha.2015.07.007

93. Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, Friston KJ. Canonical microcircuits for predictive coding. Neuron. 2012; 76(4):695–711. https://doi.org/10.1016/j.neuron.2012.10.038 PMID: 23177956

94. Meyniel F, Maheu M, Dehaene S. Human inferences about sequences: A minimal transition probability model. PLoS computational biology. 2016; 12(12). https://doi.org/10.1371/journal.pcbi.1005260 PMID: 28030543

95. Rubin J, Ulanovsky N, Nelken I, Tishby N. The representation of prediction error in auditory cortex. PLoS computational biology. 2016; 12(8). https://doi.org/10.1371/journal.pcbi.1005058 PMID: 27490251

96. Payzan-LeNestour E, Bossaerts P. Risk, unexpected uncertainty, and estimation uncertainty: Bayesian learning in unstable settings. PLoS computational biology. 2011; 7(1). https://doi.org/10.1371/journal.pcbi.1001048 PMID: 21283774

97. Meyniel F, Dehaene S. Brain networks for confidence weighting and hierarchical inference during probabilistic learning. Proceedings of the National Academy of Sciences. 2017; 114(19):E3859–E3868. https://doi.org/10.1073/pnas.1615773114 PMID: 28439014

98. Boldt A, Blundell C, De Martino B. Confidence modulates exploration and exploitation in value-based learning. Neuroscience of consciousness. 2019; 2019(1):niz004. https://doi.org/10.1093/nc/niz004 PMID: 31086679

99. Meyniel F, Schlunegger D, Dehaene S. The sense of confidence during probabilistic learning: A normative account. PLoS computational biology. 2015; 11(6). https://doi.org/10.1371/journal.pcbi.1004305 PMID: 26076466

100. Meyniel F. Brain dynamics for confidence-weighted learning. PLOS Computational Biology. 2020; 16 (6):e1007935. https://doi.org/10.1371/journal.pcbi.1007935 PMID: 32484806

101. Mathys C, Daunizeau J, Friston KJ, Stephan KE. A Bayesian foundation for individual learning under uncertainty. Frontiers in human neuroscience. 2011; 5:39. https://doi.org/10.3389/fnhum.2011.00039 PMID: 21629826

102. Mathys CD, Lomakina EI, Daunizeau J, Iglesias S, Brodersen KH, Friston KJ, et al. Uncertainty in perception and the Hierarchical Gaussian Filter. Frontiers in human neuroscience. 2014; 8:825. https://doi.org/10.3389/fnhum.2014.00825 PMID: 25477800

103. Iglesias S, Mathys C, Brodersen KH, Kasper L, Piccirelli M, den Ouden HE, et al. Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. Neuron. 2013; 80(2):519–530. https://doi.org/10.1016/j.neuron.2013.09.009 PMID: 24139048

104. Meyniel F, Sigman M, Mainen ZF. Confidence as Bayesian probability: From neural origins to behavior. Neuron. 2015; 88(1):78–92. https://doi.org/10.1016/j.neuron.2015.09.039 PMID: 26447574

105. Pleger B, Foerster AF, Ragert P, Dinse HR, Schwenkreis P, Malin JP, et al. Functional imaging of perceptual learning in human primary and secondary somatosensory cortex. Neuron. 2003; 40(3):643–653. https://doi.org/10.1016/S0896-6273(03)00677-9 PMID: 14642286

106. Kopp B, Lange F. Electrophysiological indicators of surprise and entropy in dynamic task-switching environments. Frontiers in Human Neuroscience. 2013; 7:300. https://doi.org/10.3389/fnhum.2013.00300 PMID: 23840183

107. Herding J, Ludwig S, von Lautz A, Spitzer B, Blankenburg F. Centro-parietal EEG potentials index subjective evidence and confidence during perceptual decision making. Neuroimage. 2019; 201:116011. https://doi.org/10.1016/j.neuroimage.2019.116011 PMID: 31302254

108. Kelly SP, O'Connell RG. The neural processes underlying perceptual decision making in humans: recent progress and future directions. Journal of Physiology-Paris. 2015; 109(1-3):27–37. https://doi.org/10.1016/j.jphysparis.2014.08.003 PMID: 25204272

109. Friston KJ, Glaser DE, Henson RN, Kiebel S, Phillips C, Ashburner J. Classical and Bayesian inference in neuroimaging: applications. Neuroimage. 2002; 16(2):484–512. https://doi.org/10.1006/nimg.2002.1090 PMID: 12030833

110. Gelman A, Hill J, Yajima M. Why we (usually) don't have to worry about multiple comparisons. Journal of Research on Educational Effectiveness. 2012; 5(2):189–211. https://doi.org/10.1080/19345747.2011.618213

111. Neath AA, Flores JE, Cavanaugh JE. Bayesian multiple comparisons and model selection. Wiley Interdisciplinary Reviews: Computational Statistics. 2018; 10(2):e1420. https://doi.org/10.1002/wics.1420

Original Publication of Study 2

Grundei, M., Schröder, P., Gijsen, S. and Blankenburg, F. (2023). EEG mismatch responses in a multimodal roving stimulus paradigm provide evidence for probabilistic inference across audition, somatosensation and vision. *Human Brain Mapping*, 1– 25.

Received: 14 October 2022 | Revised: 17 March 2023 | Accepted: 24 March 2023

DOI: 10.1002/hbm.26303

RESEARCH ARTICLE

WILEY

EEG mismatch responses in a multimodal roving stimulus paradigm provide evidence for probabilistic inference across audition, somatosensation, and vision

Miro Grundei[1,2] | Pia Schröder[1] | Sam Gijsen[1,2] | Felix Blankenburg[1,2]

[1]Neurocomputation and Neuroimaging Unit, Freie Universität Berlin, Berlin, 14195, Germany

[2]Berlin School of Mind and Brain, Humboldt Universität zu Berlin, Berlin, 10117, Germany

Correspondence
Miro Grundei, Neurocomputation and Neuroimaging Unit, Freie Universität Berlin, 14195 Berlin, Germany.
Email: m.grundei@fu-berlin.de

Funding information
Berlin School of Mind and Brain; Deutscher Akademischer Austauschdienst

Abstract

The human brain is constantly subjected to a multimodal stream of probabilistic sensory inputs. Electroencephalography (EEG) signatures, such as the mismatch negativity (MMN) and the P3, can give valuable insight into neuronal probabilistic inference. Although reported for different modalities, mismatch responses have largely been studied in isolation, with a strong focus on the auditory MMN. To investigate the extent to which early and late mismatch responses across modalities represent comparable signatures of uni- and cross-modal probabilistic inference in the hierarchically structured cortex, we recorded EEG from 32 participants undergoing a novel trimodal roving stimulus paradigm. The employed sequences consisted of high and low intensity stimuli in the auditory, somatosensory and visual modalities and were governed by unimodal transition probabilities and cross-modal conditional dependencies. We found modality specific signatures of MMN (~100–200 ms) in all three modalities, which were source localized to the respective sensory cortices and shared right lateralized prefrontal sources. Additionally, we identified a cross-modal signature of mismatch processing in the P3a time range (~300–350 ms), for which a common network with frontal dominance was found. Across modalities, the mismatch responses showed highly comparable parametric effects of stimulus train length, which were driven by standard and deviant response modulations in opposite directions. Strikingly, P3a responses across modalities were increased for mispredicted stimuli with low cross-modal conditional probability, suggesting sensitivity to multimodal (global) predictive sequence properties. Finally, model comparisons indicated that the observed single trial dynamics were best captured by Bayesian learning models tracking unimodal stimulus transitions as well as cross-modal conditional dependencies.

KEYWORDS
Bayesian inference, cross-modal, mismatch negativity, multisensory, P3, predictive processing, surprise

1 | INTRODUCTION

Humans inhabit a highly structured environment governed by complex regularities. The brain is subjected to such environmental regularities by a multimodal stream of sensory inputs ultimately constructing a perceptual representation of the world. The sensory system is thought to capitalize on statistical regularities to efficiently guide interaction with the world enabling anticipation and rapid detection of sensory changes (Bregman, 1994; Dehaene et al., 2015; Friston, 2005; Frost et al., 2015; Gregory, 1980; Winkler et al., 2009).

Neuronal responses to deviations from sensory regularities can be valuable windows into the brain's processing of statistical properties of the environment and corresponding sensory predictions. The presentation of rare deviant sounds within a sequence of repeating standard sounds induces well known mismatch responses (MMRs) that can be recorded with electroencephalography (EEG), such as the mismatch negativity (MMN; Naatanen et al., 1978; Naatanen et al., 2007) and the P3 (or P300; Polich, 2007; Squires et al., 1975; Sutton et al., 1965). The MMN is defined as a negative EEG component resulting from subtraction of standard from deviant trials between ~100 and 200 ms poststimulus. Although the MMN has primarily been researched in the auditory modality, similar early mismatch components have been reported for other sensory modalities, including the visual (Kimura et al., 2011; Pazo-Alvarez et al., 2003; Stefanics et al., 2014) and, to a lesser extent, the somatosensory modality (Andersen & Lundqvist, 2019; Hu et al., 2013; Kekoni et al., 1997). The P3 is a later positive going component in response to novelty between 200 and 600 ms around central electrodes, which has been described for the auditory, somatosensory, and visual modalities and is known for its modality independent characteristics (Escera et al., 2000; Friedman et al., 2001; Knight & Scabini, 1998; Polich, 2007; Schroger, 1996).

Despite being one of the most well-studied EEG components, the neuronal generation of the MMN remains subject of ongoing debate (Garrido, Kilner, Stephan, & Friston, 2009; May & Tiitinen, 2010; Naatanen et al., 2005). Two prominent but opposing accounts cast the MMN as adaptation-based or memory-based, respectively. Adaptation-based accounts argue that the observed differences between standard and deviant responses primarily result from neuronal attenuation leading to stimulus specific adaptation (SSA; Jaaskelainen et al., 2004; May et al., 1999). In animals, SSA has been shown to result in response patterns similar to the MMN (Ulanovsky et al., 2003; Ulanovsky et al., 2004) and simulation work suggests that different types of MMN-like responses can be reproduced by pure adaptation models (May & Tiitinen, 2010). However, it remains unclear if the full range of MMN characteristics can be explained by adaptation alone (Fitzgerald & Todd, 2020; Garrido, Kilner, Stephan, & Friston, 2009; Wacongne et al., 2012). The memory-based view, on the other hand, suggests that the MMN is a marker of change detection based on sensory memory trace formation (Näätänen, 1990; Naatanen et al., 2005; Naatanen & Näätänen, 1992). The memory trace stores local information on stimulus regularity and compares it to incoming sensory inputs that may signal changes in the current sensory stream.

While largely neglected by previous interpretations of the MMN, it is becoming increasingly clear that key empirical features of MMRs concern stimulus predictability rather than stimulus change per se. The MMN has been reported in response to abstract rule violations (Paavilainen, 2013), unexpected stimulus repetitions (Alain et al., 1994; Horvath & Winkler, 2004; Macdonald & Campbell, 2011) and unexpected stimulus omissions (Heilbron & Chait, 2018; Hughes et al., 2001; Salisbury, 2012; Wacongne et al., 2011; Yabe et al., 1997). Similar characteristics have been reported for P3 MMRs (Duncan et al., 2009; Prete et al., 2022) and both MMN and P3 responses have been shown to increase for unexpected compared to expected deviants (Schroger et al., 2015; Sussman, 2005; Sussman et al., 1998). Insights concerning the predictive nature of MMRs have led to further development of the memory-based account of MMN generation into the model-adjustment hypothesis (Winkler, 2007). This view assumes a perceptual model that is informed by previous stimulus exposure and continually predicts incoming sensory inputs. The model is updated whenever inputs diverge from current predictions, and the MMN is hypothesized to constitute a marker of such divergence.

The model-adjustment hypothesis is in line with the increasingly influential view that the brain is engaging in perceptual inference to anticipate future sensory inputs (Friston, 2005; Gregory, 1980; Von Helmholtz, 1867). Related theories regard the brain as an inference engine and come with neuronal implementation schemes that accomplish probabilistic (Bayesian) inference in a neurologically plausible manner (Bastos et al., 2012; Friston, 2005, 2010). Process theories such as predictive coding assume that the brain maintains a generative model of its environment which is continuously updated by comparing incoming sensory information with model predictions on different levels of hierarchical cortical organization (Friston, 2005, 2010; Rao & Ballard, 1999; Winkler & Czigler, 2012). Differential influences of SSA and change detection on the MMN are proposed to result from the same underlying process of prediction error minimization, mediated by different post-synaptic changes to (predicted) sensory inputs (Auksztulewicz & Friston, 2016; Garrido et al., 2008; Garrido, Kilner, Kiebel, & Friston, 2009). As such, the theory has the potential to unify previously opposing theories of MMN generation (Garrido et al., 2008; Garrido, Kilner, Kiebel, & Friston, 2009; Garrido, Kilner, Stephan, & Friston, 2009) while accounting for its key empirical features (Heilbron & Chait, 2018; Wacongne et al., 2012).

With regard to the proposed universal nature of predictive accounts of brain function, reports of comparable MMRs across different modalities are of particular interest. So far, mismatch signals have been primarily studied in isolation, with a strong focus on the auditory system. However, key properties of the auditory MMN, such as omission responses and modulations by predictability, have also been reported for the visual (Czigler et al., 2006; Kok et al., 2014) and the somatosensory MMN (Andersen & Lundqvist, 2019; Naeije et al., 2018), and modeling studies in all three modalities suggest that

MMRs may reflect signatures of Bayesian learning (BL; Gijsen et al., 2021; Lieder et al., 2013; Maheu et al., 2019; Ostwald et al., 2012; Stefanics et al., 2018). While studies directly investigating mismatch signals in response to multimodal sensory inputs are rare, previous research indicates a ubiquitous role for cross-modal probabilistic learning. The brain tends to automatically integrate auditory, somatosensory, and visual stimuli during sequence processing (Bresciani et al., 2006, 2008; Frost et al., 2015) and cross-modal perceptual associations can influence statistical learning of sequence regularities (Andric et al., 2017; Parmentier et al., 2011), modulate MMRs (Besle et al., 2005; Butler et al., 2012; Friedel et al., 2020; Kiat, 2018; Zhao et al., 2015) and influence subsequent unimodal processing in various ways (Shams et al., 2011). Recent advances in modeling Bayesian causal inference suggest that the main computational stages of multimodal inference evolve along a multisensory hierarchy involving early sensory segregation followed by mid-latency sensory fusion and late Bayesian causal inference (Cao et al., 2019; Rohe et al., 2019; Rohe & Noppeney, 2015). However, the extent to which the MMN and P3 reflect these stages and should be considered sensory specific signatures of regularity violation or the result of modality independent computations in an underlying predictive network is not fully understood.

The current study aimed to investigate the commonalities and differences between MMRs in different modalities in a single experiment and to elucidate in how far they reflect local, unimodal or global, cross-modal computations. To this end, we employed a roving stimulus paradigm, in which auditory, somatosensory, and visual stimuli were simultaneously presented in a probabilistic tri-modal stimulus stream.

Typically, MMRs are studied with the oddball paradigm, in which rarely presented "oddball" stimuli deviate from frequently presented standard stimuli in some physical feature, such as sound pitch or stimulus intensity. The roving stimulus paradigm, on the other hand, defines deviants and standards in terms of their local sequence position, while the frequency of occurrence of their stimulus features across the sequence is equal (Baldeweg et al., 2004; Cowan et al., 1993). The deviant is defined as the first stimulus that breaks a train of repeating (standard) stimuli. With repetition, the deviant subsequently becomes the new standard, defining a train of stimulus repetitions. Thus, the roving stimulus paradigm is an excellent tool to experimentally induce MMRs, while controlling for differences in physical stimulus features.

Based on a probabilistic model, we generated sequences of high and low intensity stimuli that were governed by unimodal transition probabilities as well as cross-modal conditional dependencies. This allowed us to test to what extent early and late MMRs are sensitive to local and global violations of statistical regularities and to draw conclusions regarding their potential role in cross-modal hierarchical inference. Specifically, we extracted the MMN and P3 MMRs for each modality and investigated their modality specific and modality general response properties regarding stimulus repetition and change, as well as their sensitivity to cross-modal predictive information. Further, we used source localization to investigate modality specific and modality general neuronal generators of MMRs. Finally, we complemented our average-based analyses with single-trial modeling to investigate if signatures of unimodal and cross-modal Bayesian inference can account for trial-to-trial fluctuations in the MMN and P3 amplitudes.

2 | MATERIALS AND METHODS

Participants underwent a novel multimodal version of the roving stimulus paradigm. Our paradigm, depicted in Figure 1, consisted of simultaneously presented auditory (A), somatosensory (S), and visual (V) stimuli, which each alternated between two different intensity levels ("low" and "high"). The tri-modal stimulus sequences originated from a single probabilistic model (described in Section 2.3), resulting in different combinations of low and high stimuli across the three modalities in each trial.

2.1 | Participants

Thirty-four healthy volunteers (19–43 years old, mean age: 26, 22 females, all right-handed), recruited from the student body of the Freie Universität Berlin and the general public, participated for monetary compensation or an equivalent in course credit. The study was approved by the local ethics committee of the Freie Universität Berlin and written informed consent was obtained from all participants prior to the experiment.

2.2 | Experimental setup

Each trial consisted of three bilateral stimuli (A, S, and V) that were presented simultaneously by triggering three instantaneous outputs of a data acquisition card (National Instruments Corporation, Austin, Texas, USA) every 1150 ms (inter-stimulus interval).

Auditory stimuli were presented via in-ear headphones (JBL, Los Angeles, California, USA) to both ears and consisted of sinusoidal waves of 500 Hz and 100 ms duration that were modulated by two different amplitudes. The amplitudes were individually adjusted with the participants to obtain two clearly distinguishable intensities (mean of the low intensity stimulus: 81.43 ± 1.22 dB; mean of the high intensity stimulus: 93.02 ± 0.98 dB).

Somatosensory stimuli were administered with two DS5 isolated bipolar constant current stimulators (Digitimer Limited, Welwyn Garden City, Hertfordshire, UK) via adhesive electrodes (GVB-geliMED GmbH, Bad Segeberg, Germany) attached to the wrists of both arms. The stimuli consisted of electrical rectangular pulses of 0.2 ms duration. To account for interpersonal differences in sensory thresholds, the two intensity levels used in the experiment were determined on an individual basis. The low intensity level (mean: 3.97 ± 0.84mA) was set in proximity to the detection threshold yet high enough to be clearly perceivable (and judged to be the same intensity on both sides). The high intensity level (mean: 6.47 ± 1.33mA) was determined

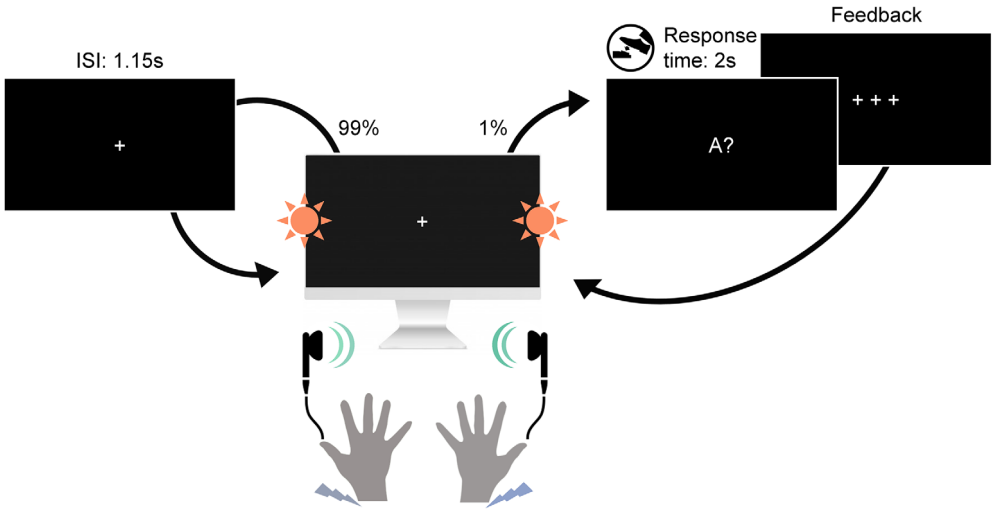


FIGURE 1 Experimental paradigm. Participants were seated in front of a screen and received sequences of simultaneously presented bilateral auditory beep stimuli (green), somatosensory electrical pulse stimuli (purple) and visual flash stimuli (orange) each at either low or high intensity. On consecutive trials, stimuli within each modality either repeated the previous stimulus intensity of that modality (standard) or alternated to the other intensity (deviant). This created tri-modal roving stimulus sequences, where the repetition/alternation probability in each modality was determined by a single probabilistic model (see Section 2.3). In 1% of trials (catch trials) the fixation cross changed to one of the three letters A, T, or V, interrupting the stimulus sequence. The letter prompted participants to indicate whether the last auditory (letter A), somatosensory (letter T for "tactile") or visual (letter V) stimulus, respectively, was of high or low intensity. Responses were given with a left or right foot pedal press using the right foot.

for each participant to be easily distinguishable from the low intensity level yet remaining non-painful and below the motor threshold.

Visual stimuli were presented via light emitting diodes (LEDs) and transmitted through optical fiber cables mounted vertically centered to both sides of a monitor. The visual flashes consisted of rectangular waves of 100 ms duration that were modulated by two different amplitudes (low intensity stimulus: 2.65 V; high intensity stimulus: 10 V) that were determined to be clearly perceivable and distinguishable prior to the experiment. Participants were seated at a distance of about 60 cm to the screen such that the LED's were placed within the visual field at a visual angle of about 67°.

In each of six experimental runs of 11.5 min, a sequence of 600 stimulus combinations was presented. To ensure that participants maintained attention throughout the experiment and to encourage monitoring of all three stimulation modalities, participants were instructed to respond to occasional catch trials (target questions) via foot pedals. In six trials randomly placed within each run the fixation cross changed to one of the letters A, T, or V followed by a question mark. This prompted participants to report if the most recent stimulus (directly before appearance of the letter) in the auditory (letter A), somatosensory (letter T for "tactile"), or visual (letter V) modality was presented with low or high intensity. The right foot was used to press either a left or a right pedal, and the pedal assignment (left = low/right = high or left = high/right = low) was counterbalanced across participants.

It should be noted that our MMR paradigm in form of an attended roving stimulus sequence with relatively long ISI (1.15 s) differs from the classic oddball protocol for MMN elicitation in which participants are engaged in a primary task, often attending a separate modality. Since this is not easily possible with our paradigm (containing auditory, somatosensory, and visual stimuli), we used catch-trials in each modality to ensure that attentional resources were distributed largely equally across the simultaneous stimulus streams. The ISI at the upper end of the range used for MMN elicitation was set during piloting such that the perceptually demanding tri-modal bilateral stimulation was deemed not to be overwhelming in terms of sensory overload.

2.3 | Probabilistic sequence generation

Each of the three sensory modalities (A, S, V) were presented as binary (low/high) stimulus sequences originating from a common probabilistic model. The model consists of a state s at time t evolving according to a Markov chain $(p(s_t|s_{t-1}))$ with each state deterministically emitting a combination of three binary observations conditional on the preceding observation combination $(p(o_{A,t},o_{S,t},o_{V,t}|o_{A,t-1},o_{S,t-1},o_{V,t-1}))$. For example, a transition expressed as [100|000] indicates a unimodal auditory change $(o_{A,t}=1,o_{A,t-1}=0)$ with repeating somatosensory and visual modalities $(o_{S,t}=o_{S,t-1}=0$ and $o_{V,t}=o_{V,t-1}=0)$. For each stimulus modality, in each state, the other two modalities form either congruent observations ([00] and [11]), or incongruent observations ([01] and [10]), which was used to manipulate the predictability of transitions in the sequences in different runs of the experiment.

Three types of stimulus sequences, depicted in Figure 2 were generated with different probability settings. The settings determine the transition probabilities within each modality given the arrangement of the other two modalities (i.e., either congruent or

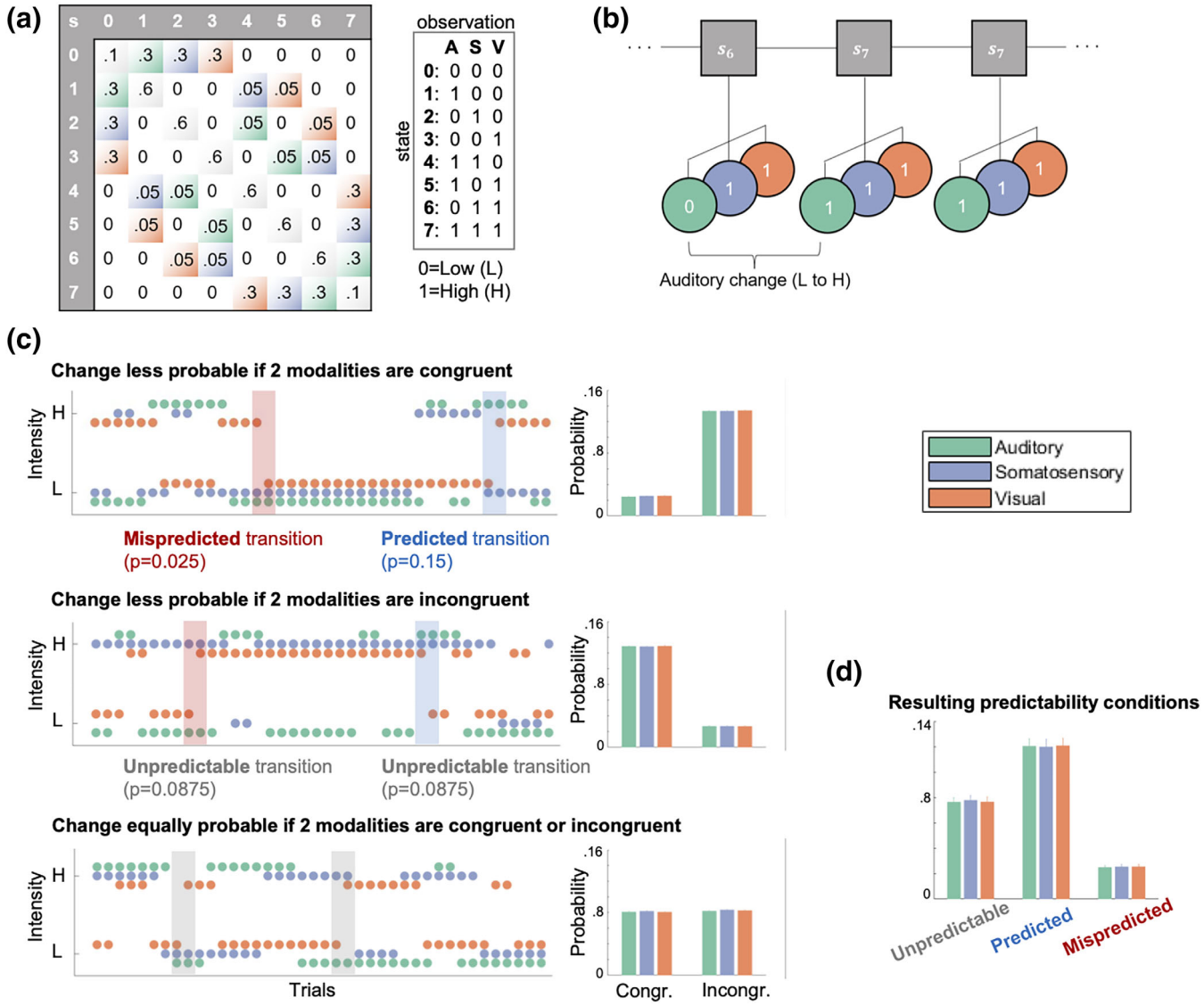


FIGURE 2 Probabilistic sequence generation. (a) Schematic of state transition matrix (left). Light color shading depicts transitions in the respective modality which were assigned specific transition probabilities: Green = auditory change, purple = somatosensory change, orange = visual change, gray diagonal = tri-modal repetition, white = multimodal change (set to zero). States 0–7 correspond to a specific stimulus combination (right), that is, eight permutations of low (0) and high (1) stimuli across three modalities (A = auditory; S = somatosensory; V = visual) as described in the main text. (b) Visualization of states (s) evolving according to a Markov chain emitting tri-modal binary outcomes. (c) Probability settings of stimulus sequences. Left column: Sequences. Right column: Averaged empirical change probabilities across all sequences. Top: Transition probabilities determine that for each modality a change is unlikely ($p = .025$) if the other two modalities are congruent (and likely if they are incongruent; $p = .15$). Middle: Transition probabilities determine that for each modality a change is likely ($p = .15$) if the other two modalities are congruent (and unlikely if they are incongruent; $p = .025$). Bottom: Transition probabilities determine that for each modality a change is equally likely ($p = .0875$) if the other two modalities are congruent or incongruent. (d) Averaged empirical change probabilities for predictability conditions.

incongruent). One setting defines lower change probability if the other two modalities are congruent (e.g., for any change in modality A from $t-1$ to t, S and V were *congruent* with $p(100|000) = p(000|100) = p(111|011) = p(011|111) = 0.025$ and S and V were *incongruent* with $p(101|001) = p(001|101) = p(110|010) = p(010|110) = 0.15$). The second setting defines lower change probability if the other two modalities are incongruent (e.g., for any change in modality A from $t-1$ to t, S and V were *incongruent* with $p(101|001) = p(001|101) = p(110|010) = p(010|110) = 0.025$ and S and V were *congruent* with $p(100|000) = p(000|100) = p(111|011) = p(011|111) = 0.15$). The third setting defines equal change probability if the other two modalities are congruent or incongruent (e.g., for any change in modality A from $t-1$ to t, S and V were *congruent* with $p(100|000) = p(000|100) = p(111|011) = p(011|111) = 0.0875$ and S and V were *incongruent* with $p(101|001) = p(001|101) = p(110|010) = p(010|110) = 0.0875$).

In each of six experimental runs, the stimulus sequence was defined by one of the three different probability settings. Each

probability setting was used twice during the experiment and the order of the six different sequences was randomized. Participants were unaware of the sequence probabilities and any learning of sequence probabilities was considered to be implicit and task irrelevant.

Following the nomenclature suggested by Arnal and Giraud (2012), the resulting stimulus transitions for each modality within the different sequences can be defined as being either *predicted* (here higher change probability conditional on congruency/incongruency), *mispredicted* (here lower change probability conditional on congruency/incongruency) or *unpredictable* (here equal change probability). For each modality, repetitions are more likely $(p = .825)$ than changes $(p = .175)$ regardless of the type of probability setting and stimulus, resulting in classic roving standard sequences for each modality (mean stimulus train length: 5, mean range of train length: 2–34 stimuli).

2.4 | EEG data collection and preprocessing

Data were collected using a 64-channel active electrode EEG system (ActiveTwo, BioSemi, Amsterdam, Netherlands) at a sampling rate of 2048 Hz, with head electrodes placed in accordance with the extended 10–20 system. Individual electrode positions were recorded using an electrode positioning system (zebris Medical GmbH, Isny, Germany).

Preprocessing of the EEG data was performed using SPM12 (Wellcome Trust Centre for Neuroimaging, Institute for Neurology, University College London, London, UK) and in-house MATLAB scripts (MathWorks, Natick, MA). First, the data were referenced against the average reference, high-pass filtered (0.01 Hz), and downsampled to 512 Hz. Subsequently, eye-blinks were corrected using a topographical confound approach (Berg & Scherg, 1994; Ille et al., 2002). The Data were epoched using a peri-stimulus time interval of -100 to 1050 ms and all trials were visually inspected and artifactual data removed. Likewise, catch trials were omitted for all further analyses. Furthermore, the EEG data of two consecutive participants were found to contain excessive noise due to hardware issues, resulting in their exclusion from further analyses and leaving data of 32 participants. Finally, a low-pass filter was applied (45 Hz) and the preprocessed EEG data were baseline corrected with respect to the pre-stimulus interval of -100 to -5 ms. To use the general linear model (GLM) implementation of SPM, the electrode data of each participant were linearly interpolated into a 32×32 grid for each time point, resulting in one three-dimensional image (with dimensions $32 \times 32 \times 590$) per trial. These images were then spatially smoothed with a 12×12mm full-width half-maximum Gaussian kernel to meet the requirements of random field theory, which the SPM software uses to control the family wise error rate.

2.5 | Event-related responses and statistical analysis

First, to extract basic MMR signals of each modality from the EEG data, we contrasted standard and deviant trials of each modality with paired t-tests corrected for multiple comparisons by using cluster-based permutation tests implemented in fieldtrip (Maris & Oostenveld, 2007). Two time windows of interest were defined based on the literature (Duncan et al., 2009) to search for earlier negative clusters between 50 and 300 ms, corresponding to the MMN, and later positive clusters between 200 and 600 ms, corresponding to the P3. Clusters were defined as adjacent electrodes with a cluster defining threshold of $p_{fwe} < .05$.

For further analyses, GLMs were set up as implemented in SPM12, which allows defining conditions on the single trial level. To test for effects of stimulus repetitions on standards, deviants, and MMRs (deviants minus standards), a *TrainLength* model was defined that consisted of 45 regressors: an intercept regressor, 36 regressors coding for the repetition train length (trials binned into 1, 2, 3, 4–5, 6–8, >8 repetitions) for standards (i.e., the position of the standard in the current train) and deviants (i.e., the number of standards preceding the deviant) in each modality, as well as four global standard and four global deviant regressors. The binning was chosen such that trials were roughly balanced across the bins, corresponding to, on average, around 100 deviant trials and around 500 standard trials in each category for each modality The global regressors captured the train length (1, 2, 3, >3 repetitions) of standards and deviants regardless of their modality, meaning that trials in which standards occurred in all three modalities were coded as global standards, whereas trials in which a deviant occurred in any of the three modalities were coded as global deviants.

To test for the implicit effect of cross-modal predictability based on the different conditional probability setting in the sequence, a *Predictability* model was defined that consisted of 37 regressors: an intercept regressor and 18 regressors coding standards and deviants of each modality for each of the three conditions described above: *unpredictable* (trials originate from sequences with no conditional dependence between modalities), *predicted* (trials originate from sequences with conditional dependence; trials defined by change being likely), *mispredicted* (trials originate from sequences with conditional dependence; trials defined by change being unlikely). On the single-participant level, these were coded for congruent and incongruent trials separately resulting in 36 regressors. By definition, the number of trials in regressors with *mispredicted* trials was lowest, on average corresponding to around 60 deviant trials and 800 standard trials per modality.

Finally, a *P3-Conjunction* model was specified that consisted of seven regressors: an intercept regressor and six regressors coding all standards and deviants for each of the three modalities. This model was used to apply SPM's second level conjunction analysis, contrasting standards and deviants across modalities in search of common P3 effects across modalities.

Each GLM was estimated on the single-trial data of each participant using restricted maximum likelihood estimation. This yielded β-parameter estimates for each model regressor over (scalp-) space and time, which were subsequently analyzed at the group level. Second level analyses consisted of a mass-univariate multiple regression analysis of the individual β scalp-time images with a design matrix specifying regressors for each condition of interest as well as a subject

factor. Second level beta estimates were contrasted for statistical inference and multiple comparison correction was achieved with SPM's random field theory-based FWE correction (Kilner et al., 2005).

2.6 | Source localization

To investigate the most likely underlying neuronal sources for the MMN and P3 MMR we applied distributed source reconstruction as implemented in SPM12 to the ERP data. For each participant, the MMN of each modality (auditory, somatosensory, visual) was source localized within a time window of 100–200 ms. For the P3, the average MMR at 330 ms was chosen for source localization as this time point showed the strongest overlap of P3 responses between modalities (based on the results of the P3 conjunction contrast).

Participant-specific forward models were created using an 8196-vertex template cortical mesh co-registered with the individual electrode positions via fiducial markers. An EEG head model based on the boundary element method was used to construct the forward model's lead field. For the participant-specific source estimates, multiple sparse priors under group constraints were applied. The source estimates were subsequently analyzed at the group level using the GLM implemented in SPM12. Second-level contrasts consisted of one-sample t tests for each modality as well as (global) conjunction contrasts across modalities. The resulting statistical parametric maps were thresholded at the peak level with $p < .05$ after FWE correction. The anatomical correspondence of the MNI coordinates of the cluster peaks were identified via cytoarchitectonic references using the SPM Anatomy toolbox (Eickhoff et al., 2005).

2.7 | Single-trial modeling of EEG data

In addition to the analysis of event-related potentials, the study aimed to compare different computational strategies of sequence processing potentially underlying neuronal generation of MMRs. To this end, we generated regressors from different BL models as well as a train length dependent change detection (TLCD) model making different predictions for the single-trial EEG data.

Theories on MMN generation hypothesize adaptation and memory-trace dependent change detection to contribute to the MMN. With prior repetition of stimuli, the response to standard stimuli tends to decrease while the response to deviant stimuli tends to increase. We defined the TLCD model to reflect such reciprocal dynamics of responses to stimulus repetition and change without invoking assumptions of probabilistic inference. The model is defined for each modality separately and tracks the stimulus train lengths c for a given modality by counting stimulus repetitions: $c_t = d_t(c_{t-1} + d_t)$, where $d_t = 1_{o_t = o_{t-1}}$ takes on the value 1 whenever the current observation o_t is a repetition of the previous observation o_{t-1} and $d_t = 0$ resets the current train length to zero. To form single-trial predictors of the EEG data, the model outputs values that increase linearly with train length and have opposite signs for standards and deviants:

$$\mathrm{TLCD}(o_t) = \begin{cases} -ct & \text{if } d_t = 1 \\ c_{t-1} & \text{if } d_t = 0 \end{cases}$$

In addition to the TLCD model, different BL models were created to contrast the static train length based TLCD model with dynamic generative models tracking transition probabilities. The BL models consist of conjugate Dirichlet-Categorical models estimating probabilities of observations read out by three different surprise functions: Bayesian surprise (BS), predictive surprise (PS), and confidence-corrected surprise (CS).

BS quantifies the degree to which an observer adapts their generative model to incorporate new observations (Baldi & Itti, 2010; Itti & Baldi, 2009) and is defined as the Kullback–Leibler (KL) divergence between the belief distribution prior and posterior to the update: $BS(y_t) = KL\Big(p(s_{t-1}|y_{t-1},\ldots,y_1)\Big\|p(s_t|y_t,\ldots,y_1)\Big)$. PS is based on (Shannon, 1948) definition of information and defined as the negative logarithm of the posterior predictive distribution, assigning high surprise to observed events y_t with low estimated probability of occurrence: $PS(y_t) = -ln p(y_t|s_t) = -ln\, p(y_t|y_{t-1},\ldots,y_1)$. CS additionally considers the commitment of the generative model and scales with the negative entropy of the prior distribution (Faraji et al., 2018). It is defined as the KL divergence between the (informed) prior distribution at the current time step and a flat prior distribution $\widehat{p}(s_t)$ updated with the most recent event y_t: $CS(y_t) = KL(p(s_t)|\widehat{p}(s_t|y_t))$.

Following Faraji et al. (2018) surprise quantifications can be categorized as puzzlement or enlightenment surprise. While puzzlement refers to the mismatch between sensory input and internal model belief, closely related to the concept of prediction error, enlightenment refers to the update of beliefs to incorporate new sensory input. In the current study, we were interested in a quantification of the model inadequacy by means of an unsigned prediction error as reflected by surprise. As such, throughout the manuscript, with prediction error we do not refer to the specific term of (signed) reward prediction error as used for example in reinforcement learning but rather use it to refer to the signaling of prediction mismatch. While both PS and CS are instances of puzzlement surprise, CS is additionally scaled by belief commitment and quantifies the concept that a low-probability event is more surprising if commitment to the belief (of this estimate) is high. BS, on the other hand, is an instance of enlightenment surprise and is considered a measure of the update to the generative model resulting from new incoming observations.

A detailed description of the Bayesian observer, its transition probability version as well as the surprise read-out functions can be found in our previous work on somatosensory MMRs (Gijsen et al., 2021). Here, we will primarily provide a brief description of the specifics of two implementations of Dirichlet-Categorical observer models, a unimodal and a cross-modal model. Both observer models receive stimulus sequences (of one respective modality) as input and iteratively update a set of parameters with each new incoming observation. In each iteration, the estimated parameters are read out by the surprise functions (BS, PS, and CS) to produce an output which is subsequently used as a predictor for the EEG data.

For each modality, the unimodal Dirichlet-Categorical model considers a binary sequence with two possible stimulus identities (low and high) estimating transition probabilities with $y_t = o_t$ for $t = 1,...,T$ with a set of hidden parameters $s^{(i)}$ for each possible transition from $o_{t-1} = i$. This unimodal model does not capture any cross-modal dependencies in the sequence (i.e., the alternation and repetition probabilities conditional on the tri-modal stimulus configuration). Therefore, we defined a cross-modal Dirichlet-Categorical model to address the question whether the conditional dependencies were used by the brain during sequence processing for prediction of stimulus change. The dependencies in the sequence were independent of stimulus identity but provide information about the probability of repetition or alternation (d_t) conditional on the congruency of the other modalities. The cross-modal model thus estimates alternation probabilities $(y_t = d_t$ for $t = 2,...,T)$ with a set of hidden parameters $s^{(i)}$ when other modalities are incongruent and $s^{(c)}$ when other modalities are congruent. Therefore, while the unimodal model learns the probability of stimulus *transitions* within modality, the cross-modal model learns the probability of stimulus *alternations* within modality conditional on the congruency of the other modalities. As such, the cross-modal model provides a minimal implementation of a Bayesian observer that captures the cross-modal dependencies in the sequences.

2.7.1 | Model fitting procedure

The technical details of the model fitting and subsequent Bayesian model selection (BMS) procedures are identical to Gijsen et al. (2021) where the interested reader is kindly referred to for further information. First, the stimulus sequence-specific regressor of each model was obtained for each participant. After z-score normalization, the regressors were fitted to the single-trial, event-related electrode data using a free-form variational inference algorithm for multiple linear regression (Flandin & Penny, 2007; Penny et al., 2003; Penny et al., 2005). The obtained model-evidence maps were subsequently subjected to the BMS procedure implemented in SPM12 (Stephan et al., 2009) to draw inferences across participants with well-established handling of the accuracy-complexity trade-off (Woolrich, 2012).

In total, eight regression models were fit: A null model (offset only), a TLCD regression model and, for each of the three surprise read-out functions, one regression model including only the unimodal regressors and one additionally including the cross-modal regressors. The purely unimodal regression model will be called UM and the regression model including unimodal and, additionally, cross-modal regressors will be called UCM. The design matrix of the TLCD regression model consisted of four regressors, an offset and the predicted parametric change responses for each of the three modality sequences (auditory, somatosensory, visual). Similarly, the design matrix of the UM regression model consisted of four regressors, an offset and the surprise responses of the unimodal model for each of the three modalities. The UCM regression model was identical to the UM regression model but with an additional three regressors containing the cross-modal surprise responses for each modality. Therefore, the UCM regression model is more complex and gets assigned higher model evidence than the reduced UM regression model only if the additional regressors contribute significantly to a better model fit (Stephan et al., 2009).

To allow for the possibility of different timescales of stimulus integration (Maheu et al., 2019; Ossmy et al., 2013; Runyan et al., 2017), the integration parameter τ of the Dirichlet-Categorical model was optimised for each model, participant and peri-stimulus time-bin before model selection. To this end, model regressors were fit for a range of 11 tau parameter configurations ([0, 0.001, 0.0015, 0.002, 0.003, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2]) corresponding to integration windows with a 0.5 stimulus weighting at (half-life of) [600, 462, 346, 231, 138, 69, 34, 13, 6, 3] stimuli, of which the parameter with the best model evidence was chosen.

2.7.2 | Bayesian model comparison

The estimated model-evidence maps were used to evaluate the models' relative performance across participants via family-wise BMS (Penny et al., 2010). The model space was partitioned into three types of families to draw inference on different aspects of the involved models. Given that the literature provides some evidence for each of the three surprise read-out functions (BS, PS, CS) to capture some aspect of EEG MMRs, we included all of them in the family wise comparisons to avoid biasing the comparison of different BL models.

The first model comparison considered the full space of BL models as a single family (BL family) and compared it to the TLCD model (TLCD family) and the null model (NULL family). Since the BL models had their tau parameter optimized, which was not possible for the TLCD model, we applied the same penalization method used in our previous study (Gijsen et al., 2021). The degree to which the optimization on average inflated model evidence was subtracted from the BL models prior to BMS. Specifically, for all parameter values, the difference between the average model evidence and that of the optimized parameter was computed and averaged across poststimulus time bins, electrodes and participants.

Subsequent analyses grouped the different BL models into separate families: The second comparison grouped the BL models into two families of UM and UCM models, as well as the null model, to test which electrodes and time points showed influences of unimodal versus cross-modal processing. The third comparison grouped the BL models into three surprise families and the null model, to test whether the observed MMRs were best captured by PS, BS, or CS.

3 | RESULTS

3.1 | Behavioral results

Participants showed consistent performance in responding to the catch trials during each experimental run, indicating their ability to

globally maintain their attention to the tri-modal stimulus stream. Of the 85.5% responses made in time, 75.3% were correct with an average reaction time of 1.4 ± 0.25s.

3.2 | Event-related potentials

3.2.1 | Unimodal MMRs

Cluster-based permutation tests confirmed the presence of early modality specific MMN components as well as later P3 MMRs for all three modalities. Both early and late MMRs showed a modulation by the number of stimulus repetitions, the details of which will be described in the following sections.

3.2.2 | Auditory MMRs

The MMN, as the classic MMR, has originally been studied in the auditory modality and is commonly described as the ERP difference wave calculated by subtraction of standard trials from deviant trials (deviants-standards). This difference wave typically shows a negative deflection at fronto-central electrodes and corresponding positivity at temporo-parietal sites, ranging from around 100 to 250 ms (Naatanen et al., 1978; Naatanen et al., 2007). Correspondingly, we find a significant negative fronto-central auditory MMN cluster between 80 and 200 ms (Figure 3a). Within the MMN cluster, deviants appear to deflect from the standard ERP around the peak of the auditory N1 component and reach their maximum difference around the peak of the subsequent P2 component. In the later time window, we observe positive MMRs at central electrodes between 200 and 400 ms, corresponding to a P3 modulation, as well as beyond 400 ms at progressively more posterior electrodes.

Within early and late auditory MMR clusters, the response to both standards and deviants was modulated by the number of standard repetitions. The auditory system is known to be sensitive to stimulus repetitions, particularly within the roving standard paradigm (Baldeweg et al., 2004; Cowan et al., 1993; Ulanovsky et al., 2003; Ulanovsky et al., 2004). Therefore, we hypothesized a gradual increase of the auditory response to standard stimuli around the time of the MMN, known as repetition positivity (Baldeweg, 2006; Baldeweg et al., 2004; Haenschel et al., 2005) as well as reciprocal negative modulation of the corresponding deviant response (Bendixen et al., 2007; Naatanen et al., 2007). Together, these effects should result in a gradual increase of the MMN amplitude with stimulus repetition. Indeed, linear contrasts applied to the GLM beta parameter estimates of the *TrainLength* model revealed that the MMN increases with the repetition of standards before a deviant was presented (94–200 ms, cluster $p_{fwe} < .001$). This effect was driven by a negative linear modulation of the deviant response (98–200 ms, cluster $p_{fwe} < .001$) as well as a repetition positivity effect on the standards (111–172 ms, cluster $p_{fwe} < .001$). Similarly, the later P3 MMR increased with standard repetitions (200–600 ms, cluster $p_{fwe} < .001$) and this effect was driven by an increase of deviant responses (200–600 ms, cluster $p_{fwe} < .001$) and a decrease of standard responses (205–359 ms, cluster $p_{fwe} < .001$). Given the temporal difference between standard (around 200–350 ms) and deviant (200–600 ms) train length effects, the parametric modulation of the late MMR beyond 350 ms seems to be primarily driven by the increase in deviant responses.

3.2.3 | Somatosensory MMRs

We hypothesized somatosensory MMRs to consist of early bilateral (fronto-) temporal negativities, resulting primarily from increased N140 components (Kekoni et al., 1997), with a corresponding central positivity extending into a later central P3 component.

After an early mismatch effect starting at ~50 ms at fronto-central electrodes, a more pronounced bilateral temporal cluster emerged that extended from ~90 to 190 ms and can be considered the somatosensory equivalent of the auditory MMN (Figure 3b). A reversed positive central component can be observed at the time of the somatosensory MMN (sMMN) and throughout the entire later time window (200–600 ms) at which point it can be considered a putative P3 MMR.

Early and late somatosensory MMRs were significantly modulated by stimulus repetition. Bilateral electrodes within the sMMN cluster show an increase of the sMMN amplitude with repetition (123–166 ms, cluster $p_{fwe} < .05$). This effect was driven by an increase of deviant negativity (135–188 ms, cluster $p_{fwe} < .05$) in combination with a positivization of the standard (86–188 ms, cluster $p_{fwe} < .05$). Similarly, the later P3 MMR increases with repetition of standards (200–600 ms, cluster $p_{fwe} < .05$), mutually driven by increasing deviant responses (221–600 ms, cluster $p_{fwe} < .05$) and decreasing standard responses (200–600 ms, cluster $p_{fwe} < .05$).

3.2.4 | Visual MMRs

We hypothesized visual MMRs to present as an early MMN at occipital to parieto-temporal electrodes and a later P3 component at central electrodes. Although less pronounced than its auditory and somatosensory counterparts, we indeed observed a negative visual mismatch component that developed from occipital to parieto-temporal electrodes between ~130 and 200 ms (Figure 3c). In the later time window, we found a central positive component between ~300 and 600 ms, corresponding to a P3 MMR.

Within the significant visual MMN (vMMN) cluster, the linear contrast testing for repetition effects did not reach significance when correcting clusters for multiple comparisons ($p_{fwe} > .05$). However, it is worth noting that some electrodes in this cluster seemed to show a similar pattern of response increases and decreases as in the auditory and somatosensory modality, which became apparent at more lenient thresholds. The vMMN tended to become more negative with repetition of standards (143–148 ms, peak $p_{uncorr} < .005$), with opposite tendencies of deviant negative increase (143–148 ms, peak $p_{uncorr} < .05$)

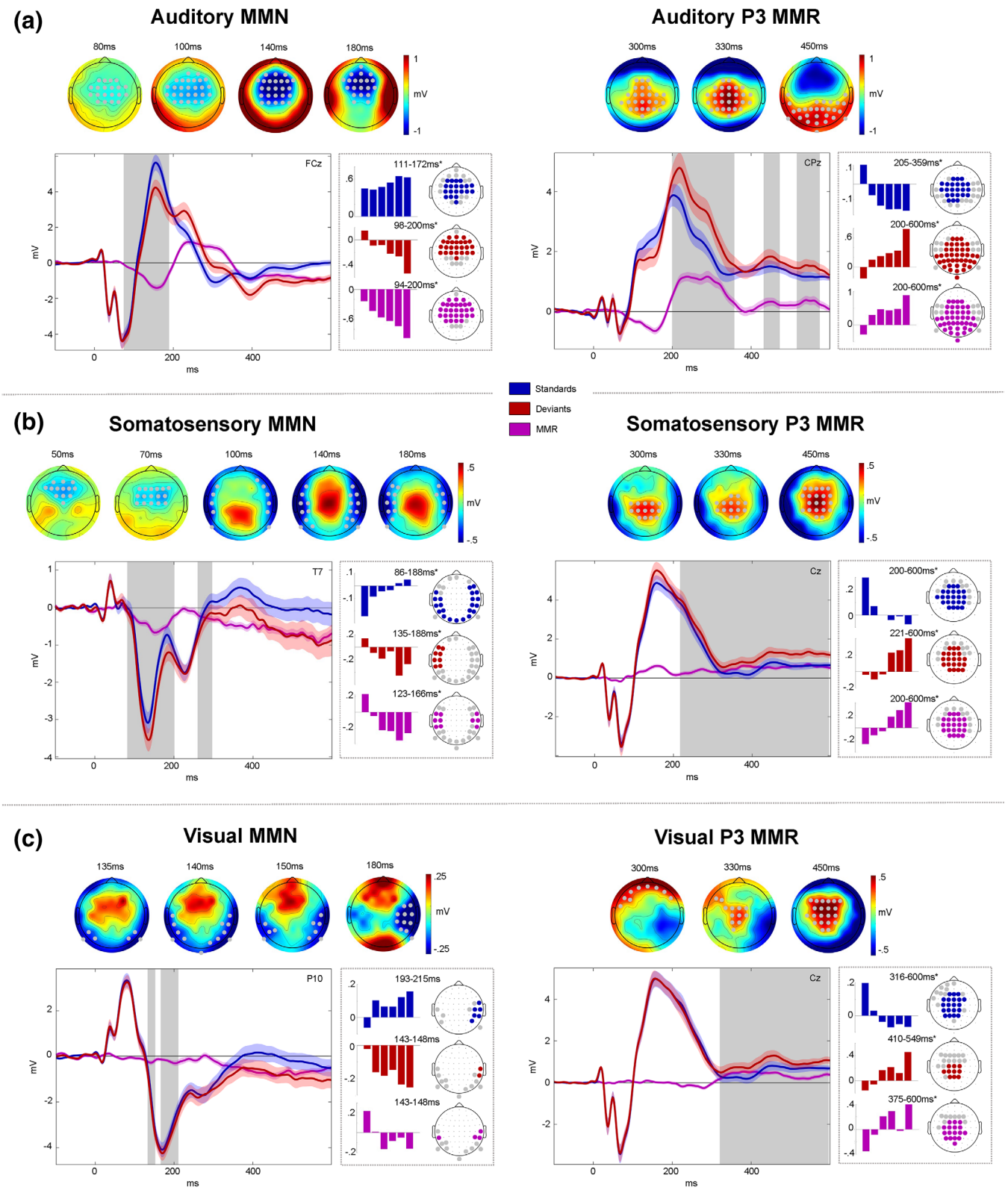


FIGURE 3 Mismatch responses. Panels (a–c) show mismatch responses (MMRs) of auditory (a), somatosensory (b), and visual (c) modalities. Within panels: Left: mismatch negativity (MMN). Right: P3 MMR. Gray dots (top) and gray boxes (bottom) indicate significant MMR electrodes and time points with $p_{fwe} < .05$. Top row: MMR scalp topographies (deviants-standards). Bottom row: Grand average ERPs (left panels) and beta parameter estimates of significant linear contrast clusters (right panels). Colored bars depict six beta parameter estimates of the *TrainLength* GLM (1, 2, 3, 4–5, 6–8, >8 repetitions) averaged across electrodes within linear contrast clusters. Asterisks indicate significance of the linear contrast ($p_{fwe} < .05$).

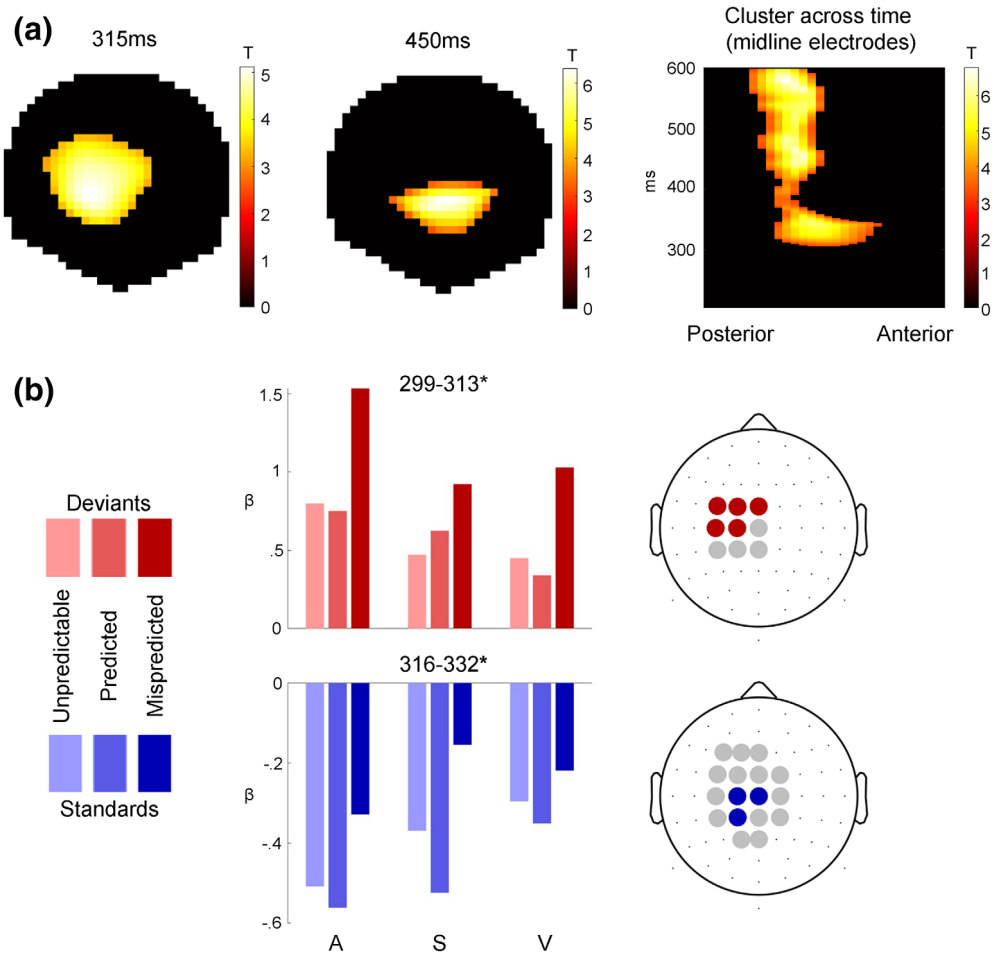


FIGURE 4 Cross-modal P3 effects. (a) T-Maps of the conjunction of deviant > standard contrasts across the auditory, somatosensory, and visual modalities. (b) Beta estimates averaged across electrodes within significant clusters with peak p_{fwe} <0.05, resulting from two-way ANOVAs testing for differences between *unpredictable*, *predicted*, and *mispredicted* deviants (red) and standards (blue).

and standard decrease (193–215 ms, peak p_{fwe} < .05). Thus, although we cannot conclude a modulation by standard repetition of the vMMN with any certainty, the observed beta parameters are in principle compatible with the effects observed in the auditory and somatosensory modalities (please see the discussion for potential reasons for the reduced vMMN in our data).

Within the P3 MMR cluster, on the other hand, we find significant clusters of linear increase of the MMR (375–600 ms, cluster p_{fwe} < .05), again constituted by an increase in deviant responses (410–549 ms, cluster p_{fwe} < .05) and concomitant decrease in standard responses (316–600 ms, cluster p_{fwe} < .05).

3.2.5 | Cross-modal P3 effects

In search of a common P3 effect to deviant stimuli, we created conjunctions of the *deviants* > *standards* contrasts across the auditory, somatosensory, and visual modalities. The conjunction revealed a common significant cluster starting at ~300 ms (cluster p_{fwe} < .05) that comprised anterior central effects around 300–350 ms followed by more posterior effects from 400 to 600 ms (Figure 4a).

To investigate the modulation of the P3 MMR by predictability, we used two-way ANOVAs with the three-level factor *modality* (*auditory*, *somatosensory*, *visual*) and the three-level factor *predictability* condition (*predicted*, *mispredicted*, *unpredictable*). Separate ANOVAs were applied to deviants and standards. We hypothesized that the cross-modal P3 MMR might be sensitive to multisensory predictive information in the sequence, as the P3 has been shown to be sensitive to global sequence statistics (Bekinschtein et al., 2009; Wacongne et al., 2011) and to be modulated by stimulus predictability (Horvath et al., 2008; Horvath & Bendixen, 2012; Max et al., 2015; Prete et al., 2022; Ritter et al., 1999; Sussman et al., 2003). Indeed, within the common P3 cluster, both deviants (299–313 ms, peak p_{fwe} < .05) and standards (316–332 ms, peak p_{fwe} < .05) show significant differences between predictability conditions. No significant interaction of predictability condition with modality was observed.

Post hoc *t* tests were applied to the peak beta estimates to investigate the differences between the three pairs of conditions. For the ANOVA concerning the deviant trials, post hoc *t* tests show a significant difference for *mispredicted* > *predicted* ($t = 14.667$; $p < .001$, Bonferroni corrected), *mispredicted* > *unpredictable* ($t = 14.76$; $p < .001$, Bonferroni corrected) and no significant difference between *unpredictable* > *predicted* conditions ($t = 0.01$; $p > .05$). Similarly, for the ANOVA concerning the standard trials, post hoc *t* tests show that there is a significant difference for *mispredicted* > *predicted* ($t = 10.67$;

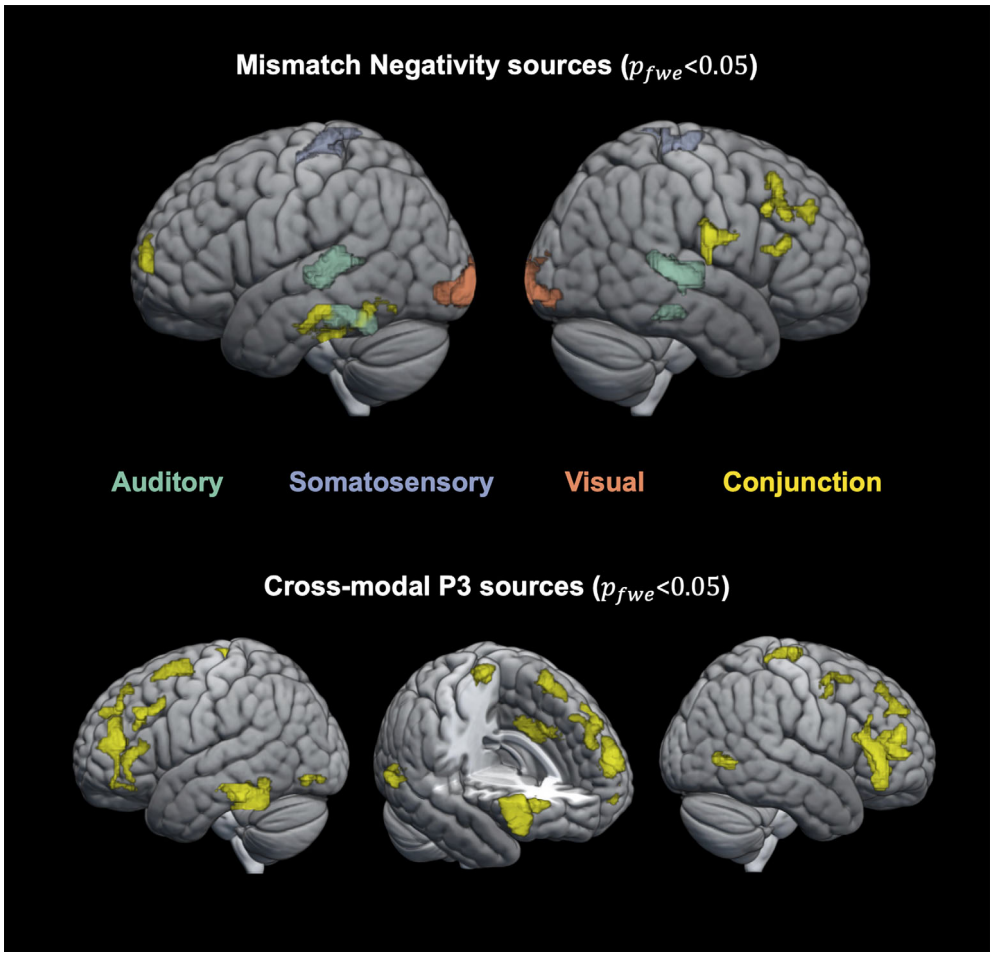


FIGURE 5 Source localization. Top row: significant sources (p_{fwe} < .05) for the auditory (green), somatosensory (purple), and visual (orange) mismatch negativity (MMN) as well as their conjunction (yellow). Bottom row: significant sources (p_{fwe} < .05) for the conjunction (yellow) of the P3 mismatch response (MMR) in the auditory, somatosensory, and visual modalities.

p < .001, Bonferroni corrected), *mispredicted* > *unpredictable* (t = 6.87; p < .001, Bonferroni corrected) and *unpredictable* > *predicted* conditions (t = 3.83; p < .001, Bonferroni corrected).

Taken together, this result suggests that stimuli which were mispredicted based on the predictive multisensory configuration resulted in increased responses within the common P3 cluster compared to predicted or unpredictable stimuli, regardless of their role as standards or deviants in the current stimulus train.

For completeness, we also tested the effect of predictability in the earlier MMN cluster, but we did not observe any significant modulations here (results not shown).

3.3 | Source localization

The source reconstruction analysis resulted in significant clusters of activation for each modality's MMN as well as the P3 MMR. The results are depicted in Figure 5 and cytoarchitectonic references are described in Table 1.

For each modality, the MMN was localized to source activations in the respective modality's sensory cortex and frontal cortex. Source localization of the auditory MMN shows the strongest activation in bilateral superior temporal areas (p_{fwe} < .05; left cluster: peak t = 6.20; right cluster: peak t = 7.64) corresponding to auditory cortex and in inferior temporal areas (p_{fwe} < .05; left cluster: peak t = 5.63; right cluster: peak t = 5.60). The somatosensory MMN shows highest source activation in postcentral gyrus (p_{fwe} < .05; left cluster: peak t = 5.22; right cluster: peak t = 4.92) corresponding to primary somatosensory cortex. Similarly, the vMMN shows highest source activation in the occipital cortex (p_{fwe} < .05; left cluster: peak t = 6.18; right cluster: peak t = 5.17), around the occipital pole, corresponding to visual areas (V1–V4). Lowering the threshold to p_{uncorr} < .001 (only shown in Table 1) suggests additional activation of hierarchically higher sensory areas such as secondary somatosensory cortex for the sMMN (p_{uncorr} < .001; left cluster: peak t = 4.21; right cluster: peak t = 5.01) and lateral occipital cortex (fusiform gyrus) for vMMN (part of the primary visual cluster). In addition to the sensory regions, common frontal sources with dominance on the right hemisphere were identified using a conjunction analysis for the MMN of all three modalities. In particular, significant common source activations were found in the right inferior frontal gyrus (IFG; p_{fwe} < .05; cluster: peak t = 3.15) and right middle frontal gyrus (MFG; p_{fwe} < .05; cluster: peak t = 2.89). Additional significant common sources include frontal pole (p_{fwe} < .05; left cluster: peak t = 2.56; right cluster: peak t = 2.28), left inferior temporal gyrus (p_{fwe} < .05; cluster: peak t = 2.52) and right inferior parietal lobe (p_{fwe} < .05; cluster: peak t = 2.85).

TABLE 1 Source localization and cytoarchitectonic reference

Contrast	Hemisphere	Cytoarchitecture (probability)	MNI coord. at cytoarch.	*t*-Statistics at cytoarch. (*p*-value)
aMMN	Left	Auditory areas:		
		TE 4 (67.3%)	−52 −26 0	6.2 (p_{fwe} < .05)
		TE 3 (15.1%)	−59.8 −17.6 5.4	4.81 (p_{fwe} < .05)
		TE 1 (50.7%)	−50.3 −19.2 5.8	4.05 (p_{uncorr} < .001)
aMMN	Right	Auditory areas:		
		TE 4 (54.1%)	−56 −26 0	7.64 (p_{fwe} < .05)
		TE 3 (33.6%)	−64.2 −16.4 5	5.85 (p_{fwe} < .05)
		TE 1 (61.9%)	53 −10.1 3.8	3.53 (p_{uncorr} < .001)
sMMN	Left	Somatosensory areas:		
		3b [S1] (31.4%)	−15.7 −33.7 68.6	4.8 (p_{fwe} < .05)
		OP4 [S2] (38.6%)	−65 −14.8 20.1	3.72 (p_{uncorr} < .001)
		OP1 [S2] (16%)		
sMMN	Right	Somatosensory areas:		
		3b [S1] (40.3%)	13.3 −33.7 68.2	4.76 (p_{fwe} < .05)
		OP4 [S2] (prob. 46.9%)	66.5 −10.6 20.9	3.51 (p_{uncorr} < .001)
		OP1 [S2] (prob. 9.1%)		
vMMN	Left	Visual areas:		
		hOc1 [V1] (84.6%)	−10 −100 0	6.18 (p_{fwe} < .05)
		hOc2 [V2] (11.1%)		
		hOc4v [V4] (51%)	−30 −87.9 -11.7	5.51 (p_{fwe} < .05)
		hOc3v [V3] (24.6%)		
		FG4 (89.5%)	−43.5 −49.4 −13.3	3.83 (p_{uncorr} < .001)
vMMN	Right	Visual areas:		
		hOc1 [V1] (86.4%)	20 −100 −4	5.17 (p_{fwe} < .05)
		hOc2 [V2] (11%)		
		hOc3 [V3] (43.7%)	34.4 −88.7 −7.7	4.7 (p_{fwe} < .05)
		hOc4 [V4] (23.8%)		
		FG4 (59.8%)	50.3 −42.5 −18.8	3.82 (p_{uncorr} < .001)
MMN conjunction	Left	Frontal pole (28%)	−14 62 8	2.56 (p_{fwe} < .05)
		Inferior temporal gyrus (49%)	−50 −24 −28	2.52 (p_{fwe} < .05)
MMN conjunction	Right	Middle frontal gyrus (36%)	42 28 30	2.89 (p_{fwe} < .05)
		Inferior frontal gyrus (53%)	53 27.6 17.3	2.96 (p_{fwe} < .05)
		Frontal pole (65%)	30 42 32	2.28 (p_{fwe} < .05)
		Inferior parietal lobe (46.8%)	62 −14 24	2.85 (p_{fwe} < .05)
P3 con-junction	(Left)	Anterior cingulate gyrus (51%)	−6 14 36	4.34 (p_{fwe} < .05)
P3 conjunction	Left	Frontal pole (74%)	−34 44 20	3.31 (p_{fwe} < .05)
		Inferior frontal gyrus (35%)	52 26 18	2.74 (p_{fwe} < .05)
		Middle frontal gyrus (78%)	−44 30 32	2.87 (p_{fwe} < .05)
		Superior frontal gyrus (43%)	−22 12 60	3.13 (p_{fwe} < .05)
		Inferior temporal gyrus (34%)	−48 −44 −24	3.21 (p_{fwe} < .05)
		Lateral occipital [hOc4la] (81%)	−46 −80 −8	2.99 (p_{fwe} < .05)
P3 conjunction	Right	Frontal pole (86%)	45 42 10	3.57 (p_{fwe} < .05)
		Inferior frontal gyrus (39.5%)	52 26 18	3.45 (p_{fwe} < .05)
		Middle frontal gyrus (38%)	36 2 54	3.0 (p_{fwe} < .05)
		Precentral gyrus [4a] (19%)	6 −32 64	2.64 (p_{fwe} < .05)
		Lateral occipital cortex [hOc5] (55%)	56 −62 0	3.45 (p_{fwe} < .05)

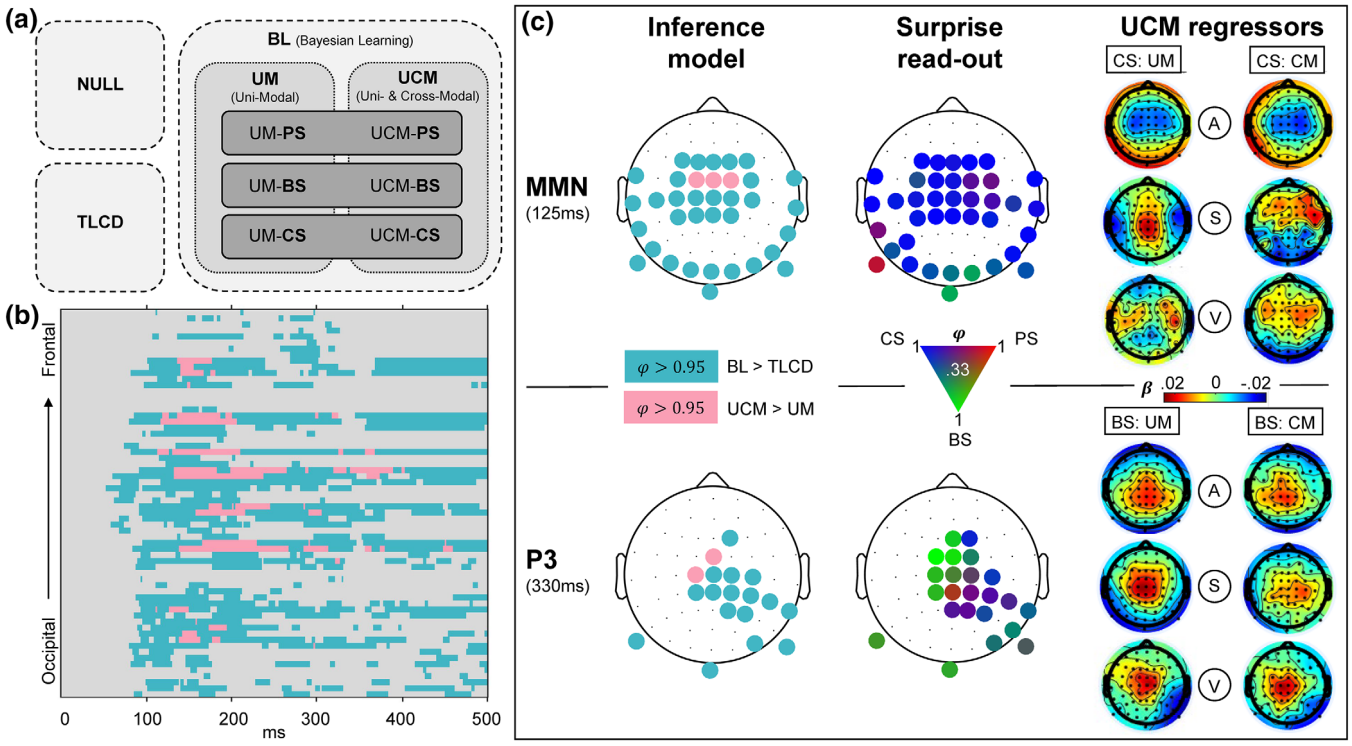


FIGURE 6 Modeling results. (a) Schematic overview of models. Model comparison 1 (light-gray box, dashed contour): Null model family (NULL), train length-dependent change detection model family (TLCD), and Bayesian learning model family (BL). Comparison 2 (gray box, dotted contour): Unimodal regression model family (UM), cross-modal regression model family (UCM). Comparison 3 (dark-gray box, line contour): Read-out model family comparison of predictive surprise family (PS), Bayesian surprise family (BS), and confidence-corrected surprise family (CS). (b) Results of comparison 1 and 2 shown for all electrodes and poststimulus time points. Color depicts exceedance probability (EP) $\varphi > 0.95$. Light-blue = BL > TLCD, pink = UCM > UM. (c) Topography of modeling results at time windows of MMN (top row) and P3 (bottom row). Left column: Results of comparison 1 (same colors as (B), depicting $\varphi > 0.95$). Middle column: Results of comparison 3. EPs between 0.33 and 1 of the three surprise functions are represented by a continuous three-dimensional RGB scale (red = predictive surprise [PS]; green = Bayesian surprise [BS]; blue = confidence-corrected surprise [CS]). Right column: Beta estimates of the model regressors of the UCM model (regressors: A = auditory; S = somatosensory; V = visual; CM = cross-modal; UM = unimodal) for CS read-out models (top) and BS read-out models (bottom).

For the late P3 MMR, a wide range of sources was expected to contribute to the EEG signal (Linden, 2005; Sabeti et al., 2016). To identify those that underlie the P3 MMR common to all modalities, we used a conjunction analysis. Significant clusters were found primarily in anterior cingulate cortex ($p_{fwe} < .05$; cluster: peak $t = 4.34$) and bilateral (pre-)frontal cortex ($p_{fwe} < .05$; left IFG cluster: peak $t = 3.57$; left superior frontal gyrus cluster: peak $t = 3.13$; left MFG cluster: peak $t = 2.87$; left frontal pole cluster: peak $t = 3.31$; right IFG cluster: peak $t = 3.45$; right MFG cluster: peak $t = 3.0$; right frontal pole cluster: peak $t = 3.57$). Additional significant sources were found in left inferior temporal gyrus ($p_{fwe} < .05$; cluster: peak $t = 3.21$), left and right lateral occipital cortex ($p_{fwe} < .05$; left cluster: peak $t = 2.99$; right cluster: peak $t = 3.45$) and right precentral gyrus ($p_{fwe} < .05$; cluster: peak $t = 2.64$).

3.4 | Single-trial modeling

As described in the previous sections, the responses of standards and deviants show specific sensitivity to (1) stimulus repetition and (2) cross-modal conditional probability. To investigate the computational principles underlying these response profiles, eight different models capturing various learning strategies were fit to the single-trial EEG data and compared via family-wise BMS. A summary of the modeling results is depicted in Figure 6.

The first model comparison aimed to further investigate observation (1) and the question whether the observed parametric modulation of standard and deviant EEG responses merely reflects a combination of neuronal adaptation and change detection dynamics or if the observed response patterns are indicative of an underlying generative model engaged in probabilistic inference. To this end, we ran family wise BMS which is schematically depicted in Figure 6a. The first comparison concerned the TLCD model, a model family containing all BL models, and a null model. In the fronto-central, temporal, occipital, and central electrodes showing MMN and P3 effects, the model comparison shows strong evidence in favor of the BL model family with an exceedance probability $\varphi > 0.95$ (corresponding to expected posterior probability $\langle r \rangle > .7$) from ~70 ms onward. On the other hand, the TLCD model did not exceed $\varphi > 0.95$ for any electrode or time point. Therefore, the TLCD model was disregarded at this

point, and we focused further investigation on the different BL models.

The second comparison set out to investigate observation (2) and evaluate the contribution of stimulus alternation tracking conditional on multimodal configurations beyond unimodal transition probability inference. Within the electrodes and time-points with sufficient evidence for BL signatures (as established by the first model comparison), a comparison of a purely unimodal model family with a cross-modally informed model family (UCM) was performed. Those electrodes and time-points where the additional inclusion of cross-modal regressors (UCM models) provided better model fits than purely unimodal models are highlighted in Figure 6b,c. The UCM family outperforms the UM family at central and fronto-central electrodes at ~100–400 ms (with $\varphi > 0.95$, corresponding to $\langle r \rangle > .7$).

Inspection of the beta estimates of auditory, somatosensory, and visual regressors of the UCM regression models shows that the beta maps of the unimodal predictors of the model resemble the ERP mismatch topographies of the respective modalities (depicted in Figure 6c). The cross-modal predictor, on the other hand, rather shows (fronto-) central activations which appear to resemble frontal aspects of the respective auditory, somatosensory, and visual MMRs.

The third comparison concerned the three surprise measures used as read-out functions for the probabilistic models. Overall, the family comparison does not show overwhelming evidence for any specific surprise function as only few electrodes reach exceedance probabilities of $\varphi > 0.95$. Nevertheless, a tendency of the MMN and the P3 to reflect different surprise dynamics can be observed. Although around the time of the MMN, only some electrodes show $\varphi > 0.95$ in favour of CS, inspection of the topographies without φ thresholding (as depicted in Figure 6c) shows CS to be dominant throughout the spatio-temporal range of the MMN (as suggested by higher EPs compared to BS and PS). On the other hand, at the time of the P3, BS appears to be the dominant surprise computation with multiple (fronto-) central electrodes showing $\varphi > 0.95$. Overall, the surprise comparison provides some evidence for a reflection of CS dynamics in the earlier mismatch signals around the time of the MMNs and suggests a tendency of the P3 to reflect BS dynamics.

In a final analysis, the optimal observation integration parameter τ was inspected. For each modality, the significant MMN clusters of the ERP analyses were used to inspect the optimal integration window of the regression models. For the UM regression model, highly similar optimal integration parameters were found within the electrodes and time-points of the different MMN clusters with no significant difference between the modalities. The optimal integration parameters were found to correspond to windows of stimulus integration with a half-life of (50% weighting at) around 5 to 25 stimuli (mean ± SD τ across participants within the MMN clusters: UM-CS = 0.0608 ± 0.0189 [8–16 stimuli], UM-BS = 0.0887 ± 0.0184 [6–9 stimuli], UCM-CS = 0.0550± 0.0179 [9–18 stimuli], UCM-BS = 0.0954 ± 0.0194 [6–9 stimuli]; mean ± SD τ across participants within the P3 clusters: UM-CS = 0.0536 ± 0.0214 [9–21 stimuli], UM-BS = 0.0811± 0.0191 [6–11 stimuli], UCM-CS = 0.0469 ± 0.0196 [10–25 stimuli], UCM-BS = 0.0862 ± 0.0215 [6–10 stimuli]). Overall, the same range of stimulus integration was found for the UM and UCM regression models and CS models tended to have higher integration windows (~10–20 stimuli) compared to Bayesian surprise models (~5–10 stimuli).

4 | DISCUSSION

The present study set out to compare mismatch signals in response to tri-modal sequence processing in the auditory, somatosensory, and visual modalities and to investigate influences of predictive cross-modal information. We found comparable but modality specific signatures of MMN-like early mismatch processing between 100 and 200 ms in all three modalities, which were source localized to their respective sensory specific cortices and shared right lateralized frontal sources. An additional cross-modal signature of mismatch processing was found in the P3 MMR for which a common network with frontal dominance was identified. With exception of the vMMN, both mismatch signals (MMN and P3) show parametric modulation by stimulus train length driven by reciprocal tendencies of standards and deviants across modalities. Strikingly, standard and deviant responses within the cross-modal P3 cluster were sensitive to predictive information carried by the tri-modal stimulus configuration. Comparisons of computational models indicated that BL models, tracking transitions between observations, captured the observed dynamics of single-trial responses to the roving stimulus sequences better than a static model reflecting TLCD. Moreover, a BL model which additionally captured cross-modal conditional dependence of stimulus alternation outperformed a purely unimodal BL model primarily at central electrodes. The comparison of different read-out functions for the BL models provides tentative evidence that the early MMN may reflect dynamics of CS whereas later P3 MMRs seem to reflect dynamics of BS.

4.1 | Modality specific mismatch signatures in response to tri-modal roving stimuli

By using a novel tri-modal roving stimulus sequence originating from an underlying Markov process of state transitions, we were able to elicit and extract unique EEG signatures in each of the three sensory modalities (auditory, somatosensory, and visual).

Of the EEG mismatch signatures, the auditory MMN is one of the most widely researched responses to deviation from an established stimulus regularity (Naatanen et al., 1978; Winkler et al., 2009). Contrasting responses to standard and deviant stimuli of the auditory sequence in the current study resulted in the expected fronto-central MMN signature with more negative responses to deviants compared to standards. The extent of the MMN might suggest an underlying negative mismatch component as proposed by Naatanen et al. (2005), which drives a more negative going ERP around the N1, extending beyond the P2 component. Such post-N1 effects of the MMN have been suggested as markers of a "genuine" mismatch component in contrast to confounds by stimulus properties modulating auditory

ERP components (Naatanen et al., 2007) and might speak against pure N1 adaptation (as suggested by Jaaskelainen et al., 2004; May et al., 1999).

The somatosensory equivalent to the auditory MMN (sMMN) reported in the current study shows negative polarity at bilateral temporal electrodes and corresponding central positivity. The sMMN likely reflects an enhanced N140 component, as suggested by Kekoni et al. (1997). However, most previous sMMN studies used oddball paradigms where some critical discussion revolves around the distinction of the sMMN from an N140 modulation by stimulus properties alone. Here, we report an sMMN around the N140 which can be assumed to be independent of stimulus confounds due to the reversed roles of standard and deviant stimuli in the roving paradigm. Although several previous studies have reported somatosensory mismatch responses, conflicting evidence exists regarding the exact components that may constitute an equivalent to the auditory MMN. Some studies report a more fronto-centrally oriented negativity (Kekoni et al., 1997; Shen et al., 2018; Spackman et al., 2007; Spackman et al., 2010) or observed such pronounced central positivity that they were led to conclude that it is in fact the central positivity that should be considered the somatosensory equivalent of the aMMN (Akatsuka et al., 2005; Shinozaki et al., 1998). However, some evidence appears to converge on a temporally centered negativity with corresponding central positivity as the primary sMMN around 140 ms (Gijsen et al., 2021; Ostwald et al., 2012).

While the auditory and somatosensory MMN's in the current study were found to be highly comparable in their signal strength, their hypothesized counterpart in the visual modality showed a comparatively weaker response. Nevertheless, we found a significant vMMN at occipital electrodes extending to temporal electrodes within a time window of 100–200 ms poststimulus, with corresponding (fronto-) central positivity. This observation is in line with previous research reporting posterior (Cleary et al., 2013; Kimura et al., 2010; Urakawa et al., 2010) and temporal (Heslenfeld, 2003; Kuldkepp et al., 2013) patterns of vMMN with corresponding central positivity (Cleary et al., 2013; Czigler et al., 2006; File et al., 2017).

4.2 | Neuronal generators of MMN signatures

Source reconstruction analyses were used to identify underlying neuronal generators of the modality specific MMN signatures. Interestingly, for each sensory modality, we found generators in the primary and higher order sensory cortices as well as additional frontal generators in IFG and MFG.

The sensory specific neuronal sources underlying the auditory MMN were identified as bilateral auditory cortex with a dominance in hierarchically higher auditory areas. With an additional modality independent contribution of right lateralized frontal sources, this set of neuronal generators identified for the aMMN is in line with previous research suggesting primary auditory cortex and higher auditory areas in superior temporal sulcus as well as right IFG as underlying the aMMN (Garrido et al., 2008; Garrido, Kilner, Kiebel, & Friston, 2009; Molholm et al., 2005; Naatanen et al., 2005; Opitz et al., 2002) with consideration of an additional frontal generator in MFG (Deouell, 2007).

The sources underlying the sMMN were identified in the current study as primary (S1) and secondary (S2) somatosensory cortices with additional frontal generators in right IFG and MFG. This finding is in accordance with previous research showing a combined response of S1 and S2 to underlie the sMMN (Akatsuka, Wasaka, Nakata, Kida, Hoshiyama, et al., 2007; Akatsuka, Wasaka, Nakata, Kida, & Kakigi, 2007; Andersen & Lundqvist, 2019; Butler et al., 2012; Gijsen et al., 2021; Naeije et al., 2016, 2018; Ostwald et al., 2012; Spackman et al., 2010) in combination with involvement of (inferior) frontal regions (Allen et al., 2016; Downar et al., 2000; Fardo et al., 2017; Huang et al., 2005; Ostwald et al., 2012).

For the visual modality, we identified sources in visual areas (V1–V4) and additional frontal activations in IFG and MFG as the neuronal generators underlying the vMMN. Previous studies have shown similar combinations of visual and prefrontal areas (Kimura et al., 2010; Kimura et al., 2011; Kimura et al., 2012; Urakawa et al., 2010; Yucel et al., 2007) and have particularly highlighted the IFG as a frontal generator of the vMMN (Downar et al., 2000; Hedge et al., 2015). Similarly, an fMRI study of perceptual sequence learning in the visual system has shown right lateralized prefrontal activation in addition to activations in visual cortex in response to regularity violations (Huettel et al., 2002). Yet another study has suggested a role for right prefrontal areas in interaction with hierarchically lower visual areas for the prediction of visual events (Kimura et al., 2012), all in line with our results.

Overall, our finding of inferior and middle frontal sources for the MMN in all three modalities provides further evidence for a modality independent role for these generators as previously suggested by Downar et al. (2000). As such, these modality-independent frontal generators might reflect higher stages of a predictive hierarchy working across modalities in interaction with lower modality specific regions, as previously suggested primarily for the auditory modality (Garrido, Kilner, Stephan, & Friston, 2009).

4.3 | Modulation of the MMN by stimulus repetition

An important feature of the MMN which theories of its generation have aimed to account for is its sensitivity to stimulus repetition. The MMN is known to increase with prior repetition of standards (Imada et al., 1993; Javitt et al., 1998; Naatanen & Näätänen, 1992; Sams et al., 1983). Correspondingly, in the current study, we find a significant increase of auditory and somatosensory MMN with the length of the preceding stimulus train as well as a comparable tendency for the vMMN. Moreover, we show that this increase was driven by a reciprocal negative modulation of deviant and positive modulation of standard responses, suggesting a combined influence of repetition dependent change detection and dynamics akin to stimulus adaptation.

The observed positive modulation of standard responses, particularly in the auditory modality, is in line with the repetition positivity account of Baldeweg and colleagues (Baldeweg, 2006; Baldeweg, 2007; Baldeweg et al., 2004; Haenschel et al., 2005). In the auditory modality, repetition positivity has been isolated as a positive slow wave that accounts for repetition-dependent increases of auditory ERPs up to the P2 component (Haenschel et al., 2005). With regard to its functional role, it has been argued to reflect auditory sensory memory trace formation (Baldeweg et al., 2004; Costa-Faidella, Baldeweg, et al., 2011; Costa-Faidella, Grimm, et al., 2011). Interestingly, MMN studies using the oddball paradigm often report an increasing MMN with standard repetition without further dissecting the contributions from standard and deviant dynamics. A contribution of the standard repetition positivity appears to be particularly dominant in roving stimulus paradigms (Cooper et al., 2013), potentially because a memory trace of the standard stimulus identity must be reestablished after each change of roles for standard and deviant stimuli. It has even been suggested that the memory trace dynamics of the standard observed in response to roving oddball sequences might in fact be the primary driver of train length effects on MMN amplitudes (Baldeweg et al., 2004; Costa-Faidella, Baldeweg, et al., 2011; Costa-Faidella, Grimm, et al., 2011; Haenschel et al., 2005). Importantly, although some evidence exists to suggest an additional role for train length dependent deviant modulation also in roving paradigms (Cowan et al., 1993; Haenschel et al., 2005), a dissection of combined standard and deviant contributions as performed here is rarely described.

Similar to the aMMN, we found the sMMN to be modulated by stimulus repetition. An early repetition positivity effect in the responses to standards was observed prior to 100 ms indicating comparable sensory adaptation dynamics as described for the aMMN. Subsequently, the negative deviant and sMMN responses increase with repetition around the N140 (i.e., around the sMMN peak). While somatosensory deviant responses have previously been shown to decrease with increasing stimulus probability (Akatsuka, Wasaka, Nakata, Kida, & Kakigi, 2007), only few other studies have reported sensitivity of the sMMN to stimulus repetition. Interestingly, in our previous study on somatosensory MMRs (Gijsen et al., 2021) we report the same reciprocal pattern found here: Negative modulation of the deviant and positive modulation of the standard response which result in an increase of the sMMN amplitude with stimulus train length.

In the visual modality, a comparable train length effect to auditory and somatosensory modalities was observed but did not reach statistical significance in the vMMN time window. Given the overall weaker response in the current study for vMMN this might not be surprising. Moreover, discussions about the repetition modulation of vMMN responses are often based on findings concerning the auditory system rather than direct findings in the visual modality. While sensory adaptation to stimulus repetition is generally found throughout the visual system (e.g., Clifford et al., 2007; Grill-Spector et al., 2006) it is rarely directly reported in vMMN studies (but see Kremlacek et al., 2016). Overall, the vMMN literature seems to suggest that the vMMN may be a rather unstable phenomenon. In fact, by controlling for confounding effects, one study has called the existence of the vMMN for low level features such as the ones used here into question entirely (Male et al., 2020). The vMMN appears to show a much less pronounced spatiotemporal pattern than auditory and somatosensory equivalents, which is reflected in larger variance in the reported topographies and time windows in studies investigating vMMN (but see Section 11 for a discussion of alternative explanations regarding the current study).

4.4 | MMN as a signature of predictive processing

Recent research supports the view that Bayesian perceptual learning mechanisms underlie the generation of mismatch responses such as the MMN (Friston, 2005, 2010; Garrido, Kilner, Stephan, & Friston, 2009). Given the proposal of Bayesian inference and predictive processing as universal principles of perception and perceptual learning in the brain (Friston, 2005, 2010), comparable mismatch responses are expected to be found across sensory modalities. Evidence for the predictive nature of mismatch responses, akin to key findings from the auditory modality, is for instance given by studies showing somatosensory (Andersen & Lundqvist, 2019; Naeije et al., 2018) and visual (Czigler et al., 2006; Kok et al., 2014) MMN in response to predicted but omitted stimuli. Moreover, Ostwald et al. (2012) and Gijsen et al. (2021) have shown that single trial somatosensory MMN and P3 MMRs can be accounted for in terms of surprise signatures of Bayesian inference models tracking stimulus transitions. Similarly, the vMMN has been described as a signature of predictive processing (Kimura et al., 2011; Stefanics et al., 2014), signaling prediction error instead of basic change detection (Stefanics et al., 2018).

Correspondingly, we found comparable mismatch signatures in auditory, somatosensory, and visual modalities. The train length effects observed in our study across modalities have previously been related to predictive processing. Repetition positivity in the auditory modality has been interpreted as a reflection of repetition suppression, resulting from fulfilled prediction (Auksztulewicz & Friston, 2016; Baldeweg, 2007; Costa-Faidella, Baldeweg, et al., 2011; Costa-Faidella, Grimm, et al., 2011). A corresponding negative modulation of deviant responses on the other hand, would signal a failure to suppress prediction error after violation of the regularity established by the current stimulus train. Under such a view, longer trains of repetitions lead to higher precision in the probability estimate which in turn results in a scaling of the prediction error in response to prediction violation (Auksztulewicz & Friston, 2016; Friston, 2005; Friston & Kiebel, 2009). In line with these hypotheses, Garrido and colleagues (Garrido et al., 2008; Garrido, Kilner, Kiebel, & Friston, 2009) used dynamic causal modeling (DCM) to show that the MMN elicited in a roving stimulus paradigm is best explained by the combined dynamics of auditory adaptation and model adjustment. Their network, proposed to underlie MMN generation, was set up as an implementation of hierarchical predictive processing involving bottom-up signals from auditory cortex and top-down modulations by

inferior frontal cortex. Similarly, another DCM study proposed a predictive coding model of pain processing in response to somatosensory oddball sequences, highlighting the role of inferior frontal cortex in top-down modulations of somatosensory potentials (Fardo et al., 2017). As we find involvement of such modality specific sensory and modality independent frontal areas for MMN responses across modalities, our results suggest comparable roles for these sources in a predictive hierarchy.

4.5 | P3 Mismatch responses reflect cross-modal processing

In addition to the modality specific MMN responses, deviants in all three modalities elicited a late positive mismatch component in the P3 time window. Despite differences in the exact latency and extent of this response between modalities, we identified a common mismatch cluster from 300 to 350 ms in central electrodes, followed by a slightly more posterior cluster extending from 400 to 600 ms. Particularly the earlier cluster may correspond to the well-known P3a response, which peaks at around 300 ms after change-onset at (fronto-) central electrodes and is thought to be elicited regardless of sensory modality (Escera et al., 2000; Friedman et al., 2001; Knight & Scabini, 1998; Polich, 2007; Schroger, 1996).

The P3a is closely related to the MMN as they are both elicited during active and passive perception of repeated stimuli interrupted by infrequent stimulus deviations (Polich, 2007; Schroger et al., 2015). While the P3a has been initially related to attentional switches to task-irrelevant but salient stimulus features (Escera et al., 2000; Friedman et al., 2001; Polich, 2007), more recent accounts suggest that the MMN and P3a might reflect two stages of a predictive hierarchy, each representing (potentially differentiable) prediction error responses (Schroger et al., 2015; Wacongne et al., 2011). Similar to the MMN, P3 responses are known to be modulated by stimulus probability (Duncan-Johnson & Donchin, 1977) and can be elicited by unexpected stimulus repetitions (Duncan et al., 2009; Squires et al., 1975) and omissions of predicted sound stimuli (Prete et al.; Sutton et al., 1967), which provides compelling evidence for a role of the P3 in predictive processing. Similar to the MMN responses described above, we found the individual P3 MMR responses in all three modalities to show reciprocal modulations of standards and deviants by stimulus repetition, which has previously only been reported for the auditory modality (Bendixen et al., 2007). This sensitivity to stimulus repetition of mismatch responses in early and late time-windows has been interpreted in terms of regularity and rule extraction in the auditory modality (Bendixen et al., 2007) and is in line with an account of repetition suppression over and above early sensory adaptation.

The MMN and P3 MMR have been shown to be differentially modulated by higher order predictability. The P3 is reduced by the presentation of visual cues preceding an auditory deviant, while the MMN is not affected by the same top-down predictability (Horvath & Bendixen, 2012; Ritter et al., 1999; Sussman et al., 2003). Similarly, explicit top-down knowledge of sequence regularities has been shown to reduce the P3, while leaving the MMN unaffected (Max et al., 2015). It has thus been suggested that the P3 reflects a higher-level deviance detection system concerned with the significance of the stimulus in providing new information for the system (Horvath et al., 2008). Interestingly, a recent study investigating mismatch responses to different auditory features showed that while the MMN response in an earlier (classical) time window was generally affected by regularity violations, only the later response (P3 range) contained information about the specific features that were violated (An et al., 2021). Furthermore, computational studies indicate that P3 responses reflect specific quantities of unexpectedness as well as updates to a prior belief (Jepma et al., 2016; Kolossa et al., 2015).

Overall, current research provides evidence for the view that the MMN reflects prediction errors at earlier hierarchical stages, primarily concerned with more local regularity extraction, whereas P3 responses reflect more global rule violations which require a certain level of abstraction and information integration (Bekinschtein et al., 2009; Wacongne et al., 2011; Winkler et al., 2005). Our findings of a sensitivity of the P3 response to cross-modal predictive information carried by the multimodal configuration of the stimulus sequence further supports such a view. Across modalities, we found an increased P3 response to mispredicted compared to predicted or unpredictable stimuli, regardless of their role as standards or deviants. Generally, the P3 deviant response in the current study likely reflects a (unsigned) prediction error to a local regularity established by stimulus repetition. However, increased P3 responses to mispredicted stimuli indicate additional violations of global, cross-modal predictions which are extracted from multimodal context information.

The observed pattern suggests influences of precision weighting on prediction errors (Friston & Kiebel, 2009). In case of both predicted and mispredicted stimuli, the cross-modal predictive context allows for more precise predictions (i.e., high prior precision) than in case of the unpredictable stimuli (low prior precision). Under such an interpretation, the precision for mispredicted deviants is high, resulting in a pronounced prediction error response. Since the precision for predicted deviants is also high, the resulting prediction error response is low because the stimulus was suppressed. Even though the size of prediction error to unpredictable deviants could generally be expected in between those of predicted and mispredicted deviants, the observed response is low (similar to that of a predicted deviant), because the prior precision in this context is low. This interpretation is in line with the fact that no significant difference was found between predicted and unpredictable deviants. A similar modulation of multimodal predictability is found for the P3 response to standards. However, interestingly, in case of the standards, the response to predicted stimuli is significantly lower than to unpredictable stimuli. This difference between standards and deviants could be due to the fact that deviants are generally surprising, even if they are more predictable in terms of their cross-modal configuration. Standards, on the other hand, are generally predicted to occur (high precision) which might result in a pronounced suppression of prediction error in case they are additionally cross-modally predicted.

The interpretation of the common P3 cluster as a cross-modal P3a response sensitive to multimodal predictive information is further supported by our source localization results, which particularly indicate prefrontal regions such as the medial frontal, inferior frontal and anterior cingulate cortex as sources of the P3 MMR. Although notoriously diverse, previous research on P3 sources has identified a fronto-parietal network of generators, particularly highlighting the role of prefrontal and anterior cingulate regions in generating the P3 novelty response (P3a; Linden, 2005; Polich, 2007), whereas parietal regions are presumed to be more involved in task-related P3b responses. The identified sources have been shown to be involved in a fronto-parietal network relevant for the supramodal processing of stimulus transitions and deviance detection (Downar et al., 2000; Huang et al., 2005). Similarly, a fronto-parietal attention network (Corbetta & Shulman, 2002) has been shown to be involved in oddball processing in the auditory and visual modalities (Kim, 2014). The network consists of two functionally and anatomically distinct parts which closely interact (Vossel et al., 2014). While the dorsal part of the network is believed to be involved in the allocation of top-down, endogenous attention (e.g., triggered by predictive information), the ventral part is involved in bottom-up, exogenous attention allocation and thus, processing of unexpected stimuli. Importantly, it has been shown that this network operates supramodally to facilitate processing of information from multimodal events (Macaluso, 2010; Macaluso & Driver, 2005). Thus, the predictive information in the multimodal sequences presented in the current study may be processed in such a fronto-parietal network to aid the perception of multimodal stimulus streams. Future research would benefit from studies further investigating such multimodal probabilistic sequences with higher spatial resolution to inform these proposed interpretations.

4.6 | Modeling single-trial EEG responses as signatures of bayesian inference

Given the results of the average- and GLM-based EEG analyses, we aimed to test if the observed modulations of standards, deviants and MMRs by local (train length) and global (cross-modal predictability) sequence properties could be captured by signatures of Bayesian inference. To this end, we compared a simple TLCD model to families of BL models capturing different aspects of the sequence statistics. In light of the literature discussed above, we hypothesized that BL models would outperform the TLCD model in explaining the recorded mismatch responses.

Overall, the BL models outperformed the static TLCD model in all electrodes in the MMN and P3 clusters indicating that these responses reflect dynamics beyond the basic repetition effects observed in the ERP analyses. This result provides evidence to suggest that the MMN and P3 MMR capture the trial-to-trial dynamics of Bayesian inference and are thus markers of probabilistic sequence processing in the brain.

Within the family of BL models, we found that a cross-modally informed model (UCM), tracking cross-modal conditional dependencies between modalities in addition to unimodal transitions, outperformed a purely unimodal transition probability model (UM) at central electrodes within an early and a late time-window. The cross-modal effects in the late time-window are directly in line with the sensitivity of the P3 cluster to cross-modal predictability discussed above and support an interpretation of P3 mismatch responses to reflect signatures of cross-modal Bayesian inference. Given that cross-modal learning was not explicitly instructed or task-relevant, the results are compatible with the view that the brain is sensitive to cross-modal information by default (Driver & Noesselt, 2008; Ghazanfar & Schroeder, 2006) and that processing multimodal information might be appropriately captured by Bayesian inference (Kording et al., 2007; Shams & Beierholm, 2022). Interestingly, however, an earlier cross-modal effect was found prior to 300 ms which was not reflected in the GLM results, suggesting that potential modulations of MMN signatures by predictability manifest in the dynamics of single trial surprise signals but not in significant mean differences between predictability conditions. Since the earlier cross-modal effect observed in the modeling results was primarily confined to central and fronto-central electrodes it may be related to activity of the frontal generators of the MMN. As discussed above, the frontal cortex is assumed to be involved in MMN generation (Deouell, 2007) in interaction with hierarchically lower sensory sources and has been hypothesized to form top-down predictions about incoming sensory stimuli (Garrido et al., 2008; Garrido, Kilner, Kiebel, & Friston, 2009; Garrido, Kilner, Stephan, & Friston, 2009). This assumption is further supported by our source reconstruction results which show modality independent frontal generators in addition to sensory specific regions to underlie the MMN in auditory, somatosensory, and visual modalities.

Regarding the surprise read-out functions of the BL models, we find a slight dominance of CS in earlier mismatch signatures prior to 200 ms, while the late clusters tend to reflect BS. This is well in line with our previous study performed in the somatosensory modality (Gijsen et al., 2021) and other studies have similarly reported a reflection of BS in P3 mismatch responses (Kolossa et al., 2015; Mars et al., 2008; Ostwald et al., 2012; Seer et al., 2016). Given their differences in reading out the probability estimates of the Bayesian observer, the different surprise signatures in the MMN and P3 MMR might provide some insight into their respective computational roles. CS has been categorized as an instantiation of puzzlement surprise (Faraji et al., 2018) reflecting a mismatch between sensory input and internal model belief which is additionally scaled by belief commitment. Low-probability events are thus more surprising if commitment to the belief (of this estimate) is high. BS reflects incorporation of new information, quantifying an update to the generative model and has been categorized as enlightenment surprise (Faraji et al., 2018). Accordingly, the MMN may be considered a marker of prediction error scaled by belief commitment, whereas the P3 may reflect the subsequent update of the predictive model.

Given that the P3 shows sensitivity to cross-modal prediction violation (GLM results) and tends to carry signatures of multimodal inference and model updating (Bayesian modeling results), we suggest that the P3 likely reflects model updates with respect to the multimodal

context. Although unnoticed by participants, given that the statistical regularities changed across experimental runs, the generative model continuously required updating. As such, our results might be reflective of a volatile sensory environment and relate to previous findings which indicate that later MMRs, such as the P3, reflect belief updates about the volatility of the underlying (hidden) statistics governing sensory observations (e.g., Weber et al., 2020; Weber et al., 2022). We leave it for future research to design experiments which are better suited to evaluate these speculations more specifically and more thoroughly.

4.7 | Limitations

Although we gained valuable insights into the commonalities and differences between mismatch responses in different modalities, our study faces certain limitations in its implementation and scope. First, although reports of weak vMMN responses can be found in the literature, an alternative explanation may lie in the stimulation protocol used in the current study. Our visual stimuli consisted of bilateral flash stimuli with two different intensities, which were presented in the periphery of the visual field. Since, to our knowledge, no other study has used visual flash stimuli to elicit vMMN, our results are not directly comparable to previous research. Moreover, due to the retinotopic organization of the visual cortex (Horton & Hoyt, 1991; Sereno et al., 1995), a "far peripheral" placement (i.e., >60°; Strasburger et al., 2011) of the LED's results in the activation of (primary) visual areas folded deep inside the cortex, in the calcarine sulcus between the hemispheres. It is therefore possible that the visual mismatch responses were not weaker per se but were merely harder to detect by means of EEG.

Further, our results concerning the comparison of the surprise read-out functions provide some indications of the computational roles for early and late MMRs, which are in line with previous research. However, it should be noted that the current study was not specifically designed to investigate their (nuanced) differences. The inclusion of three read-out functions primarily served the purpose of avoiding bias in the comparison of the BL models by prior choice of the read-out. To this end, the most prominent surprise read-out functions used in the literature were included. Research would benefit from future studies specifically designed to compare different surprise measures without the manipulation of other aspects of the underlying models. A valuable overview and suggested experiments for that purpose have been recently provided by Modirshanechi et al. (2022).

5 | CONCLUSION

With the current study, we provide evidence for modality specific and modality independent aspects of mismatch responses in audition, somatosensation, and vision resulting from a simultaneous stream of tri-modal roving stimulus sequences. Our results suggest that responses to stimulus transitions in all three modalities are based on an interaction of hierarchically lower, modality specific areas with hierarchically higher, modality independent frontal areas. We show that similar dynamics underlie these mismatch responses which likely reflect predictive processing and Bayesian inference on unimodal and multimodal sensory input streams.

ACKNOWLEDGMENT

The authors would like to thank the HPC Service of ZEDAT, Freie Universität Berlin, for computing time. Open Access funding enabled and organized by Projekt DEAL.

FUNDING INFORMATION

This work was supported by Berlin School of Mind and Brain, Humboldt Universität zu Berlin (MG and SG, http://www.mind-and-brain.de/home/), and Deutscher Akademischer Austauschdienst (SG, https://www.daad.de/en/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

CONFLICT OF INTEREST

The authors declare that no competing interests exist.

DATA AVAILABILITY STATEMENT

The raw EEG data and the code used for analysis will be made freely available in an online repository upon publication of the article.

ORCID

Miro Grundei https://orcid.org/0000-0002-6396-9674

REFERENCES

Akatsuka, K., Wasaka, T., Nakata, H., Inui, K., Hoshiyama, M., & Kakigi, R. (2005). Mismatch responses related to temporal discrimination of somatosensory stimulation. *Clinical Neurophysiology*, *116*(8), 1930–1937. https://doi.org/10.1016/j.clinph.2005.04.021

Akatsuka, K., Wasaka, T., Nakata, H., Kida, T., Hoshiyama, M., Tamura, Y., & Kakigi, R. (2007). Objective examination for two-point stimulation using a somatosensory oddball paradigm: An MEG study. *Clinical Neurophysiology*, *118*(2), 403–411. https://doi.org/10.1016/j.clinph.2006.09.030

Akatsuka, K., Wasaka, T., Nakata, H., Kida, T., & Kakigi, R. (2007). The effect of stimulus probability on the somatosensory mismatch field. *Experimental Brain Research*, *181*(4), 607–614. https://doi.org/10.1007/s00221-007-0958-4

Alain, C., Woods, D. L., & Ogawa, K. H. (1994). Brain indices of automatic pattern processing. *Neuroreport*, *6*(1), 140–144. https://doi.org/10.1097/00001756-199412300-00036

Allen, M., Fardo, F., Dietz, M. J., Hillebrandt, H., Friston, K. J., Rees, G., & Roepstorff, A. (2016). Anterior insula coordinates hierarchical processing of tactile mismatch responses. *NeuroImage*, *127*, 34–43. https://doi.org/10.1016/j.neuroimage.2015.11.030

An, H., Ho Kei, S., Auksztulewicz, R., & Schnupp, J. W. H. (2021). Do auditory mismatch responses differ between acoustic features? *Frontiers in Human Neuroscience*, *15*, 613903. https://doi.org/10.3389/fnhum.2021.613903

Andersen, L. M., & Lundqvist, D. (2019). Somatosensory responses to nothing: An MEG study of expectations during omission of tactile stimulations. *NeuroImage*, *184*, 78–89. https://doi.org/10.1016/j.neuroimage.2018.09.014

Andric, M., Davis, B., & Hasson, U. (2017). Visual cortex signals a mismatch between regularity of auditory and visual streams. *NeuroImage*, *157*, 648–659. https://doi.org/10.1016/j.neuroimage.2017.05.028

Arnal, L. H., & Giraud, A. L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences*, *16*(7), 390–398. https://doi.org/10.1016/j.tics.2012.05.003

Auksztulewicz, R., & Friston, K. (2016). Repetition suppression and its contextual determinants in predictive coding. *Cortex*, *80*, 125–140. https://doi.org/10.1016/j.cortex.2015.11.024

Baldeweg, T. (2006). Repetition effects to sounds: Evidence for predictive coding in the auditory system. *Trends in Cognitive Sciences*, *10*(3), 93–94. https://doi.org/10.1016/j.tics.2006.01.010

Baldeweg, T. (2007). ERP repetition effects and mismatch negativity generation: A predictive coding perspective. *Journal of Psychophysiology*, *21*(3–4), 204–213.

Baldeweg, T., Klugman, A., Gruzelier, J., & Hirsch, S. R. (2004). Mismatch negativity potentials and cognitive impairment in schizophrenia. *Schizophrenia Research*, *69*(2–3), 203–217. https://doi.org/10.1016/j.schres.2003.09.009

Baldi, P., & Itti, L. (2010). Of bits and wows: A Bayesian theory of surprise with applications to attention. *Neural Networks*, *23*(5), 649–666. https://doi.org/10.1016/j.neunet.2009.12.007

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, *76*(4), 695–711. https://doi.org/10.1016/j.neuron.2012.10.038

Bekinschtein, T. A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., & Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(5), 1672–1677. https://doi.org/10.1073/pnas.0809667106

Bendixen, A., Roeber, U., & Schroger, E. (2007). Regularity extraction and application in dynamic auditory stimulus sequences. *Journal of Cognitive Neuroscience*, *19*(10), 1664–1677. https://doi.org/10.1162/jocn.2007.19.10.1664

Berg, P., & Scherg, M. (1994). A multiple source approach to the correction of eye artifacts. *Electroencephalography and Clinical Neurophysiology*, *90*(3), 229–241. https://doi.org/10.1016/0013-4694(94)90094-9

Besle, J., Fort, A., & Giard, M. H. (2005). Is the auditory sensory memory sensitive to visual information? *Experimental Brain Research*, *166*(3–4), 337–344. https://doi.org/10.1007/s00221-005-2375-x

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. MIT Press.

Bresciani, J. P., Dammeier, F., & Ernst, M. O. (2006). Vision and touch are automatically integrated for the perception of sequences of events. *Journal of Vision*, *6*(5), 554–564. https://doi.org/10.1167/6.5.2

Bresciani, J. P., Dammeier, F., & Ernst, M. O. (2008). Tri-modal integration of visual, tactile and auditory signals for the perception of sequences of events. *Brain Research Bulletin*, *75*(6), 753–760. https://doi.org/10.1016/j.brainresbull.2008.01.009

Butler, J. S., Foxe, J. J., Fiebelkorn, I. C., Mercier, M. R., & Molholm, S. (2012). Multisensory representation of frequency across audition and touch: High density electrical mapping reveals early sensory-perceptual coupling. *The Journal of Neuroscience*, *32*(44), 15338–15344. https://doi.org/10.1523/JNEUROSCI.1796-12.2012

Cao, Y., Summerfield, C., Park, H., Giordano, B. L., & Kayser, C. (2019). Causal inference in the multisensory brain. *Neuron*, *102*(5), 1076–1087 e1078. https://doi.org/10.1016/j.neuron.2019.03.043

Cleary, K. M., Donkers, F. C., Evans, A. M., & Belger, A. (2013). Investigating developmental changes in sensory processing: Visual mismatch response in healthy children. *Frontiers in Human Neuroscience*, *7*, 922. https://doi.org/10.3389/fnhum.2013.00922

Clifford, C. W., Webster, M. A., Stanley, G. B., Stocker, A. A., Kohn, A., Sharpee, T. O., & Schwartz, O. (2007). Visual adaptation: Neural, psychological and computational aspects. *Vision Research*, *47*(25), 3125–3131. https://doi.org/10.1016/j.visres.2007.08.023

Cooper, R. J., Atkinson, R. J., Clark, R. A., & Michie, P. T. (2013). Event-related potentials reveal modelling of auditory repetition in the brain. *International Journal of Psychophysiology*, *88*(1), 74–81. https://doi.org/10.1016/j.ijpsycho.2013.02.003

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews. Neuroscience*, *3*(3), 201–215. https://doi.org/10.1038/nrn755

Costa-Faidella, J., Baldeweg, T., Grimm, S., & Escera, C. (2011). Interactions between "what" and "when" in the auditory system: Temporal predictability enhances repetition suppression. *The Journal of Neuroscience*, *31*(50), 18590–18597. https://doi.org/10.1523/JNEUROSCI.2599-11.2011

Costa-Faidella, J., Grimm, S., Slabu, L., Diaz-Santaella, F., & Escera, C. (2011). Multiple time scales of adaptation in the auditory system as revealed by human evoked potentials. *Psychophysiology*, *48*(6), 774–783. https://doi.org/10.1111/j.1469-8986.2010.01144.x

Cowan, N., Winkler, I., Teder, W., & Naatanen, R. (1993). Memory prerequisites of mismatch negativity in the auditory event-related potential (ERP). *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *19*(4), 909–921. https://doi.org/10.1037//0278-7393.19.4.909

Czigler, I., Winkler, I., Pato, L., Varnagy, A., Weisz, J., & Balazs, L. (2006). Visual temporal window of integration as revealed by the visual mismatch negativity event-related potential to stimulus omissions. *Brain Research*, *1104*(1), 129–140. https://doi.org/10.1016/j.brainres.2006.05.034

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., & Pallier, C. (2015). The neural representation of sequences: From transition probabilities to algebraic patterns and linguistic trees. *Neuron*, *88*(1), 2–19. https://doi.org/10.1016/j.neuron.2015.09.019

Deouell, L. Y. (2007). The frontal generator of the mismatch negativity revisited. *Journal of Psychophysiology*, *21*(3–4), 188–203.

Downar, J., Crawley, A. P., Mikulis, D. J., & Davis, K. D. (2000). A multimodal cortical network for the detection of changes in the sensory environment. *Nature Neuroscience*, *3*(3), 277–283. https://doi.org/10.1038/72991

Driver, J., & Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. *Neuron*, *57*(1), 11–23. https://doi.org/10.1016/j.neuron.2007.12.013

Duncan, C. C., Barry, R. J., Connolly, J. F., Fischer, C., Michie, P. T., Naatanen, R., Polich, J., Reinvang, I., & Van Petten, C. (2009). Event-related potentials in clinical research: Guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clinical Neurophysiology*, *120*(11), 1883–1908. https://doi.org/10.1016/j.clinph.2009.07.045

Duncan-Johnson, C. C., & Donchin, E. (1977). On quantifying surprise: The variation of event-related potentials with subjective probability. *Psychophysiology*, *14*(5), 456–467. https://doi.org/10.1111/j.1469-8986.1977.tb01312.x

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*, *25*(4), 1325–1335. https://doi.org/10.1016/j.neuroimage.2004.12.034

Escera, C., Alho, K., Schroger, E., & Winkler, I. (2000). Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiology & Neuro-Otology*, *5*(3–4), 151–166. https://doi.org/10.1159/000013877

Faraji, M., Preuschoff, K., & Gerstner, W. (2018). Balancing new against old information: The role of puzzlement surprise in learning. *Neural Computation*, *30*(1), 34–83. https://doi.org/10.1162/neco_a_01025

Fardo, F., Auksztulewicz, R., Allen, M., Dietz, M. J., Roepstorff, A., & Friston, K. J. (2017). Expectation violation and attention to pain jointly

modulate neural gain in somatosensory cortex. *NeuroImage*, *153*, 109–121. https://doi.org/10.1016/j.neuroimage.2017.03.041

File, D., File, B., Bodnar, F., Sulykos, I., Kecskes-Kovacs, K., & Czigler, I. (2017). Visual mismatch negativity (vMMN) for low- and high-level deviances: A control study. *Attention, Perception, & Psychophysics*, *79*(7), 2153–2170. https://doi.org/10.3758/s13414-017-1373-y

Fitzgerald, K., & Todd, J. (2020). Making sense of mismatch negativity. *Frontiers in Psychiatry*, *11*, 468. https://doi.org/10.3389/fpsyt.2020.00468

Flandin, G., & Penny, W. D. (2007). Bayesian fMRI data analysis with sparse spatial basis function priors. *NeuroImage*, *34*(3), 1108–1125. https://doi.org/10.1016/j.neuroimage.2006.10.005

Friedel, E. B. N., Bach, M., & Heinrich, S. P. (2020). Attentional interactions between vision and hearing in event-related responses to Crossmodal and conjunct oddballs. *Multisensory Research*, *33*(3), 251–275. https://doi.org/10.1163/22134808-20191329

Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: An event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience and Biobehavioral Reviews*, *25*(4), 355–373. https://doi.org/10.1016/s0149-7634(01)00019-7

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

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews. Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1521), 1211–1221. https://doi.org/10.1098/rstb.2008.0300

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences*, *19*(3), 117–125. https://doi.org/10.1016/j.tics.2014.12.010

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

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, *10*(6), 278–285. https://doi.org/10.1016/j.tics.2006.04.008

Gijsen, S., Grundei, M., Lange, R. T., Ostwald, D., & Blankenburg, F. (2021). Neural surprise in somatosensory Bayesian learning. *PLoS Computational Biology*, *17*(2), e1008068. https://doi.org/10.1371/journal.pcbi.1008068

Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *290*(1038), 181–197. https://doi.org/10.1098/rstb.1980.0090

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, *10*(1), 14–23. https://doi.org/10.1016/j.tics.2005.11.006

Haenschel, C., Vernon, D. J., Dwivedi, P., Gruzelier, J. H., & Baldeweg, T. (2005). Event-related brain potential correlates of human auditory sensory memory-trace formation. *The Journal of Neuroscience*, *25*(45), 10494–10501. https://doi.org/10.1523/JNEUROSCI.1227-05.2005

Hedge, C., Stothart, G., Todd Jones, J., Rojas Frias, P., Magee, K. L., & Brooks, J. C. (2015). A frontal attention mechanism in the visual mismatch negativity. *Behavioural Brain Research*, *293*, 173–181. https://doi.org/10.1016/j.bbr.2015.07.022

Heilbron, M., & Chait, M. (2018). Great expectations: Is there evidence for predictive coding in auditory cortex? *Neuroscience*, *389*, 54–73. https://doi.org/10.1016/j.neuroscience.2017.07.061

Heslenfeld, D. J. (2003). Visual mismatch negativity. In *Detection of change: Event-related potential and fMRI findings* (pp. 41–59). Kluwer Academic Publishers.

Horton, J. C., & Hoyt, W. F. (1991). The representation of the visual field in human striate cortex. A revision of the classic Holmes map. *Archives of Ophthalmology*, *109*(6), 816–824. https://doi.org/10.1001/archopht.1991.01080060080030

Horvath, J., & Bendixen, A. (2012). Preventing distraction by probabilistic cueing. *International Journal of Psychophysiology*, *83*(3), 342–347. https://doi.org/10.1016/j.ijpsycho.2011.11.019

Horvath, J., & Winkler, I. (2004). How the human auditory system treats repetition amongst change. *Neuroscience Letters*, *368*(2), 157–161. https://doi.org/10.1016/j.neulet.2004.07.004

Horvath, J., Winkler, I., & Bendixen, A. (2008). Do N1/MMN, P3a, and RON form a strongly coupled chain reflecting the three stages of auditory distraction? *Biological Psychology*, *79*(2), 139–147. https://doi.org/10.1016/j.biopsycho.2008.04.001

Hu, L., Zhao, C., Li, H., & Valentini, E. (2013). Mismatch responses evoked by nociceptive stimuli. *Psychophysiology*, *50*(2), 158–173. https://doi.org/10.1111/psyp.12000

Huang, M. X., Lee, R. R., Miller, G. A., Thoma, R. J., Hanlon, F. M., Paulson, K. M., Martin, K., Harrington, D. L., Weisend, M. P., Edgar, J. C., & Canive, J. M. (2005). A parietal-frontal network studied by somatosensory oddball MEG responses, and its cross-modal consistency. *NeuroImage*, *28*(1), 99–114. https://doi.org/10.1016/j.neuroimage.2005.05.036

Huettel, S. A., Mack, P. B., & McCarthy, G. (2002). Perceiving patterns in random series: Dynamic processing of sequence in prefrontal cortex. *Nature Neuroscience*, *5*(5), 485–490. https://doi.org/10.1038/nn841

Hughes, H. C., Darcey, T. M., Barkan, H. I., Williamson, P. D., Roberts, D. W., & Aslin, C. H. (2001). Responses of human auditory association cortex to the omission of an expected acoustic event. *NeuroImage*, *13*(6 Pt 1), 1073–1089. https://doi.org/10.1006/nimg.2001.0766

Ille, N., Berg, P., & Scherg, M. (2002). Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. *Journal of Clinical Neurophysiology*, *19*(2), 113–124. https://doi.org/10.1097/00004691-200203000-00002

Imada, T., Hari, R., Loveless, N., McEvoy, L., & Sams, M. (1993). Determinants of the auditory mismatch response. *Electroencephalography and Clinical Neurophysiology*, *87*(3), 144–153. https://doi.org/10.1016/0013-4694(93)90120-k

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research*, *49*(10), 1295–1306. https://doi.org/10.1016/j.visres.2008.09.007

Jaaskelainen, I. P., Ahveninen, J., Bonmassar, G., Dale, A. M., Ilmoniemi, R. J., Levanen, S., Lin, F. H., May, P., Melcher, J., Stufflebeam, S., Tiitinen, H., & Belliveau, J. W. (2004). Human posterior auditory cortex gates novel sounds to consciousness. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(17), 6809–6814. https://doi.org/10.1073/pnas.0303760101

Javitt, D. C., Grochowski, S., Shelley, A. M., & Ritter, W. (1998). Impaired mismatch negativity (MMN) generation in schizophrenia as a function of stimulus deviance, probability, and interstimulus/interdeviant interval. *Electroencephalography and Clinical Neurophysiology*, *108*(2), 143–153. https://doi.org/10.1016/s0168-5597(97)00073-7

Jepma, M., Murphy, P. R., Nassar, M. R., Rangel-Gomez, M., Meeter, M., & Nieuwenhuis, S. (2016). Catecholaminergic regulation of learning rate

in a dynamic environment. *PLoS Computational Biology*, *12*(10), e1005171. https://doi.org/10.1371/journal.pcbi.1005171

Kekoni, J., Hamalainen, H., Saarinen, M., Grohn, J., Reinikainen, K., Lehtokoski, A., & Naatanen, R. (1997). Rate effect and mismatch responses in the somatosensory system: ERP-recordings in humans. *Biological Psychology*, *46*(2), 125–142. https://doi.org/10.1016/s0301-0511(97)05249-6

Kiat, J. E. (2018). Assessing cross-modal target transition effects with a visual-auditory oddball. *International Journal of Psychophysiology*, *129*, 58–66. https://doi.org/10.1016/j.ijpsycho.2018.04.010

Kilner, J. M., Kiebel, S. J., & Friston, K. J. (2005). Applications of random field theory to electrophysiology. *Neuroscience Letters*, *374*(3), 174–178. https://doi.org/10.1016/j.neulet.2004.10.052

Kim, H. (2014). Involvement of the dorsal and ventral attention networks in oddball stimulus processing: A meta-analysis. *Human Brain Mapping*, *35*(5), 2265–2284. https://doi.org/10.1002/hbm.22326

Kimura, M., Kondo, H., Ohira, H., & Schroger, E. (2012). Unintentional temporal context-based prediction of emotional faces: An electrophysiological study. *Cerebral Cortex*, *22*(8), 1774–1785. https://doi.org/10.1093/cercor/bhr244

Kimura, M., Ohira, H., & Schroger, E. (2010). Localizing sensory and cognitive systems for pre-attentive visual deviance detection: An sLORETA analysis of the data of Kimura et al. (2009). *Neuroscience Letters*, *485*(3), 198–203. https://doi.org/10.1016/j.neulet.2010.09.011

Kimura, M., Schroger, E., & Czigler, I. (2011). Visual mismatch negativity and its importance in visual cognitive sciences. *Neuroreport*, *22*(14), 669–673. https://doi.org/10.1097/WNR.0b013e32834973ba

Knight, R. T., & Scabini, D. (1998). Anatomic bases of event-related potentials and their relationship to novelty detection in humans. *Journal of Clinical Neurophysiology*, *15*(1), 3–13. https://doi.org/10.1097/00004691-199801000-00003

Kok, P., Failing, M. F., & de Lange, F. P. (2014). Prior expectations evoke stimulus templates in the primary visual cortex. *Journal of Cognitive Neuroscience*, *26*(7), 1546–1554. https://doi.org/10.1162/jocn_a_00562

Kolossa, A., Kopp, B., & Fingscheidt, T. (2015). A computational analysis of the neural bases of Bayesian inference. *NeuroImage*, *106*, 222–237. https://doi.org/10.1016/j.neuroimage.2014.11.007

Kording, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS One*, *2*(9), e943. https://doi.org/10.1371/journal.pone.0000943

Kremlacek, J., Kreegipuu, K., Tales, A., Astikainen, P., Poldver, N., Naatanen, R., & Stefanics, G. (2016). Visual mismatch negativity (vMMN): A review and meta-analysis of studies in psychiatric and neurological disorders. *Cortex*, *80*, 76–112. https://doi.org/10.1016/j.cortex.2016.03.017

Kuldkepp, N., Kreegipuu, K., Raidvee, A., Naatanen, R., & Allik, J. (2013). Unattended and attended visual change detection of motion as indexed by event-related potentials and its behavioral correlates. *Frontiers in Human Neuroscience*, *7*, 476. https://doi.org/10.3389/fnhum.2013.00476

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

Linden, D. E. (2005). The p300: Where in the brain is it produced and what does it tell us? *The Neuroscientist*, *11*(6), 563–576. https://doi.org/10.1177/1073858405280524

Macaluso, E. (2010). Orienting of spatial attention and the interplay between the senses. *Cortex*, *46*(3), 282–297. https://doi.org/10.1016/j.cortex.2009.05.010

Macaluso, E., & Driver, J. (2005). Multisensory spatial interactions: A window onto functional integration in the human brain. *Trends in Neurosciences*, *28*(5), 264–271. https://doi.org/10.1016/j.tins.2005.03.008

Macdonald, M., & Campbell, K. (2011). Effects of a violation of an expected increase or decrease in intensity on detection of change within an auditory pattern. *Brain and Cognition*, *77*(3), 438–445. https://doi.org/10.1016/j.bandc.2011.08.014

Maheu, M., Dehaene, S., & Meyniel, F. (2019). Brain signatures of a multiscale process of sequence learning in humans. *eLife*, *8*, e41541. https://doi.org/10.7554/eLife.41541

Male, A. G., O'Shea, R. P., Schroger, E., Muller, D., Roeber, U., & Widmann, A. (2020). The quest for the genuine visual mismatch negativity (vMMN): Event-related potential indications of deviance detection for low-level visual features. *Psychophysiology*, *57*(6), e13576. https://doi.org/10.1111/psyp.13576

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Mars, R. B., Debener, S., Gladwin, T. E., Harrison, L. M., Haggard, P., Rothwell, J. C., & Bestmann, S. (2008). Trial-by-trial fluctuations in the event-related electroencephalogram reflect dynamic changes in the degree of surprise. *The Journal of Neuroscience*, *28*(47), 12539–12545. https://doi.org/10.1523/JNEUROSCI.2925-08.2008

Max, C., Widmann, A., Schroger, E., & Sussman, E. (2015). Effects of explicit knowledge and predictability on auditory distraction and target performance. *International Journal of Psychophysiology*, *98*(2 Pt 1), 174–181. https://doi.org/10.1016/j.ijpsycho.2015.09.006

May, P., Tiitinen, H., Ilmoniemi, R. J., Nyman, G., Taylor, J. G., & Naatanen, R. (1999). Frequency change detection in human auditory cortex. *Journal of Computational Neuroscience*, *6*(2), 99–120. https://doi.org/10.1023/a:1008896417606

May, P. J., & Tiitinen, H. (2010). Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology*, *47*(1), 66–122. https://doi.org/10.1111/j.1469-8986.2009.00856.x

Modirshanechi, A., Brea, J., & Gerstner, W. (2022). A taxonomy of surprise definitions. *Journal of Mathematical Psychology*, *110*, 102712.

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

Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences*, *13*(2), 201–233.

Naatanen, R., Gaillard, A. W., & Mantysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, *42*(4), 313–329. https://doi.org/10.1016/0001-6918(78)90006-9

Naatanen, R., Jacobsen, T., & Winkler, I. (2005). Memory-based or afferent processes in mismatch negativity (MMN): A review of the evidence. *Psychophysiology*, *42*(1), 25–32. https://doi.org/10.1111/j.1469-8986.2005.00256.x

Naatanen, R., & Näätänen, R. (1992). *Attention and brain function*. Psychology Press.

Naatanen, 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

Naeije, G., Vaulet, T., Wens, V., Marty, B., Goldman, S., & De Tiege, X. (2016). Multilevel cortical processing of somatosensory novelty: A magnetoencephalography study. *Frontiers in Human Neuroscience*, *10*, 259. https://doi.org/10.3389/fnhum.2016.00259

Naeije, G., Vaulet, T., Wens, V., Marty, B., Goldman, S., & De Tiege, X. (2018). Neural basis of early somatosensory change detection: A magnetoencephalography study. *Brain Topography*, *31*(2), 242–256. https://doi.org/10.1007/s10548-017-0591-x

Opitz, B., Rinne, T., Mecklinger, A., von Cramon, D. Y., & Schroger, 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

Ossmy, O., Moran, R., Pfeffer, T., Tsetsos, K., Usher, M., & Donner, T. H. (2013). The timescale of perceptual evidence integration can be adapted to the environment. *Current Biology*, *23*(11), 981–986. https://doi.org/10.1016/j.cub.2013.04.039

Ostwald, D., Spitzer, B., Guggenmos, M., Schmidt, T. T., Kiebel, S. J., & Blankenburg, F. (2012). Evidence for neural encoding of Bayesian surprise in human somatosensation. *NeuroImage*, *62*(1), 177–188. https://doi.org/10.1016/j.neuroimage.2012.04.050

Paavilainen, P. (2013). The mismatch-negativity (MMN) component of the auditory event-related potential to violations of abstract regularities: A review. *International Journal of Psychophysiology*, *88*(2), 109–123. https://doi.org/10.1016/j.ijpsycho.2013.03.015

Parmentier, F. B., Elsley, J. V., Andres, P., & Barcelo, F. (2011). Why are auditory novels distracting? Contrasting the roles of novelty, violation of expectation and stimulus change. *Cognition*, *119*(3), 374–380. https://doi.org/10.1016/j.cognition.2011.02.001

Pazo-Alvarez, P., Cadaveira, F., & Amenedo, E. (2003). MMN in the visual modality: A review. *Biological Psychology*, *63*(3), 199–236. https://doi.org/10.1016/s0301-0511(03)00049-8

Penny, W., Kiebel, S., & Friston, K. (2003). Variational Bayesian inference for fMRI time series. *NeuroImage*, *19*(3), 727–741. https://doi.org/10.1016/s1053-8119(03)00071-5

Penny, W. D., Stephan, K. E., Daunizeau, J., Rosa, M. J., Friston, K. J., Schofield, T. M., & Leff, A. P. (2010). Comparing families of dynamic causal models. *PLoS Computational Biology*, *6*(3), e1000709. https://doi.org/10.1371/journal.pcbi.1000709

Penny, W. D., Trujillo-Barreto, N. J., & Friston, K. J. (2005). Bayesian fMRI time series analysis with spatial priors. *NeuroImage*, *24*(2), 350–362. https://doi.org/10.1016/j.neuroimage.2004.08.034

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Prete, D. A., Heikoop, D., McGillivray, J. E., Reilly, J. P., & Trainor, L. J. (2022). The sound of silence: Predictive error responses to unexpected sound omission in adults. *The European Journal of Neuroscience*, *55*, 1972–1985. https://doi.org/10.1111/ejn.15660

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

Ritter, W., Sussman, E., Deacon, D., Cowan, N., & Vaughan, H. G., Jr. (1999). Two cognitive systems simultaneously prepared for opposite events. *Psychophysiology*, *36*(6), 835–838 Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/10554596

Rohe, T., Ehlis, A. C., & Noppeney, U. (2019). The neural dynamics of hierarchical Bayesian causal inference in multisensory perception. *Nature Communications*, *10*(1), 1907. https://doi.org/10.1038/s41467-019-09664-2

Rohe, T., & Noppeney, U. (2015). Cortical hierarchies perform Bayesian causal inference in multisensory perception. *PLoS Biology*, *13*(2), e1002073. https://doi.org/10.1371/journal.pbio.1002073

Runyan, C. A., Piasini, E., Panzeri, S., & Harvey, C. D. (2017). Distinct timescales of population coding across cortex. *Nature*, *548*(7665), 92–96. https://doi.org/10.1038/nature23020

Sabeti, M., Katebi, S. D., Rastgar, K., & Azimifar, Z. (2016). A multi-resolution approach to localize neural sources of P300 event-related brain potential. *Computer Methods and Programs in Biomedicine*, *133*, 155–168. https://doi.org/10.1016/j.cmpb.2016.05.013

Salisbury, D. F. (2012). Finding the missing stimulus mismatch negativity (MMN): Emitted MMN to violations of an auditory gestalt. *Psychophysiology*, *49*(4), 544–548. https://doi.org/10.1111/j.1469-8986.2011.01336.x

Sams, M., Alho, K., & Naatanen, R. (1983). Sequential effects on the ERP in discriminating two stimuli. *Biological Psychology*, *17*(1), 41–58. https://doi.org/10.1016/0301-0511(83)90065-0

Schroger, E. (1996). A neural mechanism for involuntary attention shifts to changes in auditory stimulation. *Journal of Cognitive Neuroscience*, *8*(6), 527–539. https://doi.org/10.1162/jocn.1996.8.6.527

Schroger, E., Marzecova, A., & SanMiguel, I. (2015). Attention and prediction in human audition: A lesson from cognitive psychophysiology. *The European Journal of Neuroscience*, *41*(5), 641–664. https://doi.org/10.1111/ejn.12816

Seer, C., Lange, F., Boos, M., Dengler, R., & Kopp, B. (2016). Prior probabilities modulate cortical surprise responses: A study of event-related potentials. *Brain and Cognition*, *106*, 78–89. https://doi.org/10.1016/j.bandc.2016.04.011

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., Rosen, B. R., & Tootell, R. B. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, *268*(5212), 889–893. https://doi.org/10.1126/science.7754376

Shams, L., & Beierholm, U. (2022). Bayesian causal inference: A unifying neuroscience theory. *Neuroscience and Biobehavioral Reviews*, *137*, 104619. https://doi.org/10.1016/j.neubiorev.2022.104619

Shams, L., Wozny, D. R., Kim, R., & Seitz, A. (2011). Influences of multisensory experience on subsequent unisensory processing. *Frontiers in Psychology*, *2*, 264. https://doi.org/10.3389/fpsyg.2011.00264

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423.

Shen, G., Smyk, N. J., Meltzoff, A. N., & Marshall, P. J. (2018). Using somatosensory mismatch responses as a window into somatotopic processing of tactile stimulation. *Psychophysiology*, *55*(5), e13030. https://doi.org/10.1111/psyp.13030

Shinozaki, N., Yabe, H., Sutoh, T., Hiruma, T., & Kaneko, S. (1998). Somatosensory automatic responses to deviant stimuli. *Brain Research. Cognitive Brain Research*, *7*(2), 165–171. https://doi.org/10.1016/s0926-6410(98)00020-2

Spackman, L. A., Boyd, S. G., & Towell, A. (2007). Effects of stimulus frequency and duration on somatosensory discrimination responses. *Experimental Brain Research*, *177*(1), 21–30. https://doi.org/10.1007/s00221-006-0650-0

Spackman, L. A., Towell, A., & Boyd, S. G. (2010). Somatosensory discrimination: An intracranial event-related potential study of children with refractory epilepsy. *Brain Research*, *1310*, 68–76. https://doi.org/10.1016/j.brainres.2009.10.072

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, *38*(4), 387–401. https://doi.org/10.1016/0013-4694(75)90263-1

Stefanics, G., Heinzle, J., Horvath, 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

Stefanics, G., Kremlacek, J., & Czigler, I. (2014). Visual mismatch negativity: A predictive coding view. *Frontiers in Human Neuroscience*, *8*, 666. https://doi.org/10.3389/fnhum.2014.00666

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., & Friston, K. J. (2009). Bayesian model selection for group studies. *NeuroImage*, *46*(4), 1004–1017. https://doi.org/10.1016/j.neuroimage.2009.03.025

Strasburger, H., Rentschler, I., & Juttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision*, *11*(5), 13. https://doi.org/10.1167/11.5.13

Sussman, E., Ritter, W., & Vaughan, H. G., Jr. (1998). Predictability of stimulus deviance and the mismatch negativity. *Neuroreport*, *9*(18), 4167–4170. https://doi.org/10.1097/00001756-199812210-00031

Sussman, E., Winkler, I., & Schroger, E. (2003). Top-down control over involuntary attention switching in the auditory modality. *Psychonomic*

Bulletin & Review, *10*(3), 630–637. https://doi.org/10.3758/bf03196525

Sussman, E. S. (2005). Integration and segregation in auditory scene analysis. *The Journal of the Acoustical Society of America*, *117*(3 Pt 1), 1285–1298. https://doi.org/10.1121/1.1854312

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, *150*(3700), 1187–1188. https://doi.org/10.1126/science.150.3700.1187

Sutton, S., Tueting, P., Zubin, J., & John, E. R. (1967). Information delivery and the sensory evoked potential. *Science*, *155*(3768), 1436–1439. https://doi.org/10.1126/science.155.3768.1436

Ulanovsky, N., Las, L., Farkas, D., & Nelken, I. (2004). Multiple time scales of adaptation in auditory cortex neurons. *The Journal of Neuroscience*, *24*(46), 10440–10453. https://doi.org/10.1523/JNEUROSCI.1905-04.2004

Ulanovsky, N., Las, L., & Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nature Neuroscience*, *6*(4), 391–398. https://doi.org/10.1038/nn1032

Urakawa, T., Inui, K., Yamashiro, K., & Kakigi, R. (2010). Cortical dynamics of the visual change detection process. *Psychophysiology*, *47*(5), 905–912. https://doi.org/10.1111/j.1469-8986.2010.00987.x

Von Helmholtz, H. (1867). *Handbuch der physiologischen Optik* (Vol. 9). Voss.

Vossel, S., Geng, J. J., & Fink, G. R. (2014). Dorsal and ventral attention systems: Distinct neural circuits but collaborative roles. *The Neuroscientist*, *20*(2), 150–159. https://doi.org/10.1177/1073858413494269

Wacongne, C., Changeux, J. P., & Dehaene, S. (2012). A neuronal model of predictive coding accounting for the mismatch negativity. *The Journal of Neuroscience*, *32*(11), 3665–3678. https://doi.org/10.1523/JNEUROSCI.5003-11.2012

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., Schobi, 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

Winkler, I. (2007). Interpreting the mismatch negativity. *Journal of Psychophysiology*, *21*(3–4), 147–163.

Winkler, I., & Czigler, I. (2012). Evidence from auditory and visual event-related potential (ERP) studies of deviance detection (MMN and vMMN) linking predictive coding theories and perceptual object representations. *International Journal of Psychophysiology*, *83*(2), 132–143. https://doi.org/10.1016/j.ijpsycho.2011.10.001

Winkler, I., Czigler, I., Sussman, E., Horvath, J., & Balazs, L. (2005). Preattentive binding of auditory and visual stimulus features. *Journal of Cognitive Neuroscience*, *17*(2), 320–339. https://doi.org/10.1162/0898929053124866

Winkler, I., Denham, S. L., & Nelken, I. (2009). Modeling the auditory scene: Predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences*, *13*(12), 532–540. https://doi.org/10.1016/j.tics.2009.09.003

Woolrich, M. W. (2012). Bayesian inference in FMRI. *NeuroImage*, *62*(2), 801–810. https://doi.org/10.1016/j.neuroimage.2011.10.047

Yabe, H., Tervaniemi, M., Reinikainen, K., & Naatanen, R. (1997). Temporal window of integration revealed by MMN to sound omission. *Neuroreport*, *8*(8), 1971–1974. https://doi.org/10.1097/00001756-199705260-00035

Yucel, G., McCarthy, G., & Belger, A. (2007). fMRI reveals that involuntary visual deviance processing is resource limited. *NeuroImage*, *34*(3), 1245–1252. https://doi.org/10.1016/j.neuroimage.2006.08.050

Zhao, C., Valentini, E., & Hu, L. (2015). Functional features of crossmodal mismatch responses. *Experimental Brain Research*, *233*(2), 617–629. https://doi.org/10.1007/s00221-014-4141-4

How to cite this article: Grundei, M., Schröder, P., Gijsen, S., & Blankenburg, F. (2023). EEG mismatch responses in a multimodal roving stimulus paradigm provide evidence for probabilistic inference across audition, somatosensation, and vision. *Human Brain Mapping*, 1–25. https://doi.org/10.1002/hbm.26303

Original Publication of Study 3

Grundei, M., Schmidt T. T., and Blankenburg, F. (2023). A multimodal cortical network of sensory expectation violation revealed by fMRI. *Human Brain Mapping, 44*(17), 5871-5891.

Received: 21 February 2023 | Revised: 4 July 2023 | Accepted: 29 August 2023

DOI: 10.1002/hbm.26482

RESEARCH ARTICLE

WILEY

A multimodal cortical network of sensory expectation violation revealed by fMRI

Miro Grundei[1,2] | Timo Torsten Schmidt[1] | Felix Blankenburg[1,2]

[1]Neurocomputation and Neuroimaging Unit, Freie Universität Berlin, Berlin, Germany

[2]Berlin School of Mind and Brain, Humboldt Universität zu Berlin, Berlin, Germany

Correspondence
Miro Grundei, Neurocomputation and Neuroimaging Unit, Freie Universität Berlin, 14195 Berlin, Germany.
Email: miro.grundei@fu-berlin.de

Funding information
Berlin School of Mind and Brain, Humboldt Universität zu Berlin

Abstract

The brain is subjected to multi-modal sensory information in an environment governed by statistical dependencies. Mismatch responses (MMRs), classically recorded with EEG, have provided valuable insights into the brain's processing of regularities and the generation of corresponding sensory predictions. Only few studies allow for comparisons of MMRs across multiple modalities in a simultaneous sensory stream and their corresponding cross-modal context sensitivity remains unknown. Here, we used a tri-modal version of the roving stimulus paradigm in fMRI to elicit MMRs in the auditory, somatosensory and visual modality. Participants ($N = 29$) were simultaneously presented with sequences of low and high intensity stimuli in each of the three senses while actively observing the tri-modal input stream and occasionally reporting the intensity of the previous stimulus in a prompted modality. The sequences were based on a probabilistic model, defining transition probabilities such that, for each modality, stimuli were more likely to repeat ($p = .825$) than change ($p = .175$) and stimulus intensities were equiprobable ($p = .5$). Moreover, each transition was conditional on the configuration of the other two modalities comprising global (cross-modal) predictive properties of the sequences. We identified a shared mismatch network of modality general inferior frontal and temporo-parietal areas as well as sensory areas, where the connectivity (psychophysiological interaction) between these regions was modulated during mismatch processing. Further, we found deviant responses within the network to be modulated by local stimulus repetition, which suggests highly comparable processing of expectation violation across modalities. Moreover, hierarchically higher regions of the mismatch network in the temporo-parietal area around the intraparietal sulcus were identified to signal cross-modal expectation violation. With the consistency of MMRs across audition, somatosensation and vision, our study provides insights into a shared cortical network of uni- and multi-modal expectation violation in response to sequence regularities.

KEYWORDS
connectivity, cross-modal, mismatch negativity, mismatch responses, multisensory, P3, predictive processing, statistical regularity

Miro Grundei and Timo Torsten Schmidt shared first authorship.

1 | INTRODUCTION

The brain is constantly subjected to a multi-modal stream of sensory inputs. As humans are encountering sensory information in an environment governed by statistical dependencies, the brain is engaging in probabilistic inference within and across sensory modalities (Barascud et al., 2016; Friston, 2005; Frost et al., 2015; Geisler, 2008; Gregory, 1980; Perruchet & Pacton, 2006; Summerfield & de Lange, 2014; Winkler et al., 2009). Neuronal mismatch responses (MMRs) to regularity violations such as the mismatch negativity (MMN; Näätänen et al., 1978, 2007) and the P3 (or P300; Polich, 2007; Squires et al., 1975; Sutton et al., 1965) have proven to provide valuable insights into the processing of probabilistic sensory input at various scales (Bekinschtein et al., 2009; Dehaene et al., 2015; Heilbron & Chait, 2018; Näätänen et al., 2001; Paavilainen, 2013; Schröger et al., 2014; Squires et al., 1975; Wacongne et al., 2011; Yaron et al., 2012). Although MMRs are among the most researched neural signatures (Näätänen et al., 2019; Polich, 2007), only few studies allow their direct comparison across sensory modalities and the mechanisms of modality specific and modality general MMRs to probabilistic *multi-modal* inputs are largely unknown.

If sensory regularities, for example, repeating (standard) stimuli or stimulus patterns, are occasionally violated by rare (deviant) stimuli, brain signals typically recorded with EEG in form of MMRs can be observed. The most well-known MMR is the auditory MMN, an early EEG component between ~100 and 200 ms post-stimulus onset which results from contrasting responses to deviant and standard stimuli. The MMN is elicited independent of attentional focus and task-related top-down processes (Alain & Woods, 1997; Näätänen et al., 1993; Ritter et al., 1999), even though attention can increase its amplitude (Trejo et al., 1995; Woldorff et al., 1991). Although primarily researched in the auditory modality, similar early MMN responses have been reported for other sensory modalities, including somatosensation (Andersen & Lundqvist, 2019; Gijsen et al., 2021; Hu et al., 2013; Kekoni et al., 1997; Ostwald et al., 2012; Shinozaki et al., 1998) and vision (Czigler, 2007; Kimura et al., 2011; Pazo-Alvarez et al., 2003; Stefanics et al., 2014). The major neuronal generator of the auditory MMN is found in the auditory cortex where the exact location depends on the eliciting sound features and their complexity (Alho, 1995; Giard et al., 1990; Molholm et al., 2005; Sabri et al., 2004), with additional contributions from inferior frontal cortex (Deouell, 2007; Doeller et al., 2003; Molholm et al., 2005; Opitz et al., 2002; Rinne et al., 2005; Shalgi & Deouell, 2007). Similarly, the neuronal sources underlying the somatosensory MMN are found in primary and secondary somatosensory cortices (Akatsuka, Wasaka, Nakata, Kida, Hoshiyama, et al., 2007; Akatsuka, Wasaka, Nakata, Kida, & Kakigi, 2007; Andersen & Lundqvist, 2019; Butler et al., 2012; Gijsen et al., 2021; Grundei et al., 2023; Naeije et al., 2016; Ostwald et al., 2012; Spackman et al., 2010), in combination with inferior frontal cortex (Allen et al., 2016; Chen et al., 2008; Fardo et al., 2017; Grundei et al., 2023; Ostwald et al., 2012). The combination of sensory and frontal sources is also indicated to underlie the visual MMN (Hedge et al., 2015; Iglesias et al., 2013; Kimura et al., 2011; Pazo-Alvarez et al., 2003; Yucel et al., 2007). Overall, ample evidence is pointing to modality-specific MMN generators in sensory regions and more modality-general contributions from (right inferior) prefrontal cortex.

A second well researched MMR is the P3, a later positive signal in response to stimulus deviance between 250 and 500 ms. The P3 is generally considered to be an attention-dependent response (Duncan et al., 2009; Duncan-Johnson & Donchin, 1982; Näätänen & Gaillard, 1983; Polich, 2007). While the earlier P3a sub-component is task-independent, drawing observers' attention to novel or unexpected stimuli (Escera et al., 2000; Friedman et al., 2001; Knight & Scabini, 1998), the later P3b response is more sensitive to task-related target stimuli (Duncan et al., 2009; Polich, 2007). Although extensively researched in the auditory domain, the P3 is known for its modality independent characteristics (Linden, 2005; Polich, 2007) and has been equivalently reported for somatosensation (Andersen & Lundqvist, 2019; Gijsen et al., 2021; Ostwald et al., 2012; Shen et al., 2018a, 2018b; Yamaguchi & Knight, 1991, 1992) and vision (Conill, 1998; Duncan et al., 2009; Picton, 1992; Zhang et al., 2022), and has been described across senses in response to multi-modal sequences (Grundei et al., 2023). The generating sources underlying the P3 are thought to be distributed in a fronto-parietal network, involving inferior frontal, anterior cingulate and temporo-parietal regions, with some indications for pronounced frontal contributions for the P3a and parietal dominance for the P3b (Linden, 2005; Polich, 2007). Thus, research on the P3 supports a modality-general role for fronto-parietal network activations in the processing of expectation violation and novelty alerting (Linden, 2005; Polich, 2007; Squires et al., 1975; Sutton et al., 1965).

Brain connectivity analyses based on fMRI and electrophysiological recordings support comparable network mechanisms underlying MMRs across modalities. Studies using dynamic causal models (DCM) indicate modulations in bidirectional connectivity in a fronto-sensory network hierarchy underlying MMRs in the auditory (Chennu et al., 2016; Garrido et al., 2007; Garrido et al., 2008; Garrido, Kilner, Kiebel, & Friston, 2009; Hughes et al., 2013; Phillips et al., 2015; Phillips et al., 2016) and somatosensory modality (Allen et al., 2016; Fardo et al., 2017), and propose that feedforward and feedback connections carry sensory errors and top-down expectations respectively. Similar mechanisms are hypothesized for visual MMRs (Stefanics et al., 2014). Moreover, parietal contributions to this network have been indicated for MMRs in somatosensation via DCM (Fardo et al., 2017) as well as to auditory MMRs via psychophysiological interaction (PPI) analyses in fMRI (Uhrig et al., 2014). Beyond the auditory modality only few studies have investigated MMR-related connectivity modulations (Allen et al., 2016; Fardo et al., 2017; Kellermann et al., 2017; Sherman et al., 2016), and findings for multi-modal inputs are largely lacking.

In a series of seminal articles, Downar and colleagues have investigated MMRs to multi-modal stimulus sequences (Downar et al., 2000, 2001, 2002). These studies provided first indications for a multi-modal mismatch network of sensory specific activation in

hierarchically higher sensory cortices and shared activations in inferior frontal and temporo-parietal regions in line with converging evidence from the auditory modality (Chennu et al., 2016; Dürschmid et al., 2016; El Karoui et al., 2015; Phillips et al., 2015; Phillips et al., 2016; Uhrig et al., 2014). A meta-analysis of neuroimaging studies corroborated these reports, revealing similar fronto-parietal MMR-related activation for the auditory and visual modality (Kim, 2014). Furthermore, current research on statistical learning points to domain-general computations underlying associative and probabilistic learning across different senses (Frost et al., 2015; Saffran & Thiessen, 2007). It has been shown that intraparietal and inferior frontal cortex encode the abstract structure of auditory and visual sequence patterns (Dehaene et al., 2015; Planton & Dehaene, 2021; Wang et al., 2015; Wang et al., 2019), highlighting their modality independent role during the extraction of regularities. Moreover, multimodal integration has been indicated during sequence processing (Bresciani et al., 2006, 2008), in terms of modulatory influences on the MMN (Besle et al., 2005; Butler et al., 2012; Friedel et al., 2020; Kiat, 2018; Zhao et al., 2015) and during multi-modal causal inference in a fronto-parietal network (Cao et al., 2019; Noppeney, 2021). Therefore, current research suggests a role for modality general fronto-parietal activation during probabilistic sensory inference and it is of great interest to characterize the processing of multi-modal statistical regularities and the underlying network in more detail.

An early finding, which has become a focus of the research on statistical sensory learning, is the modulation of the auditory MMN by standard repetition, that is, its amplitude increase with the number of prior standard presentations (Cowan et al., 1993; Haenschel et al., 2005; Imada et al., 1993; Sams et al., 1983). More recently, such modulation was shown beyond the auditory modality for both, MMN and P3 (Gijsen et al., 2021; Grundei et al., 2023). Although discussion persists about the contribution of stimulus specific adaptation in early sensory regions to these effects (Jääskelainen et al., 2004; May & Tiitinen, 2010), evidence converges to the view that increasing top-down modulations in response to the repeating stimulus accounts best for this observation (Auksztulewicz et al., 2017; Auksztulewicz & Friston, 2016; Baldeweg, 2006; Ewbank et al., 2011; Garrido, Kilner, Stephan, & Friston, 2009; Langner et al., 2011; Summerfield et al., 2008; Todorovic & de Lange, 2012), particularly as studies have highlighted the dependence of MMRs on stimulus expectation based on statistical regularities as opposed to mere changes in stimulus properties (Bendixen et al., 2012; Heilbron & Chait, 2018; Wacongne et al., 2011). Computational modeling of EEG dynamics has indicated that probabilistic learning of environmental statistics underlies MMRs in the auditory (Lecaignard et al., 2022; Lieder, Daunizeau, et al., 2013; Weber et al., 2020), somatosensory (Gijsen et al., 2021; Grundei et al., 2023; Ostwald et al., 2012) and visual modality (Stefanics et al., 2018). This view is supported by studies showing deviant responses to abstract rule violations (Näätänen et al., 2010; Paavilainen, 2013; Schröger et al., 2007), unexpected stimulus repetitions (Alain et al., 1999; Horvath & Winkler, 2004; Macdonald & Campbell, 2011; Nordby et al., 1988) and unexpected stimulus omissions (Bendixen et al., 2009; Chennu et al., 2016; SanMiguel et al., 2013; Suda et al., 2022; Yabe et al., 1997). Such key properties of the auditory MMN (Wacongne et al., 2012) have similarly been reported for the somatosensory (Andersen & Lundqvist, 2019; Naeije et al., 2018) and the visual MMN (Czigler, 2007; Czigler et al., 2006) as well as for the (auditory) P3 (Duncan et al., 2009; Prete et al., 2022). Furthermore, it has been proposed that the deviance detection system of MMN and P3 differentially responds to expectation violation to sequence regularities on different levels of complexity. Studies employing the "local–global" paradigm (Bekinschtein et al., 2009) in which stimulus sequences are defined by local regularities (e.g., the tendency of a stimulus to repeat) and additional global regularities (e.g., every fifth stimulus in a repeated sequence is a deviant) show that the MMN is only elicited by the local regularity violations whereas the later P3 is additionally sensitive to violations of the global deviant regularity (Bekinschtein et al., 2009; Chennu et al., 2013; Chennu et al., 2016; Dürschmid et al., 2016; El Karoui et al., 2015; King et al., 2014; Niedernhuber et al., 2022; Shirazibeheshti et al., 2018; Wacongne et al., 2011). Strikingly, this dichotomy for MMRs was recently shown to hold for the auditory, somatosensory and visual modality alike (Niedernhuber et al., 2022). Evidence converges to the view that the MMN, induced by local deviants primarily activates sensory regions, while the P3 MMR after global deviance is accompanied by frontal (Chao et al., 2018; Chennu et al., 2013; El Karoui et al., 2015) and fronto-parietal activations (Bekinschtein et al., 2009; Uhrig et al., 2014), in line with the neuronal sources thought to underlie the P3 (Linden, 2005; Polich, 2007). These results suggest that the P3 reflects expectation violation on a more global scale of sequence processing, indicating increasing levels of information integration in the hierarchy of a putative mismatch network.

Mechanistic accounts of universal principles of perception and perceptual learning in the brain, such as predictive processing (Clark, 2013; Friston, 2005, 2010), imply a modality independent role for MMRs reflecting error signals during expectation violation. Under such a view, the brain maintains a generative model of its environment which is continuously updated by comparing incoming sensory information with model predictions on different levels of hierarchical cortical organization (Friston, 2005; Kiebel et al., 2008; Mumford, 1992; Rao & Ballard, 1999). MMRs are interpreted as signatures of sensory prediction error in response to violations of top-down predictions (Friston, 2005; Garrido, Kilner, Stephan, & Friston, 2009; Winkler & Czigler, 2012). Predictive coding models can account for key features of MMRs (Heilbron & Chait, 2018; Lieder, Stephan, et al., 2013; Wacongne et al., 2012) and the dichotomy of MMN and P3 identified by the local–global paradigm is thought to reflect differential processing stages in a predictive hierarchy operating on different levels of complexity and information integration (Bekinschtein et al., 2009; Chennu et al., 2016; Dürschmid et al., 2016; Uhrig et al., 2014; Wacongne et al., 2011). Similarly, the repetition modulation of MMRs reflect prediction error responses scaled by an increasing sensory expectation to repeat the current stimulus train (Auksztulewicz & Friston, 2016; Baldeweg, 2006). The results of prior empirical work, showing interactions in the fronto-parietal and fronto-sensory hierarchy of the cortex during

mismatch processing across different modalities, are well in line with such mechanistic predictive processing accounts of MMR generation (Garrido, Kilner, Stephan, & Friston, 2009; Heilbron & Chait, 2018; Wacongne et al., 2012).

Overall, comparable dynamics of brain responses reflecting expectation violation within and across sensory modalities at different scales of complexity are of great interest for a mechanistic understanding of MMRs and fMRI investigations complementing the large body of work done in EEG remain rare. In the current study, we use a tri-modal version of the roving stimulus paradigm (Grundei et al., 2023) to elicit MMRs in fMRI. The paradigm allows to study responses to rare stimulus transitions independent of their equiprobable features (Baldeweg et al., 2004; Cowan et al., 1993; Garrido et al., 2008). As such, we take into account the consensus that MMRs reflect mismatching sensory expectations rather than stimulus properties as well as a suggested fundamental role for probabilistic representations of stimulus *transitions* in sequence perception and statistical learning (Dehaene et al., 2015; Maheu et al., 2019; Meyniel et al., 2016; Mittag et al., 2016). Based on a probabilistic model, we generated multi-modal sequences of low and high intensity stimuli for the auditory, somatosensory and visual modality which were governed by transition probabilities specifying cross-modal conditional dependencies. As each stimulus transition was conditional on the prior tri-modal stimulus configuration, the sequences exhibit global predictive properties in form of the multi-modal context. The aim of the study was to identify a mismatch network of modality specific sensory cortices and modality independent hubs of mismatch processing in frontal and parietal cortices, as suggested by previous research (e.g., Downar et al., 2000). Moreover, we intended to show equivalences of MMRs across the senses, particularly in terms of the modulation of deviant responses with increasing expectation established by a local stimulus train (Baldeweg, 2006; Haenschel et al., 2005). Finally, our manipulation of the global stimulus predictability based on the other senses was expected to reveal potential higher-level loci within the mismatch network sensitive cross-modal expectation violation.

2 | MATERIALS AND METHODS

Twenty-nine participants with no history of psychiatric or neurological disorders completed the experiment (14 female; 15 male; age range 18–43, M = 28, SD = 5.9). Prior to the experiment, written informed consent was obtained from each participant. The study was approved by the ethics committee at the Freie Universität Berlin (003/2021).

Participants underwent a multi-modal version of the roving stimulus paradigm. Our paradigm, depicted in Figure 1a, consisted of simultaneously presented bilateral auditory (A), somatosensory (S) and visual (V) stimuli, which each alternated between two different intensity levels ('low' and 'high'). The tri-modal stimulus sequences originated from a single probabilistic model resulting in different combinations of low and high stimuli across the three modalities in each trial. The sequence generation process is described in detail for an EEG study in Grundei et al. (2023). In short, it consists of a state s at time t evolving according to a Markov chain ($p(s_t|s_{t-1})$). Each state s represents 1 of 8 possible patterns as shown in Figure 1a with each state corresponding to a combination of three binary observations o. Therefore, each observation combination is conditional on the preceding observation combination ($p(o_{A,t}, o_{S,t}, o_{V,t}|o_{A,t-1}, o_{S,t-1}, o_{V,t-1})$). For each modality, the binary observations have equal probability of occurrence ($p = .5$) and stimulus transition probabilities were defined such that repetitions are more likely ($p = .825$) than changes ($p = .175$). For each modality, the overall configuration of change probabilities results in classic roving stimulus sequences, where trains of stimulus repetitions with different lengths alternate between the two stimulus intensities (depicted in Figure 1b). The range of train lengths was between 1 and 51 repetitions with an equal distribution across modalities (right skewed with expected value of 5 repetitions).

In each trial, for each of the three stimulus modalities, the other two modalities can be either both in low or high intensity (congruent) or one of them is low and the other high (incongruent). This property of the configuration of stimuli across the modalities was used to manipulate the predictability of stimulus transitions in the sequences. Three different types of stimulus sequences were generated with different settings (conditions C1, C2 and C3) determining the transition probabilities of each modality given the arrangement of the stimuli in the other two modalities (i.e., either congruent or incongruent; depicted in Figure 1c exemplified by a visual sequence for condition C1). Setting C1 defines higher change probability if the other two modalities are congruent (p(change|congruent) = 0.15, p(change|incongruent) = 0.025). Setting C2 defines lower change probability if the other two modalities are congruent (p(change|congruent) = 0.025, p(change|incongruent) = 0.15). Setting C3 defines equal change probability if the other two modalities are congruent or incongruent (p(change|congruent) = p(change|incongruent) = 0.0875). Per definition, the repetition probability follows the same principle such that for C1, p(repetition|congruent) = 0.85 and p(repetition|incongruent) = 0.975, for C2, p(repetition|congruent) = 0.975 and p(repetition|incongruent) = 0.85 and for C3, p(repetition|congruent) = p(repetition|incongruent) 0.925. Thus, in addition to the basic roving rule (i.e., repetitions are more likely than changes), for each modality, the settings C1 and C2 result in a tendency for stimuli to be more or less predicted in context of the two other modalities: In setting C1 a sequential stimulus tends to change more often (shows higher volatility) if accompanied by two congruent stimuli and tends to repeat more often (i.e., is more stable) if accompanied by two incongruent stimuli, and v.v. for C2. In other words, the multi-modal context predicts the tendency for volatile and stable phases of the stimulus sequence. Using a terminology along the lines of Arnal and Giraud (2012), the resulting stimulus transitions for each modality within the different sequences can therefore be defined as being rather *predicted* (here: higher transition probability conditional on congruency/incongruency of the other modalities), rather *mispredicted* (here: lower transition probability) or *unpredictable* (here: equal transition probability) in context of the tri-modal stimulus presentation. During each of 6 experimental runs, a sequence of stimuli with one of the three different probability settings (C1, C2, C3)

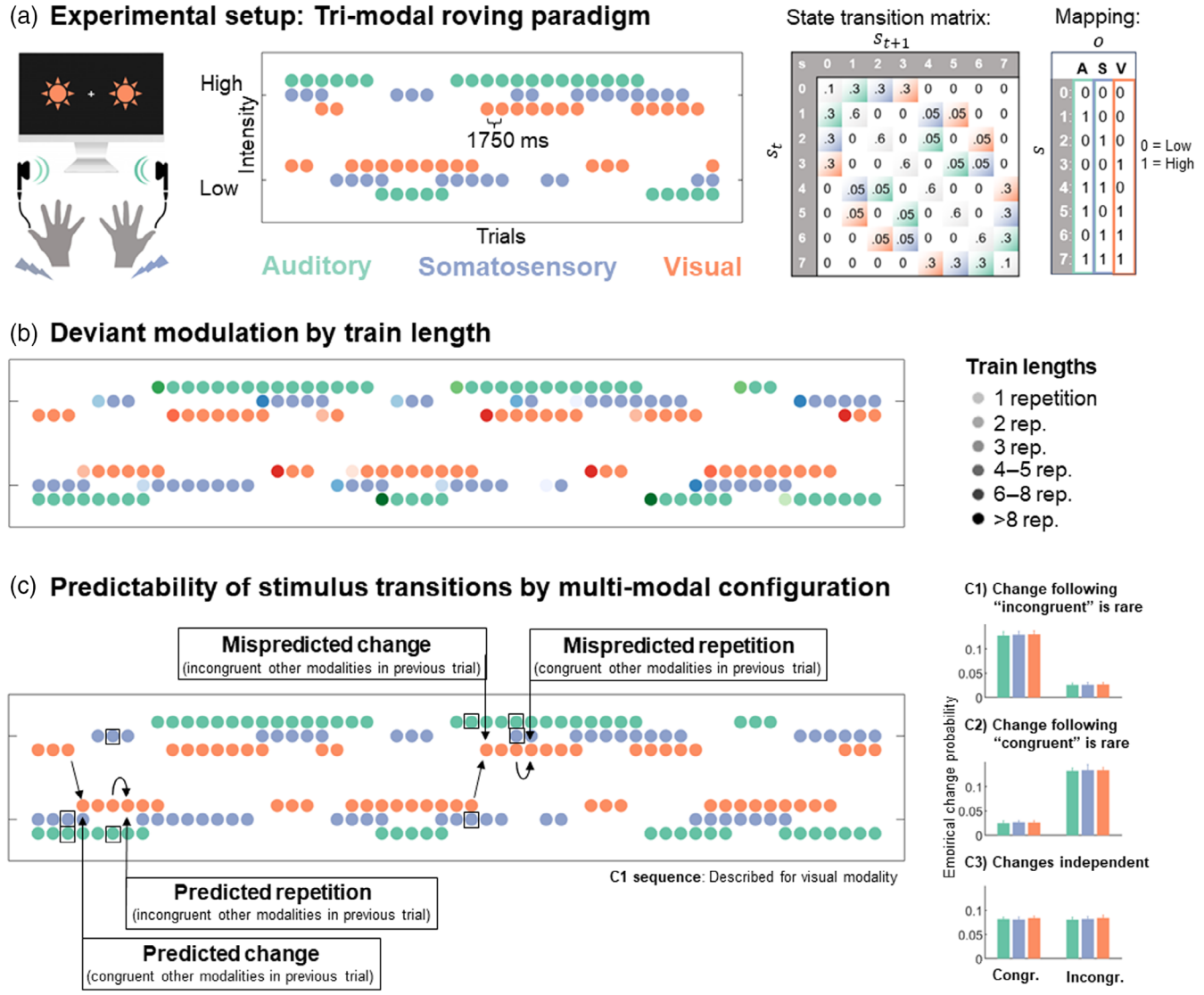


FIGURE 1 Experimental design. (a) Participants were presented with sequences of simultaneously presented bilateral auditory (A; green) "beep" stimuli, somatosensory (S; purple) electrical pulse stimuli and visual (V; orange) flash stimuli each at either low or high intensity. On consecutive trials, stimuli within each modality either repeated the previous stimulus intensity of that modality (standard) or alternated to the other intensity (deviant), corresponding to roving stimulus sequences for each modality (middle). Sequences were sampled according to a state transition matrix specifying the probabilities for states 0–7 at time $t + 1$ given the previous state s_t. Three different transition probability settings were used to define sequences for conditions C1, C2 and C3 (see Methods). Transition matrix here shown for probability setting C1. Light color shading depicts changes in a respective modality (A, S, V); light-gray diagonal defines the probability of tri-modal repetitions; white entries define the probability of multi-modal changes which are set to zero. Each state maps to a specific tri-modal observation (o) combination of low and high intensity stimuli (right table). (b) Deviant modulation by train length. Within the sequence of each modality stimulus repetitions form trains of standards of different lengths. The deviant following a specific standard train is labeled as falling in one of six categories depicted by color shading of the deviant (categories: 1, 2, 3, 4–5, 6–8, >8 repetitions). In our analyses we were interested in the modulation of the deviant response as a function of the standard train length preceding it. (c) Predictability of stimulus transitions by multi-modal configuration. Left: Exemplary shown for a visual sequence with probability setting C1. Right: Empirical change probabilities for sequences of conditions C1, C2 and C3. Bars depict the mean percentage of occurrence ±standard error for the three modalities. In C1, if two modalities are congruent (both low or both high) a change in the third modality is more likely than when the other modalities are incongruent with each other (one low and one high, or v.v.). In C2, if two modalities are congruent a change in the third modality is less likely. In C3, changes are equally likely if the other two modalities are congruent or incongruent.

was presented. For each participant unique sequences were sampled with randomly assigned conditions. The conditions were randomized across participants and a comparable number was presented overall, although condition C1 was presented more often (C1 = 63, C2 = 56, C3 = 55). In an alternative GLM setup presented in the Supplementary Materials which indicated highly comparable results (see Figure S1), the number of trials was balanced more rigorously, indicating that this slight imbalance in conditions is unlikely to have affected

our main findings. The empirical change probabilities present in the applied sequences were ensured to be representative of the true underlying probabilities (see Figure 1c). Sequences were resampled if deviations were deemed too large. The range of deviations was set to ±0.005 for the probability differences in congruent/incongruent conditions. For the overall change probability as well as the probability of low and high intensity stimuli this range was set to ±0.025. Finally, a minimum number of one repetition was ensured by discarding trials during the sampling process which corresponded to changes following changes. Participants were uninformed about the sequence probabilities and they were task irrelevant. Upon completion of the experiment, participants were debriefed and asked if they noticed any regularities in the sequences.

2.1 | Experimental setup

Each trial consisted of bilateral stimuli in three modalities (A, S, and V) that were presented simultaneously by triggering three instantaneous outputs of a data acquisition card (NI-USB 6343; National Instruments Corporation, Austin, Texas, USA) with an inter-stimulus interval (ISI) of 1750 ms. At each trial the device received and stored the corresponding stimulus waveform of each modality and released these to the stimulation devices upon a trigger signal, ensuring simultaneous stimulation.

Auditory stimuli were presented via in-ear MRI compatible headphones (Siemens Healthcare GmbH, Erlangen, Germany) to both ears. The MRI internal auditory system was set to maximum and received auditory inputs from the data acquisition card consisting of sinusoidal waves of 500 Hz and 100 ms duration modulated in their amplitudes by two different voltage factors. These were individually adjusted with the participants prior to the experiment to obtain two clearly perceivable and distinguishable intensities (mean intensity across participants ± SD: low = 0.29 ± 0.3 V, high = 1.58 ± 1.28 V). As a different set of headphones was used for the first four participants, these were not included in the average intensity calculation.

Somatosensory stimuli were administered with two DS5 isolated bipolar constant current stimulators (Digitimer Limited, Welwyn Garden City, Hertfordshire, UK) via adhesive electrodes (GVB-geliMED GmbH, Bad Segeberg, Germany) attached to the wrists of both arms to stimulate the median nerve. The stimuli consisted of electrical rectangular pulses of 0.2 ms duration, modulated by two different amplitudes. The two intensity levels were determined on an individual basis to obtain two clearly perceivable and distinguishable intensities (mean intensity across participants ± SD: low = 4.47 ± 0.99 mA, high = 7.47 ± 1.56 mA).

Visual stimuli were presented via light emitting diodes (LEDs) and transmitted through optical fiber cables (Loptec GmbH, Berlin, Germany). The LEDs were mounted 10 cm to the left and to the right of a fixation cross along the horizontal meridian (10°, eccentricity) presented on a display board at the base of the magnet bore at approximately 110 cm. The visual flashes consisted of rectangular waves of 100 ms duration which were modulated by two different voltage amplitudes (low intensity stimulus: 2.65 V, corresponding to approximately 0.4 lux; high intensity stimulus: 10 V, corresponding to approximately 91.5 lux). The visual stimuli were determined to be clearly perceivable and distinguishable by all participants so that no individual intensity adjustments were applied.

In each of 6 experimental runs, a sequence of 400 trials was presented. To ensure that participants maintained attention throughout the experiment and to encourage monitoring of all three stimulation modalities, participants were instructed to respond to occasional target questions (catch trials) via button presses with the foot. In six trials (<1%), randomly placed within each run, the fixation cross changed to one of the letters A, T or V followed by a question mark. This prompted participants to report if the most recent stimulus (before the appearance of the letter) in the auditory (letter A), somatosensory (letter T for "tactile") or visual (letter V) modality was presented with low or high intensity. To minimize motion during responses, the hallux of the right foot was used by participants to press either a left or a right button on an MRI compatible button-box, and the button assignment (left = low/right = high or left = high/right = low) was counterbalanced across participants.

2.2 | fMRI data acquisition and preprocessing

Functional MRI data was acquired in 6 runs of 11.9 min on a 3 T Magnetom Prisma Fit Scanner (Siemens Healthcare GmbH, Erlangen, Germany) at the Center for Cognitive Neuroscience Berlin (CCNB), using a 64 channel head coil. Four hundred and seventy-five functional volumes were acquired per run using a T2*-weighted gradient-echo EPI multiband 1 sequence (SMS factor = 3), with interleaved acquisition order and whole brain coverage (TR = 1.5 s; TE = 33 ms; 2.5 × 2.5 × 2.5 mm^3 voxel; matrix size = 80 × 80, FOV = 200 mm, flip angle = 70°; 48 slices; gap = 10%). Additionally, a T1-weighted MPRAGE with 208 sagittal slices, TR = 1930 ms, TE = 3.52 ms, 0.8 × 0.8 × 0.8 mm^3 voxel size was acquired.

FMRI data were pre-processed with SPM12 (Wellcome Trust Centre for Neuroimaging, Institute for Neurology, University College London, UK). Functional data were realigned to the mean image, normalized to MNI space using unified segmentation, interpolated to 2 × 2 × 2 mm^3 voxel size, spatially smoothed with an 8 mm FWHM Gaussian kernel, and temporally detrended (Macey et al., 2004).

2.3 | GLM analyses

Statistical analysis was performed according to a standard general linear model (GLM) approach with SPM12. For the experimental conditions, two separate first-level GLMs were computed for each participant. Each analysis was applied on the full data set and all computed first level GLMs contained regressors of no interest for the motion parameters.

One first-level GLM comprised a regressor modeling the onsets of all trials complemented with parametric regressors for each

modality coding the trials with 1 and −1 for intensity (High > Low), mismatch responses (Deviants > Standards) and cross-modal predictability (Mispredicted > Predicted). Correspondingly, three first-level contrasts for each modality were computed. Please note that, contrary to conventional MMN analyses in EEG which only consider the pre-deviant standard stimulus, in our analysis all standards were modeled in the GLM to avoid collinearity in the contrasts.

An additional subject-level GLM was computed to test for parametric effects of deviant responses, dependent on the number of repetitions of standards before the deviant (train length; Figure 1b). We binned the deviant trials into six categories of train length (repetitions before the deviant: 1, 2, 3, 4–5, 6–8, >8) and created a model with a separate regressor for each of the train length categories for each modality. Computing separate contrast estimates for each corresponding train length allowed us to compute a linear contrast on the second level and plot the respective contrast estimates (shown in Results Figure 3).

The second level GLM analyses were performed as ANOVA models in terms of the flexible factorial design specification implemented in SPM. First-level contrast images for each modality with one factor coding for modality, one factor coding for experimental condition were included supplemented with a subject factor. All second level conjunction analyses were computed as a conjunction against the global null hypothesis as implemented in SPM (Friston et al., 1999; Friston et al., 2005). Activation clusters were labeled according to anatomical and functional assignments provided by the cytoarchitectonic maps of the SPM Anatomy (Eickhoff et al., 2005).

2.4 | Psychophysiological-interaction analyses

To model changes in connectivity, we used PPI analyses as implemented in SPM (Friston et al., 1997). PPIs indicate if the contribution of one brain area to another changes significantly with an experimental factor and, as such, can be viewed as an event-related connectivity measure. We tested for PPI-connectivity changes during mismatch responses from seed regions in sensory cortices to the remaining brain voxels. The seed regions were based on the analysis of mismatch responses described above. From the seed regions we extracted time-series data from volumes of interest defined as 8 mm radius spheres around peak voxels identified in the GLM analysis (somatosens: left $x = -62$, $y = -16$, $z = 16$; right: $x = 58$, $y = -18$, $z = 22$; auditory: left $x = -60$, $y = -42$, $z = 12$; right: $x = 62$, $y = -38$, $z = 8$; visual: left $x = -44$, $y = -68$, $z = 2$; right: $x = 46$, $y = -58$, $z = 14$). Following SPM's implementation of PPI analyses, for each modality, the interaction of the extracted source signal with the respective Deviants > Standards regressor of that modality was formed and included in a first level GLM analysis together with the source region's signal and all remaining experimental regressors. Subsequently, first level contrasts were computed for the interaction regressors and included in a second level GLM which included all PPI estimates (left and right of all modalities) and a subject factor.

3 | RESULTS

3.1 | Behavioral results

Participants performances in the "catch-trials" indicates their ability to globally maintain their attention to the tri-modal stimulus stream. Of the 70.9 ± 21% (M ± SD) responses that were given within the short response window (of 2.3 s), 73.5 ± 15.6% were correct with an average reaction time of 1.49 ± 0.21 s. One-way repeated measures ANOVAs indicated no difference between modalities for response evaluation ($F(2,50) = 0.02$, $p = .98$) or reaction time ($F(2,50) = 0.75$, $p = .48$). Exclusion of a sample of participants ($n = 5$) showing bad response performance (who responded to ≤ 50% of questions in ≥ 50% of experimental runs) resulted in virtually identical fMRI results as presented below and no participants were removed from the analysis. Upon completion of the experiment, participants were debriefed and none of the participants identified the cross-modal regularities in the sequences.

3.2 | Mismatch responses across modalities

To reveal activation related to MMRs for each modality we computed the contrast Deviants > Standards against the same contrasts in the other two modalities, to delineate sensory specific activations. The results are presented in Figure 2, thresholded with $p < .05$ FWE corrected on the peak level and unthresholded SPMs of all results are available at https://www.neurovault.org/collections/LECDZXPI.

The auditory cluster spans bilaterally across the superior temporal gyrus (STG). Strongest activity was found in STG regions identified as auditory association areas in temporal cortex TE3 (right peak: $x = 66$, $y = -32$, $z = 4$, t-value = 15.4; left peak: $x = -64$, $y = -38$, $z = 12$, t-value = 13.5), temporal area TE4 in the upper bank of the superior temporal sulcus STS1 (right peak: $x = 60$, $y = -20$, $z = 0$, t-value = 13.8; left peak: $x = -62$, $y = -22$, $z = 2$, t-value = 10.1) and temporal area TE5 in the lower bank of the superior temporal sulcus STS2 (right peak: $x = 54$, $y = 2$, $z = -12$, t-value = 10.7; left peak: $x = -50$, $y = -12$, $z = -8$, t-value = 5.4). Moreover, some extensions into planum temporale and Heschl's gyrus were found on the left, encompassing primary auditory cortex in temporal area TE1 (left peak: $x = -36$, $y = -32$, $z = 8$, t-value = 8.6).

The somatosensory cluster extends bilaterally across the postcentral gyrus and the operculum. Strongest activity was found in opercular cortex OP1, functionally identified as the secondary somatosensory cortex (SII; right peak: $x = 48$, $y = -18$, $z = 18$, t-value = 10.2; left peak: $x = -58$, $y = -18$, $z = 16$, t-value = 9.2), as well as the insular cortex (IC/Id4; right peak: $x = 38$, $y = 0$, $z = 10$, t-value = 9.5; left peak: $x = -40$, $y = -6$, $z = 10$, t-value = 7.9). The functional regions in postcentral gyrus were attributed to the primary somatosensory cortex (SI) in Brodmann areas BA2 (right peak: $x = 26$, $y = -40$, $z = 68$, t-value = 7.4; left peak: $x = -24$, $y = -42$, $z = 66$, t-value = 6.7) and BA3b (right peak: $x = 48$, $y = -16$, $z = 36$, t-value = 6.2; left peak: $x = -50$, $y = -16$, $z = 32$, t-value = 5.5).

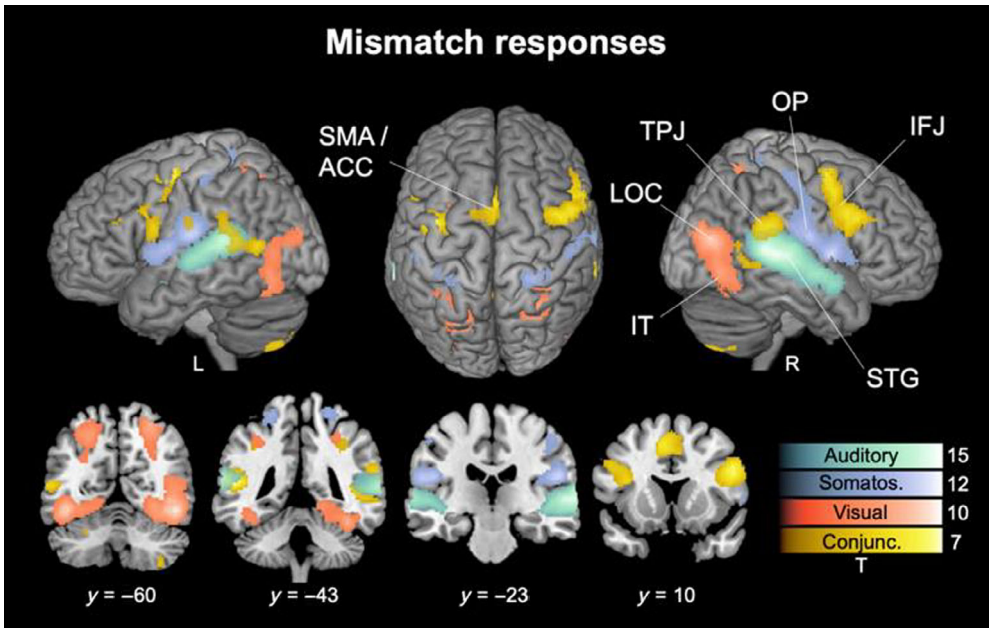


FIGURE 2 Mismatch responses. Contrasts of Deviants > Standards are shown for the auditory (green), somatosensory (purple) and visual (orange) modality as well as their conjunction (yellow), $p < .05$ FWE corrected on the peak level. Unthresholded SPMs are available at https://www.neurovault.org/collections/LECDZXPI. Abbreviations of region labels: IFJ: inferior frontal junction; IT: inferior temporal cortex; LOC: lateral occipital cortex; OP: Opercular cortex; SMA/ACC: (anterior) supplementary motor area/ anterior cingulate gyrus; STG: superior temporal gyrus.

The visual cluster extends bilaterally across the lateral occipital cortex (LOC) and inferior temporal cortex (IT) around the fusiform gyrus (FG). Strongest activity was found in FG (FG2; right peak: $x = 46$, $y = -58$, $z = -14$, t-value = 11.9; left peak: $x = -44$, $y = -64$, $z = -12$, t-value = 12.2) as well as LOC containing functional regions identified as higher order visual areas V4 (hOc4; right peak: $x = 44$, $y = -74$, $z = 14$, t-value = 10.6) and V5 (hOc5; left peak: $x = -44$, $y = -68$, $z = 2$, t-value = 7.5). Additional activation was found around the calcarine sulcus identified as visual areas V1 (hOc1; right peak: $x = 6$, $y = -80$, $z = 6$, t-value = 5) and V2 (hOc2; right peak: $x = 8$, $y = -84$, $z = 12$, t-value = 5.1).

The conjunction contrast across modalities shows bilateral clusters around the inferior frontal junction (IFJ) in the inferior frontal gyrus (IFG; right peak: $x = 50$, $y = 12$, $z = 24$, t-value = 6.7; left peak: $x = -50$, $y = -8$, $z = 26$, t-value = 5.3) and middle frontal gyrus (MFG; right peak: $x = 42$, $y = 0$, $z = 52$, t-value = 5.8; left peak: $x = -40$, $y = -4$, $z = 52$, t-value = 6.9), and is most pronounced on the right side where it extends into frontal operculum and IC (Id7; right peak: $x = 32$, $y = 26$, $z = 2$, t-value = 2.7). Additional pronounced conjunction effects were found across the anterior portion of the supplementary motor area, extending into the anterior cingulate gyrus (SMA/ACC; left peak: $x = -6$, $y = 10$, $z = 50$, t-value = 7.1), as well as bilaterally at the intersection of supramarginal gyrus (SMG), angular gyrus (AG) and superior and middle temporal gyrus (MTG) around the temporo-parietal junction (TPJ; right peak: $x = 54$, $y = -40$, $z = 22$, t-value = 5.2; left peak: $x = -54$, $y = -44$, $z = 10$, t-value = 6.9).

3.3 | Mismatch modulated connectivity

Results of seed-based PPI-connectivity analyses are presented in Figure 3, thresholded with $p < .05$ FWE corrected on the cluster level, showing connectivity increases from the three respective bilateral sensory seed regions to the rest of the brain, modulated by the mismatch contrast Deviants > Standards. Seed-regions were in modality-specific higher order sensory cortices of both hemispheres based on the strongest effects of the Deviants > Standards contrast presented above (see Figure 2), consisting of STG (auditory; left: $x = -60$, $y = -42$, $z = 12$; right: $x = 62$, $y = -38$, $z = 8$), OP/SII (somatosensory; left: $x = -58$, $y = -18$, $z = 22$; right: $x = 58$, $y = -18$, $z = 22$) and IT cortex (visual; left: $x = -44$, $y = -68$, $z = 2$; right: $x = 46$, $y = -58$, $z = -14$).

The conjunction contrast revealed a common increase in connectivity with brain areas found in the extended mismatch network comprised of bilateral frontal and temporo-parietal regions with pronounced clusters on the right hemisphere. The frontal clusters are located bilaterally across the IFJ, including MFG (right peak: $x = 44$, $y = 28$, $z = 22$, t-value = 3.6; left peak: $x = -42$, $y = 4$, $z = 36$, t-value = 3.6) and IFG (right peak: $x = 50$, $y = 12$, $z = 30$, t-value = 3.9; left peak: $x = -40$, $y = 28$, $z = 20$, t-value = 3.4), with extensions into IC (Id6; right peak: $x = 34$, $y = 20$, $z = 0$, t-value = 3.7; left peak: $x = -34$, $y = 14$, $z = -2$, t-value = 3) and right frontal pole (FP) (peak: $x = 42$, $y = 42$, $z = 22$, t-value = 4.1). The parietal clusters locate around the intraparietal sulcus (IPS; right peak: $x = 36$, $y = -52$, $z = 50$, t-value = 4.6; left peak: $x = -32$, $y = -56$, $z = 46$, t-value = 3.6), the TPJ (right peak: $x = 54$, $y = 38$, $z = 18$, t-value = 3.2; left peak: $x = -56$, $y = -42$, $z = 14$, t-value = 2.5) as well as precuneus (right peak: $x = 8$, $y = -64$, $z = 44$, t-value = 3.4; left peak: $x = -4$, $y = -62$, $z = 50$, t-value = 1.8).

3.4 | Modulation of deviant response by train length

The contrast of a parametric deviant increase with train length (defined by the number of standards presented prior to the deviant)

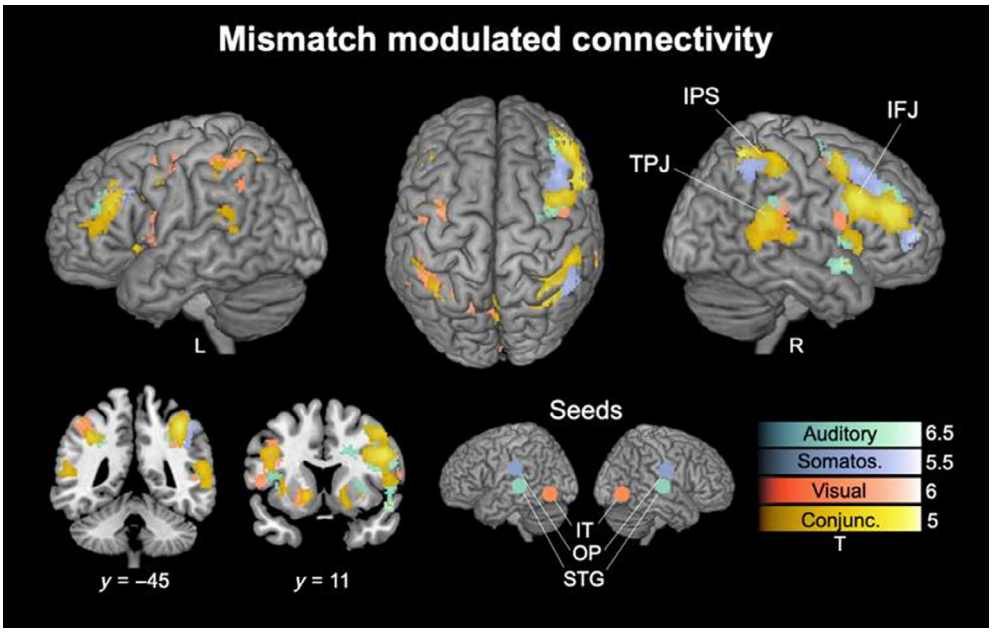


FIGURE 3 Mismatch modulated connectivity. Seed-based psychophysiological interaction (PPI) connectivity analyses show connectivity modulations by the Deviants > Standards contrast from the three respective sensory seed regions to the rest of the brain. Clusters are shown for the auditory (green), somatosensory (purple) and visual (orange) modality and their conjunction (yellow). $p < .05$ FWE corrected on the cluster level. Unthresholded SPMs are available at https://www.neurovault.org/collections/LECDZXPI. Abbreviations of region labels: IFJ: inferior frontal junction; IPS: intraparietal sulcus; IT: inferior temporal cortex; OP: opercular cortex; STG: superior temporal cortex; TPJ: temporo-parietal junction.

for each modality is shown in Figure 4, thresholded with $p < .05$ FWE corrected on the cluster level. Significant clusters largely overlap with the effects in the mismatch network described above (Figure 2).

The auditory clusters are found bilaterally in STG with strongest activity in regions identified as auditory association areas TE3 (right peak: $x = 66$, $y = -30$, $z = 6$, t-value = 12.5; left peak: $x = -62$, $y = -36$, $z = 10$, t-value = 8.8) and TE4 (right peak: $x = 52$, $y = -22$, $z = -2$, t-value = 7; left peak: $x = -60$, $y = -16$, $z = 0$, t-value = 6.7) as well as the precuneus (right peak: $x = 10$, $y = -74$, $z = -40$, t-value = 4.6; left peak: $x = -6$, $y = -72$, $z = 46$, t-value = 4.9).

The somatosensory clusters primarily extend across the OP cortex including OP1, functionally corresponding to SII (right peak: $x = 56$, $y = -16$, $z = 20$, t-value = 8.2; left peak: $x = -58$, $y = -16$, $z = 20$, t-value = 7.1), and the IC (Id4; right peak: $x = 40$, $y = -2$, $z = 10$, t-value = 7.6; left peak: $x = -38$, $y = -6$, $z = 10$, t-value = 5.2). Additional activation was found in left postcentral sulcus identified as BA 2, containing SI (left peak: $x = -48$, $y = -26$, $z = 42$, t-value = 4.3) and around the SMA/ACC (left peak: $x = -2$, $y = 6$, $z = 36$, t-value = 4.3).

The visual clusters were found bilaterally in IT and FG (FG3; right peak: $x = 32$, $y = -54$, $z = -14$, t-value = 5.2; left peak: $x = -36$, $y = -56$, $z = -8$, t-value = 4.4), and most pronounced in LOC regions functionally corresponding to higher order visual areas V4 (hOc4; right peak: $x = 48$, $y = -74$, $z = 8$, t-value = 5.8; left peak: $x = -36$, $y = -80$, $z = 14$, t-value = 4.2) and V5 (hOc5; right peak: $x = 42$, $y = -64$, $z = -2$, t-value = 5.8; left peak: $x = -48$, $y = -70$, $z = 4$, t-value = 5.5). Additional clusters are observed in the calcarine sulcus identified as early visual areas V1/V2 (hOc1/hOc2; right peak: $x = 4$, $y = -78$, $z = -12$, t-value = 4.6; left peak: $x = -4$, $y = -86$, $z = 10$, t-value = 4.2) as well as round the right IFJ (right peak: $x = 40$, $y = 8$, $z = 26$, t-value = 4.7).

The conjunction contrast across modalities shows most pronounced activation bilaterally around the IFJ, that is, IFG (right peak: $x = 42$, $y = 4$, $z = 22$, t-value = 2.7; left peak: $x = -46$, $y = 4$, $z = 30$, t-value = 2.3) and MFG (right peak: $x = 48$, $y = 4$, $z = 42$, t-value = 2.1; left peak: $x = -46$, $y = 2$, $z = 34$, t-value = 2.2), as well as the TPJ (right peak: $x = 60$, $y = -42$, $z = 18$, t-value = 2.4; left peak: $x = -54$, $y = -38$, $z = 18$, t-value = 2.9). Additional clusters are found around the SMA/ACC (right peak: $x = 4$, $y = 4$, $z = 56$, t-value = 2).

Since the train lengths of a given modality are longer if accompanied by congruent other modalities in condition C1, and v.v. in C2, we performed an additional train length analysis using only trials of the unpredictable condition C3 (corresponding to around 1/3 of trials) with identical distributions of train lengths across congruent and incongruent trials. Although less significant ($p < .001$ uncorrected; likely due to the reduced number of trials), the results for trials of C3 (not shown) revealed the same main clusters sensitive to parametric increase in deviant responses as presented in Figure 4.

3.5 | Cross-modal predictability

Activations in response to mispredicted stimuli (with respect to the multi-modal stimulus configuration) were identified by contrasting

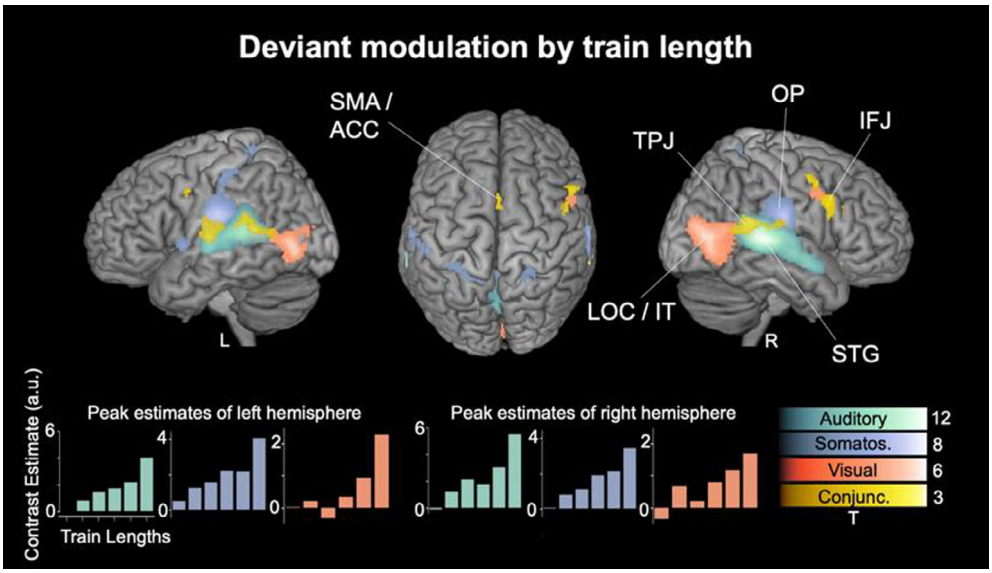


FIGURE 4 Modulation of deviant responses by prior repetition of standards. Parametric modulation of BOLD activity across six levels of stimulus train length (number of standards before the deviant; binned repetitions: 1, 2, 3, 4–5, 6–8, >8) for the auditory (green) somatosensory (purple) and visual (orange) modality and their conjunction (yellow). $p < .05$ FWE corrected on the cluster level. Unthresholded SPMs are available at https://www.neurovault.org/collections/LECDZXPI. Bottom row: The respective contrast estimates of the train length contrasts at the peak voxel of the main cluster of each modality (right and left). Abbreviations of region labels: IFJ: inferior frontal junction; IT: inferior temporal cortex; LOC: Lateral occipital cortex; OP: opercular cortex; SMA/ACC: (anterior) supplementary motor area/anterior cingulate cortex; STG: superior temporal cortex; TPJ: Temporo-parietal junction.

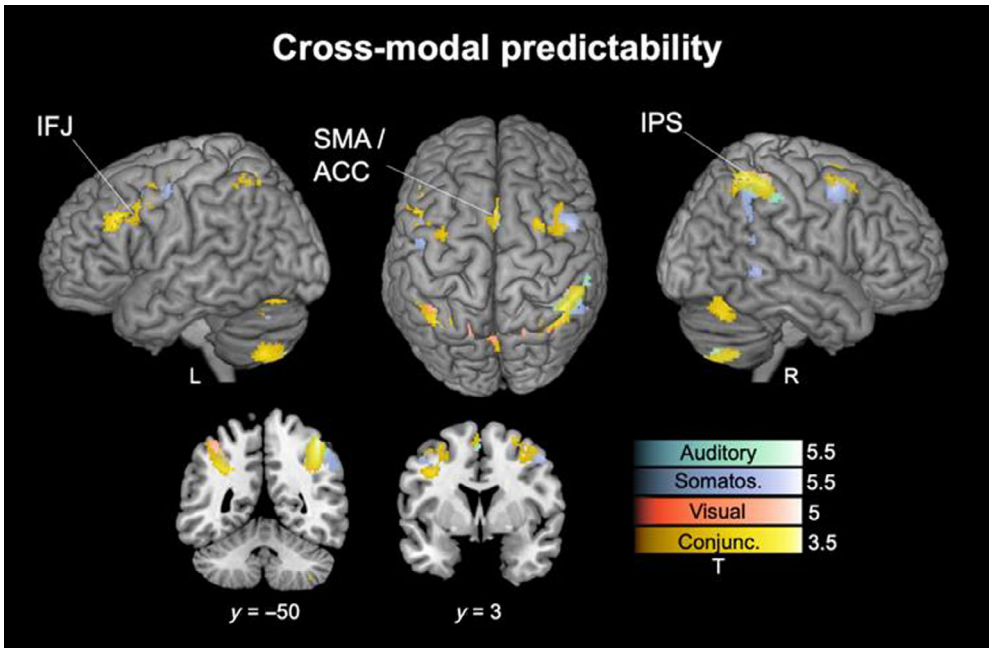


FIGURE 5 Cross-modal expectation violation. Significant clusters of activation for the predictability regressor (contrasting Mispredicted > Predicted trials) for the auditory (green), somatosensory (purple) and visual (orange) modality and their conjunction (yellow). $p < .05$ FWE corrected on the cluster level. Unthresholded SPMs are available at https://www.neurovault.org/collections/LECDZXPI. Abbreviations of region labels: IFJ: inferior frontal junction; IPS: intraparietal sulcus; SMA/ACC: (anterior) supplementary motor area/anterior cingulate cortex.

Mispredicted > Predicted trials for each modality. The contrast revealed increased activity in frontal and parietal clusters which are shown in Figure 5, thresholded with $p < .05$ FWE corrected on the cluster level. The clusters show overlap with clusters of the extended mismatch network described above. The auditory, somatosensory and visual modality showed the most pronounced clusters around the IPS, identified as a dorsal extension of the TPJ: The conjunction contrast across modalities shows bilateral clusters at the intersection of AG (right peak: $x = 42$, $y = -50$, $z = 54$, t-value $= 3.4$; left peak: $x = -30$, $y = -54$, $z = 34$, t-value $= 2.6$) and SMG (right peak: $x = 50$, $y = -42$, $z = 52$, t-value $= 2.9$; left peak: $x = -52$, $y = -44$, $z = 38$, t-value $= 2.3$) with extension into precuneus on the right (peak: $x = 16$, $y = -64$, $z = 48$, t-value $= 2.5$). Additional clusters were found in the SMA/ACC (left peak: $x = -2$, $y = 12$, $z = 50$, t-

value = 2.8) and bilaterally around the IFJ, primarily in MFG (right peak: $x = 40$, $y = 0$, $z = 54$, t-value = 1.9; left peak: $x = -48$, $y = 26$, $z = 30$, t-value = 2.9).

4 | DISCUSSION

Using a tri-modal version of the roving stimulus paradigm in combination with fMRI, we induced MMRs in the auditory, somatosensory and visual modality corresponding to modality specific activation in sensory cortices and modality independent clusters of activation in inferior frontal and temporo-parietal cortex. In addition to confirming initial fMRI work on multi-modal MMRs (Downar et al., 2000), our results expand the previous description by showing deviance related modulation of PPI-connectivity from each sensory region to the modality independent hubs of the putative cortical mismatch network. Moreover, across modalities, we showed increasing deviant responses within the identified network with prior standard repetition, most pronounced in higher order sensory regions. Strikingly, our novel experimental manipulation of cross-modal stimulus predictability revealed a parietal contribution to mismatch processing selectively sensitive to cross-modal regularity violation.

4.1 | Modality specific activations in sensory cortices

In accordance with prior research, mismatch effects specific to the auditory input sequence were found in the STG (Doeller et al., 2003; Downar et al., 2000; Garrido, Kilner, Stephan, & Friston, 2009; Molholm et al., 2005; Näätänen et al., 2005; Opitz et al., 2002; Rinne et al., 2005; Yucel et al., 2005). Interestingly, the most pronounced activation was observed bilaterally in higher order auditory processing areas with only little overlap with primary auditory cortex, which is often specified in EEG source modeling as a separate source in Heschl's gyrus (e.g., Garrido, Kilner, Stephan, & Friston, 2009). In contrast, here we find sensory specific activations in secondary auditory areas across the STG (TE3) and in the upper (STS1/TE4) and lower bank of the superior temporal sulcus (STS2/TE5) which are considered high level auditory processing regions (Zachlod et al., 2020). Our results are in accordance with a recent comparative fMRI study showing STG (TE3) activation for MMR to intensity changes among other stimulus features (Zvyagintsev et al., 2020). Of the few studies, which have used the roving stimulus paradigm in fMRI, one showed activations of higher order auditory regions in STG for pattern ("what") and location ("where") changes (Altmann et al., 2007). Similarly, another study differentially located responses to duration and frequency deviants in STG as well as within inferior frontal and posterior parietal regions (Molholm et al., 2005). The authors find a tendency for left lateralization of duration deviants (temporal information) and right lateralization for frequency deviants (tonal information). Here, we supplement these findings with intensity changes in roving stimulus sequences resulting in sensory specific activations in higher order auditory and parietal regions with right hemispheric dominance. While it is obvious that primary auditory cortex contributes to the MMR as it receives sensory input signals, our results suggest that the primary generator signaling sensory specific mismatch lies in non-primary auditory regions in STG. These results are in line with DCM studies modeling the STG as the intermediate stage of MMR processing, receiving feedforward input from primary auditory and feedback input from pre-frontal cortex (Auksztulewicz & Friston, 2015; Chennu et al., 2016; Garrido et al., 2007; Garrido et al., 2008; Garrido, Kilner, Kiebel, & Friston, 2009; Phillips et al., 2015; Phillips et al., 2016) and support their interpretations of auditory MMRs reflecting prediction error signals in response to violation of top-down predictions.

Mismatch effects specific to the somatosensory sequence were observed in SI and SII extending into IC. Most pronounced activation was found in SII and activation in SI was primarily found in BA1 and BA2 which are not major input regions (Purves et al., 2008). Previous research on somatosensory MMRs was primarily done with EEG and MEG, where source modeling suggests underlying neuronal generators in SI and SII (Akatsuka, Wasaka, Nakata, Kida, Hoshiyama, et al., 2007; Akatsuka, Wasaka, Nakata, Kida, & Kakigi, 2007; Andersen & Lundqvist, 2019; Butler et al., 2012; Gijsen et al., 2021; Grundei et al., 2023; Naeije et al., 2016, 2018; Ostwald et al., 2012; Spackman et al., 2010) and only very few studies investigated somatosensory MMRs with fMRI (Allen et al., 2016; Chen et al., 2008; Downar et al., 2000). Although the findings of Downar et al. (2000) in their tri-modal fMRI study overlap largely with our mismatch network, their somatosensory specific activation was restricted to SII. Our results, on the other hand, showed additional activation in IC which was distinct from its multi-modal activation in anterior portions (Id7) close to the IFJ cluster. Similarly, Chen et al. (2008) identified a comparable network of SII, IC and fronto-parietal activations for somatosensory MMRs in fMRI to attended and unattended uni-modal deviants during electrical median nerve stimulation (as used here), showing that SII activation was unmodulated by the attentional focus, in contrast to the higher-level processing stages, indicating SII as a primary driver for the first stage of mismatch processing reflected in the MMN response. Moreover, the (right) IC has been suggested to be involved in the integration of ascending sensory information with descending signals from higher level regions in prefrontal cortex during somatosensory processing (Cerliani et al., 2012; Lovero et al., 2009; Seth et al., 2011), in line with our finding. Correspondingly, in a DCM study using somatosensory roving stimulus sequences, Allen et al. (2016) identified modulations in a network comprised of S1, IC and MFG during somatosensory mismatch processing, providing evidence for a role of IC in the coordination of hierarchical predictive interactions. In correspondence with the results of our PPI-connectivity analysis, the authors showed increasing feedforward connectivity from somatosensory cortex to insular and prefrontal cortices for unexpected stimuli, while feedback projections were found between IC and somatosensory cortex, suggesting that the IC mediates the reciprocal exchange with hierarchically higher areas (Allen et al., 2016). Similarly, in another DCM study, Fardo et al. (2017) demonstrated that expectation violation reflected in somatosensory MMRs was

accompanied by intrinsic modulations within the somatosensory system in SI and SII as well as extrinsic recurrent connectivity modulations between somatosensory, pre-frontal and parietal regions. Taken together, our results are in support of interactions between SII and IC to underlie the sensory specific aspects of somatosensory mismatch processing.

Mismatch effects specific to the visual sequence were found in LOC and IT in hierarchically higher brain areas of visual processing such as V4, V5 and the fusiform gyrus. In parallel to the somatosensory modality, most previous research on visual MMRs was done using M/EEG recordings, generally indicating activation of the visual cortices (Czigler, 2007; Kimura et al., 2011; Pazo-Alvarez et al., 2003). Nevertheless, some more direct evidence has been provided for hierarchically higher visual regions in occipito-temporal cortex (Egner et al., 2010; Kimura et al., 2010; Stefanics et al., 2019; Urakawa et al., 2010; Yucel et al., 2007). A recent fMRI study identified MMRs to roving face stimuli (with emotion and color changes) in lateral occipital and posterior parietal cortex (Stefanics et al., 2019). Interestingly, although we used vastly different stimulus types (flashes as opposed to faces), we characterized highly similar higher order visual regions coding for the stimulus transitions. The authors identified a perceptual model reflecting precision weighted prediction errors as best explaining their results (Stefanics et al., 2018; Stefanics et al., 2019). Top-down projections from higher cortical processing stages are thought to modulate responses in non-primary visual areas in particular (Buffalo et al., 2010; Johnson et al., 2007; Kastner et al., 1998; Mohr et al., 2009) and visual research highlights the modulatory effects of sensory expectations in hierarchical visual processing (de Lange et al., 2018; Ferrari et al., 2022; Summerfield & de Lange, 2014; Summerfield & Egner, 2009). Given that activation in monkey IT cortex has shown to be reflective of expectation violation based on probabilistic information (Bell et al., 2016) with respect to the learning of transition probabilities (Meyer et al., 2014; Meyer & Olson, 2011), our results are well in line with current visual research on probabilistic sequence processing and in support of the notion that sensory specific visual MMRs reflect prediction error in higher visual areas violating top-down predictions (Stefanics et al., 2014; Stefanics et al., 2018).

4.2 | A modality general fronto-parietal mismatch network

In addition to the sensory specific clusters, we identified a shared network of mismatch processing across modalities with activations around the IFJ, TPJ and around the SMA/ACC. Most pronounced activation was found in the right hemisphere, especially around the IFJ. The involvement of frontal cortex in deviance detection integrates well with evidence from MMN research where source modeling suggests a combined activation of sensory and (right) inferior frontal regions. Primarily researched in the auditory modality, this combination of neuronal generators has been repeatedly demonstrated using a large variety of electrophysiological methods such as EEG (Garrido et al., 2008; Garrido, Kilner, Stephan, & Friston, 2009; Giard et al., 1990; Giard et al., 1995; Rinne et al., 2000; Shalgi & Deouell, 2007), MEG (Rinne et al., 2000), intracranial EEG (Dürschmid et al., 2016; Phillips et al., 2016) as well as optical imaging (Tse et al., 2006; Tse et al., 2013). Comparable evidence is increasingly provided across modalities (Grundei et al., 2023), for example, in somatosensation (Allen et al., 2016; Fardo et al., 2017; Huang et al., 2005) and vision (Hedge et al., 2015; Tse et al., 2021).

Using fMRI, to our knowledge, only Downar et al. (2000) have investigated MMRs to similar multi-modal sequences used here. The authors applied naturalistic stimuli, such as sounds of frogs and running water, visual shapes and tactile unilaterally applied brush strokes. Therefore, it is noteworthy that we replicated their neuroimaging results, showing the same uni- and multi-modal activation patterns using a larger sample ($N = 29$ compared to $N = 10$) in response to stimulus sequences more commonly used in current MMR research. With a heavy focus on the auditory modality, a considerable number of studies have identified a similar fronto-parietal mismatch network in uni-modal fMRI experiments (Diekhof et al., 2009; Doeller et al., 2003; Molholm et al., 2005; Opitz et al., 2002; Rinne et al., 2005; Shalgi & Deouell, 2007). Particularly the IFJ and, as opposed to most implications from the EEG literature, additionally the TPJ have appeared as consistent findings in fMRI oddball-studies across different modalities (Doricchi et al., 2022; Downar et al., 2000, 2001, 2002; Huang et al., 2005; Kim, 2014).

Given the temporal resolution of fMRI, it is impossible to isolate the MMN response from the effects of the later P3 MMR, although it is likely that fronto-sensory activation primarily reflects the first stage of mismatch processing underlying the MMN, while parietal generators are involved at later stages (Bekinschtein et al., 2009; Phillips et al., 2016; Uhrig et al., 2014). Both prominent EEG mismatch signatures have been described as involuntary attention orienting responses to unexpected sensory input (Escera et al., 2000; Näätänen, 1992; Schröger et al., 2015; Wetzel & Schröger, 2014). The identified fronto-parietal activations might therefore be reflective of the attention network identified by Corbetta and Shulman (2002) which has been suggested to orient attention to salient exogenous events and might represent a modality independent system for novelty alerting (Corbetta et al., 2008; Kim, 2014; Macaluso, 2010). The ventral part of the network includes IFJ, TPJ, IC and SMA/ACC (Corbetta & Shulman, 2002; Eckert et al., 2009; Yeo et al., 2011) and has been specifically associated with expectation violation (Vossel et al., 2014). As such, prior research highlights a central role for stimulus expectancy for an involvement of the identified network and suggests an interplay of expectation violation and attentional recruitment at later stages of mismatch processing.

Our PPI results indicate that increased projections from sensory regions to the modality general mismatch network contribute to the processing of regularity violation. Strikingly, we found that the left and the right seed regions similarly converged to clusters with right hemispheric dominance, in line with prior research on the auditory mismatch network (Dietz et al., 2014, 2021). As highlighted by a recent review, the larger TPJ area, including its dorsal extension

around SMG into AG and IPS (Igelstrom & Graziano, 2017), is thought to code matches (left lateralized) and mismatches (right lateralized) between expected and actual events across sensory, motor and cognitive operations and keeps track of their statistical contingencies (Doricchi et al., 2022; Parr et al., 2023). With the current study, we provide further evidence for such wider TPJ activation as a common signature of mismatch effects across the senses in addition to the classically demonstrated involvement of the (right) pre-frontal cortex in MMN generation. Previous MMR studies have indicated connectivity modulations within and between the identified regions in accordance with hierarchical predictive processing (Allen et al., 2016; Auksztulewicz & Friston, 2015; Chennu et al., 2016; El Karoui et al., 2015; Fardo et al., 2017; Garrido, Kilner, Stephan, & Friston, 2009; Phillips et al., 2015; Phillips et al., 2016; Uhrig et al., 2014). Our results complement these findings by showing the domain generality of connectivity modulations in the fronto-parietal mismatch network underlying expectation violation. While PPI-connectivity provides the advantage of an assumption-free exploration of functional correlation within the brain with the seed regions, future studies would benefit from DCM analyses to provide further insights into the directed modulations between the nodes identified here across modalities.

4.3 | Modulation of deviant responses by stimulus expectancy

Our parametric contrasts revealed that deviant responses within the identified mismatch network were increasing as a function of the length of the preceding standard stimulus train. The underlying sources of such deviance modulation during auditory, somatosensory and visual sequence processing have not been rendered by fMRI up to today, although indicated by similar EEG responses in different modalities (Baldeweg, 2006; Cowan et al., 1993; Gijsen et al., 2021; Grundei et al., 2023; Haenschel et al., 2005; Imada et al., 1993; Sams et al., 1983). While previous fMRI studies have shown that deviant responses increase with the relative mismatch magnitude in terms of deviant properties (Doeller et al., 2003; Opitz et al., 2002; Rinne et al., 2005), we provide additional evidence for a stimulus-independent mismatch increase with prior stimulus repetitions which is comparable across the senses. Moreover, our results show that higher order sensory regions were most reflective of the parametric modulation, largely overlapping with the sensory clusters of the identified mismatch network. Although less pronounced, additional train length effects were found in the identified modality general regions such as IFJ, TPJ and SMA/ACC, which suggests that the fronto-parietal recruitment during mismatch processing is not a binary process (e.g., attention switch or not), but rather related to the degree of expectancy induced by the prior stimulus train. Under predictive processing accounts of MMN generation (Friston, 2005; Garrido, Kilner, Stephan, & Friston, 2009; Stefanics et al., 2014), this modulation of the deviant response reflects an increasing prediction error to an established sensory regularity and increasing precision of the top-down prediction to repeat the current stimulus train (Auksztulewicz et al., 2017; Auksztulewicz & Friston, 2016). Our finding of expectancy modulated mismatch responses in higher order sensory cortices aligns well with research in rats showing that prediction error responses to regularity violations increase along the auditory processing pathway (Parras et al., 2017) with error responses predominantly found in hierarchically higher auditory regions (Luo et al., 2023; Parras et al., 2021) and adaptation dominating in primary auditory cortex (Parras et al., 2021). The additional (reduced) modulation of frontal and parietal mismatch clusters might indicate projections of remaining prediction errors to the modality independent hubs of the mismatch network, as suggested by the results of our PPI analysis. Overall, our results provide evidence for highly comparable dynamics of deviant responses based on stimulus expectancy across modalities most pronounced in sensory specific regions of the mismatch network.

4.4 | The parietal hub of the mismatch network reflects cross-modal expectation violation

We used tri-modal probabilistic sequences to create cross-modal regularities by defining the transitions in one modality conditional on the configuration of the other two modalities. As such, stimulus transitions were more or less likely based on the multi-modal context and thus rather predicted or mispredicted. With the cross-modal predictability implicit in the sequences we go beyond any previous attempts to locate MMRs to multi-modal sequences. Across sensory modalities, we identified a dorsal part of the larger TPJ area around IPS (Doricchi et al., 2022; Igelstrom & Graziano, 2017) to be particularly sensitive to cross-modal expectation violation. This is striking given that the area is well known as a major connection hub for different senses (Damasio, 1989; Hagmann et al., 2008; Tomasi & Volkow, 2011), mapping multi-modal inputs in both human and non-human primates (Avillac et al., 2007; Sereno & Huang, 2014). This *convergence zone* for multi-modal information integration (Damasio, 1989) has been proposed to provide a critical gateway to transform sensory information into cognitively relevant functions (Mesulam, 1998). Specifically, the extended TPJ area, including the IPS, forms the major parietal network hub for multi-modal integration and higher-order cognition (Igelstrom & Graziano, 2017) and is particularly known for coding unexpected events across a variety of sensory and cognitive processes (Doricchi et al., 2022). In line with these indications, our seed-based PPI-connectivity results show connectivity modulations from the sensory region of each modality to TPJ and IPS during mismatch processing. Therefore, our results suggest that the extended TPJ area forms an integrative processing hub in the mismatch network with IPS specifically signaling expectation violation based on cross-modal contingencies.

In the hierarchical structure of the cortex, the parietal convergence zone is proposed to provide an amodal interface between bottom-up sensory inputs and hierarchically higher levels (Seghier, 2013), such as the frontal cortex which is considered to transform accumulated context dependent sensory evidence from

parietal cortex into choice (Erlich et al., 2015; Hanks et al., 2015). The IPS activation found here might indicate accumulation and integration of multi-modal sensory evidence which is further projected to frontal cortex forming predictions about multi-modal regularities and ultimately informing decision making. Support for such a role of the IPS in a hierarchy of multi-modal perceptual inference comes from recent advances in research on Bayesian causal inference showing that reliability weighted sensory estimates are integrated in IPS and used in interaction with frontal areas to infer the hidden causes of sensory inputs (Cao et al., 2019; Noppeney, 2021; Rohe & Noppeney, 2015). Moreover, it has been suggested that the IPS is involved in a modality-general representation of sequences (Planton & Dehaene, 2021) as it is found during regularity violations across different modalities and presentation formats (Planton & Dehaene, 2021; Wang et al., 2015; Wang et al., 2019) which is in line with our results.

In a previous study, using the same paradigm in EEG (Grundei et al., 2023), we reported an increase of the P3 to mispredicted trials, indicating sensitivity of the response to the predictive multi-modal stimulus configuration. We showed that early and late MMRs in our EEG data were best explained by a Bayesian observer tracking stimulus transition probabilities and that more central and later responses around the P3 appeared to specifically track stimulus transitions across multiple modalities. In recent MMR research, the local–global paradigm has revealed that the early MMN and the later P3 reflect two hierarchical stages of mismatch processing signaling regularity violation on increasing levels of sequence complexity (Bekinschtein et al., 2009; Chennu et al., 2013; Chennu et al., 2016; Dehaene et al., 2015; Dürschmid et al., 2016; El Karoui et al., 2015; King et al., 2014; Niedernhuber et al., 2022; Shirazibeheshti et al., 2018; Wacongne et al., 2011): While the MMN is primarily sensitive to local regularities, such as basic stimulus repetition, the P3 is additionally sensitive to global regularities such as a repeating pattern over an extended period of time. In addition to global sequence monitoring with respect to temporal regularities, the P3 appears to additionally be sensitive to global sequence regularities in terms of the multi-modal stimulus configuration (Grundei et al., 2023). Therefore, we suggest that the late P3 MMR might signal violation of global sequence contingencies on multiple spatio-temporal scales. Given that the P3 has fronto-parietal generators, including the TPJ around SMG and IPL (Linden, 2005; Polich, 2007), the fMRI results of the current study indicate a correspondence of the intraparietal cluster sensitive to cross-modal expectation violation with our previous EEG results. Such an interpretation is supported by fMRI studies using the local–global paradigm which show parietal activations during the global MMR in humans (Bekinschtein et al., 2009) and in macaque monkey IPS (Uhrig et al., 2014). Taken together, the IPS appears to keep track of global cross-modal sequence regularities, potentially by estimation of transition probabilities across modalities.

The current work provides evidence that the activation of fronto-parietal network nodes to stimulus changes is rather independent of the input modality. In particular activation of the IPS shows sensitivity to multi-modal probabilistic stimulus combinations, as we found it modulated by cross-modal regularities. However, an alternative interpretation to cross-modal learning cannot be ruled out, namely that the three synchronously presented stimuli are bound together to a tri-modal object or triplet. If this was the case, the brain might extract the transition probabilities of tri-modal states specified in the transition matrix (see Figure 1) rather than tracking each uni-modal input stream and learning the cross-modal regularities on top. While most previous MMN literature is focused on uni-modal inputs, such different approaches of multi-modal mismatch processing have by now not been broadly explored and our work motivates further research in this direction. Therefore, future modeling studies should directly compare observer models tracking different sequence properties to evaluate if and at which stage probabilistic uni-modal inputs are integrated to, and represented as, combinations of multi-modal objects.

The IPS is also integral part of the (dorsal) network of attentional control where it is found to be a marker of memory related expectation violation (O'Connor et al., 2010) provoking attention allocation related to uncertainty in information retrieval (Hutchinson & Turk-Browne, 2012). In addition to the general activation of the attention network by unexpected events (Vossel et al., 2014), the IPS in particular might reflect expectations related to stored memory traces which further supports the idea that the area might operate on a global scale of sequence processing. Moreover, in the fronto-parietal attention network the IPS is involved in the selection of sensory expectations in a multi-dimensional environment with co-existing contingencies (Ferrari et al., 2022; Leong et al., 2017; Niv et al., 2015), directly in line with our results.

Overall, our finding of an involvement of the dorsal TPJ around the IPS in cross-modal expectation violation integrates well with current research converging to a modality-general role for the extended TPJ region in signaling a divergence between expected and actual events at various scales (Doricchi et al., 2022). Moreover, we provide evidence for a novel parietal contribution to the multi-modal mismatch network and suggest that the IPS tracks cross-modal probabilistic associations during global sequence monitoring.

4.5 | Conclusion

With the current study we substantiate previous evidence for a shared mismatch network across modalities, involving modality specific sensory cortices as well as modality independent inferior frontal and temporo-parietal areas. Additionally, we demonstrated PPI-connectivity modulations from sensory regions to common multi-modal network hubs during mismatch processing and show that deviant responses within the network were modulated by local stimulus repetition, suggesting highly comparable organization of the computation of expectation violation across the senses. Moreover, hierarchically higher regions of the mismatch network in the extended TPJ area around IPS were identified to signal cross-modal expectation violation and might keep track of global multi-modal sequence regularities. Overall, these findings shed light on mismatch responses to

multi-modal probabilistic inputs in a shared cortical network of expectation violation.

ACKNOWLEDGMENT

Open Access funding enabled and organized by Projekt DEAL.

FUNDING INFORMATION

This work was supported by Berlin School of Mind and Brain, Humboldt Universität zu Berlin (Miro Grundei; http://www.mind-and-brain.de/home/). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Miro Grundei https://orcid.org/0000-0002-6396-9674

REFERENCES

Akatsuka, K., Wasaka, T., Nakata, H., Kida, T., Hoshiyama, M., Tamura, Y., & Kakigi, R. (2007). Objective examination for two-point stimulation using a somatosensory oddball paradigm: An MEG study. *Clinical Neurophysiology*, *118*(2), 403–411. https://doi.org/10.1016/j.clinph.2006.09.030

Akatsuka, K., Wasaka, T., Nakata, H., Kida, T., & Kakigi, R. (2007). The effect of stimulus probability on the somatosensory mismatch field. *Experimental Brain Research*, *181*(4), 607–614. https://doi.org/10.1007/s00221-007-0958-4

Alain, C., Achim, A., & Woods, D. L. (1999). Separate memory-related processing for auditory frequency and patterns. *Psychophysiology*, *36*(6), 737–744.

Alain, C., & Woods, D. L. (1997). Attention modulates auditory pattern memory as indexed by event-related brain potentials. *Psychophysiology*, *34*(5), 534–546. https://doi.org/10.1111/j.1469-8986.1997.tb01740.x

Alho, K. (1995). Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. *Ear and Hearing*, *16*(1), 38–51. https://journals.lww.com/ear-hearing/Fulltext/1995/02000/Cerebral_Generators_of_Mismatch_Negativity__MMN_.4.aspx

Allen, M., Fardo, F., Dietz, M. J., Hillebrandt, H., Friston, K. J., Rees, G., & Roepstorff, A. (2016). Anterior insula coordinates hierarchical processing of tactile mismatch responses. *NeuroImage*, *127*, 34–43. https://doi.org/10.1016/j.neuroimage.2015.11.030

Altmann, C. F., Bledowski, C., Wibral, M., & Kaiser, J. (2007). Processing of location and pattern changes of natural sounds in the human auditory cortex. *NeuroImage*, *35*(3), 1192–1200. https://doi.org/10.1016/j.neuroimage.2007.01.007

Andersen, L. M., & Lundqvist, D. (2019). Somatosensory responses to nothing: An MEG study of expectations during omission of tactile stimulations. *NeuroImage*, *184*, 78–89. https://doi.org/10.1016/j.neuroimage.2018.09.014

Arnal, L. H., & Giraud, A. L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences*, *16*(7), 390–398. https://doi.org/10.1016/j.tics.2012.05.003

Auksztulewicz, R., Barascud, N., Cooray, G., Nobre, A. C., Chait, M., & Friston, K. (2017). The cumulative effects of predictability on synaptic gain in the auditory processing stream. *The Journal of Neuroscience*, *37*(28), 6751–6760. https://doi.org/10.1523/JNEUROSCI.0291-17.2017

Auksztulewicz, R., & Friston, K. (2015). Attentional enhancement of auditory mismatch responses: A DCM/MEG study. *Cerebral Cortex*, *25*(11), 4273–4283. https://doi.org/10.1093/cercor/bhu323

Auksztulewicz, R., & Friston, K. (2016). Repetition suppression and its contextual determinants in predictive coding. *Cortex*, *80*, 125–140. https://doi.org/10.1016/j.cortex.2015.11.024

Avillac, M., Ben Hamed, S., & Duhamel, J. R. (2007). Multisensory integration in the ventral intraparietal area of the macaque monkey. *The Journal of Neuroscience*, *27*(8), 1922–1932. https://doi.org/10.1523/JNEUROSCI.2646-06.2007

Baldeweg, T. (2006). Repetition effects to sounds: Evidence for predictive coding in the auditory system. *Trends in Cognitive Sciences*, *10*(3), 93–94. https://doi.org/10.1016/j.tics.2006.01.010

Baldeweg, T., Klugman, A., Gruzelier, J., & Hirsch, S. R. (2004). Mismatch negativity potentials and cognitive impairment in schizophrenia. *Schizophrenia Research*, *69*(2–3), 203–217. https://doi.org/10.1016/j.schres.2003.09.009

Barascud, N., Pearce, M. T., Griffiths, T. D., Friston, K. J., & Chait, M. (2016). Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(5), E616–E625. https://doi.org/10.1073/pnas.1508523113

Bekinschtein, T. A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., & Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(5), 1672–1677. https://doi.org/10.1073/pnas.0809667106

Bell, A. H., Summerfield, C., Morin, E. L., Malecek, N. J., & Ungerleider, L. G. (2016). Encoding of stimulus probability in macaque inferior temporal cortex. *Current Biology*, *26*(17), 2280–2290. https://doi.org/10.1016/j.cub.2016.07.007

Bendixen, A., SanMiguel, I., & Schröger, E. (2012). Early electrophysiological indicators for predictive processing in audition: A review. *International Journal of Psychophysiology*, *83*(2), 120–131. https://doi.org/10.1016/j.ijpsycho.2011.08.003

Bendixen, A., Schröger, E., & Winkler, I. (2009). I heard that coming: Event-related potential evidence for stimulus-driven prediction in the auditory system. *The Journal of Neuroscience*, *29*(26), 8447–8451. https://doi.org/10.1523/JNEUROSCI.1493-09.2009

Besle, J., Fort, A., & Giard, M. H. (2005). Is the auditory sensory memory sensitive to visual information? *Experimental Brain Research*, *166*(3–4), 337–344. https://doi.org/10.1007/s00221-005-2375-x

Bresciani, J. P., Dammeier, F., & Ernst, M. O. (2006). Vision and touch are automatically integrated for the perception of sequences of events. *Journal of Vision*, *6*(5), 554–564. https://doi.org/10.1167/6.5.2

Bresciani, J. P., Dammeier, F., & Ernst, M. O. (2008). Tri-modal integration of visual, tactile and auditory signals for the perception of sequences of events. *Brain Research Bulletin*, *75*(6), 753–760. https://doi.org/10.1016/j.brainresbull.2008.01.009

Buffalo, E. A., Fries, P., Landman, R., Liang, H., & Desimone, R. (2010). A backward progression of attentional effects in the ventral stream. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(1), 361–365. https://doi.org/10.1073/pnas.0907658106

Butler, J. S., Foxe, J. J., Fiebelkorn, I. C., Mercier, M. R., & Molholm, S. (2012). Multisensory representation of frequency across audition and touch: High density electrical mapping reveals early sensory-perceptual coupling. *The Journal of Neuroscience*, *32*(44), 15338–15344. https://doi.org/10.1523/JNEUROSCI.1796-12.2012

Cao, Y., Summerfield, C., Park, H., Giordano, B. L., & Kayser, C. (2019). Causal inference in the multisensory brain. *Neuron*, *102*(5), 1076–1087. e1078. https://doi.org/10.1016/j.neuron.2019.03.043

Cerliani, L., Thomas, R. M., Jbabdi, S., Siero, J. C., Nanetti, L., Crippa, A., Gazzola, V., D'Arceuil, H., & Keysers, C. (2012). Probabilistic tractography recovers a rostrocaudal trajectory of connectivity variability in the human insular cortex. *Human Brain Mapping*, *33*(9), 2005–2034. https://doi.org/10.1002/hbm.21338

Chao, Z. C., Takaura, K., Wang, L., Fujii, N., & Dehaene, S. (2018). Large-scale cortical networks for hierarchical prediction and prediction error in the primate brain. *Neuron*, *100*(5), 1252–1266. e1253. https://doi.org/10.1016/j.neuron.2018.10.004

Chen, T. L., Babiloni, C., Ferretti, A., Perrucci, M. G., Romani, G. L., Rossini, P. M., Tartaro, A., & Del Gratta, C. (2008). Human secondary somatosensory cortex is involved in the processing of somatosensory rare stimuli: An fMRI study. *NeuroImage*, *40*(4), 1765–1771. https://doi.org/10.1016/j.neuroimage.2008.01.020

Chennu, S., Noreika, V., Gueorguiev, D., Blenkmann, A., Kochen, S., Ibanez, A., Owen, A. M., & Bekinschtein, T. A. (2013). Expectation and attention in hierarchical auditory prediction. *The Journal of Neuroscience*, *33*(27), 11194–11205. https://doi.org/10.1523/JNEUROSCI.0114-13.2013

Chennu, S., Noreika, V., Gueorguiev, D., Shtyrov, Y., Bekinschtein, T. A., & Henson, R. (2016). Silent expectations: Dynamic causal modeling of cortical prediction and attention to sounds that weren't. *The Journal of Neuroscience*, *36*(32), 8305–8316. https://doi.org/10.1523/JNEUROSCI.1125-16.2016

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *The Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Conill, J. (1998). P300 potentials evoked by visual stimulation. *Revista de Neurologia*, *26*(151), 448–451. (Potencial P300 provocado por estimulos visuales.).

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: From environment to theory of mind. *Neuron*, *58*(3), 306–324. https://doi.org/10.1016/j.neuron.2008.04.017

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews. Neuroscience*, *3*(3), 201–215. https://doi.org/10.1038/nrn755

Cowan, N., Winkler, I., Teder, W., & Naatanen, R. (1993). Memory prerequisites of mismatch negativity in the auditory event-related potential (ERP). *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *19*(4), 909–921. https://doi.org/10.1037//0278-7393.19.4.909

Czigler, I. (2007). Visual mismatch negativity. *Journal of Psychophysiology*, *21*(3–4), 224–230. https://doi.org/10.1027/0269-8803.21.34.224

Czigler, I., Winkler, I., Pato, L., Varnagy, A., Weisz, J., & Balazs, L. (2006). Visual temporal window of integration as revealed by the visual mismatch negativity event-related potential to stimulus omissions. *Brain Research*, *1104*(1), 129–140. https://doi.org/10.1016/j.brainres.2006.05.034

Damasio, A. R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition*, *33*(1–2), 25–62. https://doi.org/10.1016/0010-0277(89)90005-x

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception? *Trends in Cognitive Sciences*, *22*(9), 764–779. https://doi.org/10.1016/j.tics.2018.06.002

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., & Pallier, C. (2015). The neural representation of sequences: From transition probabilities to algebraic patterns and linguistic trees. *Neuron*, *88*(1), 2–19. https://doi.org/10.1016/j.neuron.2015.09.019

Deouell, L. Y. (2007). The frontal generator of the mismatch negativity revisited. *Journal of Psychophysiology*, *21*(3–4), 188–203.

Diekhof, E. K., Biedermann, F., Ruebsamen, R., & Gruber, O. (2009). Top-down and bottom-up modulation of brain structures involved in auditory discrimination. *Brain Research*, *1297*, 118–123. https://doi.org/10.1016/j.brainres.2009.08.040

Dietz, M. J., Friston, K. J., Mattingley, J. B., Roepstorff, A., & Garrido, M. I. (2014). Effective connectivity reveals right-hemisphere dominance in audiospatial perception: implications for models of spatial neglect. *Journal of Neuroscience*, *34*(14), 5003–5011.

Dietz, M. J., Nielsen, J. F., Roepstorff, A., & Garrido, M. I. (2021). Reduced effective connectivity between right parietal and inferior frontal cortex during audiospatial perception in neglect patients with a right-hemisphere lesion. *Hearing Research*, *399*, 108052.

Doeller, C. F., Opitz, B., Mecklinger, A., Krick, C., Reith, W., & Schröger, E. (2003). Prefrontal cortex involvement in preattentive auditory deviance detection: Neuroimaging and electrophysiological evidence. *NeuroImage*, *20*(2), 1270–1282. https://doi.org/10.1016/S1053-8119(03)00389-6

Doricchi, F., Lasaponara, S., Pazzaglia, M., & Silvetti, M. (2022). Left and right temporal-parietal junctions (TPJs) as "match/mismatch" hedonic machines: A unifying account of TPJ function. *Physics of Life Reviews*, *42*, 56–92. https://doi.org/10.1016/j.plrev.2022.07.001

Downar, J., Crawley, A. P., Mikulis, D. J., & Davis, K. D. (2000). A multimodal cortical network for the detection of changes in the sensory environment. *Nature Neuroscience*, *3*(3), 277–283. https://doi.org/10.1038/72991

Downar, J., Crawley, A. P., Mikulis, D. J., & Davis, K. D. (2001). The effect of task relevance on the cortical response to changes in visual and auditory stimuli: An event-related fMRI study. *NeuroImage*, *14*(6), 1256–1267. https://doi.org/10.1006/nimg.2001.0946

Downar, J., Crawley, A. P., Mikulis, D. J., & Davis, K. D. (2002). A cortical network sensitive to stimulus salience in a neutral behavioral context across multiple sensory modalities. *Journal of Neurophysiology*, *87*(1), 615–620. https://doi.org/10.1152/jn.00636.2001

Duncan, C. C., Barry, R. J., Connolly, J. F., Fischer, C., Michie, P. T., Naatanen, R., Polich, J., Reinvang, I., & Van Petten, C. (2009). Event-related potentials in clinical research: Guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clinical Neurophysiology*, *120*(11), 1883–1908. https://doi.org/10.1016/j.clinph.2009.07.045

Duncan-Johnson, C. C., & Donchin, E. (1982). The P300 component of the event-related brain potential as an index of information processing. *Biological Psychology*, *14*(1–2), 1–52. https://doi.org/10.1016/0301-0511(82)90016-3

Dürschmid, S., Edwards, E., Reichert, C., Dewar, C., Hinrichs, H., Heinze, H. J., Kirsch, H. E., Dalal, S. S., Deouell, L. Y., & Knight, R. T. (2016). Hierarchy of prediction errors for auditory events in human temporal and frontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(24), 6755–6760. https://doi.org/10.1073/pnas.1525030113

Eckert, M. A., Menon, V., Walczak, A., Ahlstrom, J., Denslow, S., Horwitz, A., & Dubno, J. R. (2009). At the heart of the ventral attention system: The right anterior insula. *Human Brain Mapping*, *30*(8), 2530–2541. https://doi.org/10.1002/hbm.20688

Egner, T., Monti, J. M., & Summerfield, C. (2010). Expectation and surprise determine neural population responses in the ventral visual stream. *The Journal of Neuroscience*, *30*(49), 16601–16608. https://doi.org/10.1523/JNEUROSCI.2770-10.2010

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*, *25*(4), 1325–1335. https://doi.org/10.1016/j.neuroimage.2004.12.034

El Karoui, I., King, J. R., Sitt, J., Meyniel, F., Van Gaal, S., Hasboun, D., Adam, C., Navarro, V., Baulac, M., Dehaene, S., Cohen, L., & Naccache, L. (2015). Event-related potential, time-frequency, and functional connectivity facets of local and global auditory novelty processing: An intracranial study in humans. *Cerebral Cortex*, *25*(11), 4203–4212. https://doi.org/10.1093/cercor/bhu143

Erlich, J. C., Brunton, B. W., Duan, C. A., Hanks, T. D., & Brody, C. D. (2015). Distinct effects of prefrontal and parietal cortex inactivations on an accumulation of evidence task in the rat. *eLife*, *4*, e05457. https://doi.org/10.7554/eLife.05457

Escera, C., Alho, K., Schröger, E., & Winkler, I. (2000). Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiology & Neuro-Otology*, *5*(3–4), 151–166. https://doi.org/10.1159/000013877

Ewbank, M. P., Lawson, R. P., Henson, R. N., Rowe, J. B., Passamonti, L., & Calder, A. J. (2011). Changes in "top-down" connectivity underlie repetition suppression in the ventral visual pathway. *The Journal of Neuroscience*, *31*(15), 5635–5642. https://doi.org/10.1523/JNEUROSCI.5013-10.2011

Fardo, F., Auksztulewicz, R., Allen, M., Dietz, M. J., Roepstorff, A., & Friston, K. J. (2017). Expectation violation and attention to pain jointly modulate neural gain in somatosensory cortex. *NeuroImage*, *153*, 109–121. https://doi.org/10.1016/j.neuroimage.2017.03.041

Ferrari, A., Richter, D., & de Lange, F. P. (2022). Updating contextual sensory expectations for adaptive behavior. *The Journal of Neuroscience*, *42*(47), 8855–8869. https://doi.org/10.1523/JNEUROSCI.1107-22.2022

Friedel, E. B. N., Bach, M., & Heinrich, S. P. (2020). Attentional interactions between vision and hearing in event-related responses to crossmodal and conjunct oddballs. *Multisensory Research*, *33*(3), 251–275. https://doi.org/10.1163/22134808-20191329

Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: An event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience and Biobehavioral Reviews*, *25*(4), 355–373. https://doi.org/10.1016/s0149-7634(01)00019-7

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

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews. Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage*, *6*(3), 218–229. https://doi.org/10.1006/nimg.1997.0291

Friston, K. J., Holmes, A. P., Price, C. J., Buchel, C., & Worsley, K. J. (1999). Multisubject fMRI studies and conjunction analyses. *NeuroImage*, *10*(4), 385–396. https://doi.org/10.1006/nimg.1999.0484

Friston, K. J., Penny, W. D., & Glaser, D. E. (2005). Conjunction revisited. *NeuroImage*, *25*(3), 661–667. https://doi.org/10.1016/j.neuroimage.2005.01.013

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences*, *19*(3), 117–125. https://doi.org/10.1016/j.tics.2014.12.010

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. (2007). Evoked brain responses are generated by feedback loops. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(52), 20961–20966. https://doi.org/10.1073/pnas.0706274105

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

Geisler, W. S. (2008). Visual perception and the statistical properties of natural scenes. *Annual Review of Psychology*, *59*, 167–192. https://doi.org/10.1146/annurev.psych.58.110405.085632

Giard, M. H., Lavikahen, J., Reinikainen, K., Perrin, F., Bertrand, O., Pernier, J., & Naatanen, R. (1995). Separate representation of stimulus frequency, intensity, and duration in auditory sensory memory: An event-related potential and dipole-model analysis. *Journal of Cognitive Neuroscience*, *7*(2), 133–143. https://doi.org/10.1162/jocn.1995.7.2.133

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

Gijsen, S., Grundei, M., Lange, R. T., Ostwald, D., & Blankenburg, F. (2021). Neural surprise in somatosensory Bayesian learning. *PLoS Computational Biology*, *17*(2), e1008068. https://doi.org/10.1371/journal.pcbi.1008068

Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *290*(1038), 181–197. https://doi.org/10.1098/rstb.1980.0090

Grundei, M., Schroder, P., Gijsen, S., & Blankenburg, F. (2023). EEG mismatch responses in a multimodal roving stimulus paradigm provide evidence for probabilistic inference across audition, somatosensation, and vision. *Human Brain Mapping*, *44*(9), 3644–3668. https://doi.org/10.1002/hbm.26303

Haenschel, C., Vernon, D. J., Dwivedi, P., Gruzelier, J. H., & Baldeweg, T. (2005). Event-related brain potential correlates of human auditory sensory memory-trace formation. *The Journal of Neuroscience*, *25*(45), 10494–10501. https://doi.org/10.1523/JNEUROSCI.1227-05.2005

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., & Sporns, O. (2008). Mapping the structural core of human cerebral cortex. *PLoS Biology*, *6*(7), e159. https://doi.org/10.1371/journal.pbio.0060159

Hanks, T. D., Kopec, C. D., Brunton, B. W., Duan, C. A., Erlich, J. C., & Brody, C. D. (2015). Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature*, *520*(7546), 220–223. https://doi.org/10.1038/nature14066

Hedge, C., Stothart, G., Todd Jones, J., Rojas Frias, P., Magee, K. L., & Brooks, J. C. (2015). A frontal attention mechanism in the visual mismatch negativity. *Behavioural Brain Research*, *293*, 173–181. https://doi.org/10.1016/j.bbr.2015.07.022

Heilbron, M., & Chait, M. (2018). Great expectations: Is there evidence for predictive coding in auditory cortex? *Neuroscience*, *389*, 54–73. https://doi.org/10.1016/j.neuroscience.2017.07.061

Horvath, J., & Winkler, I. (2004). How the human auditory system treats repetition amongst change. *Neuroscience Letters*, *368*(2), 157–161. https://doi.org/10.1016/j.neulet.2004.07.004

Hu, L., Zhao, C., Li, H., & Valentini, E. (2013). Mismatch responses evoked by nociceptive stimuli. *Psychophysiology*, *50*(2), 158–173. https://doi.org/10.1111/psyp.12000

Huang, M. X., Lee, R. R., Miller, G. A., Thoma, R. J., Hanlon, F. M., Paulson, K. M., Martin, K., Harrington, D. L., Weisend, M. P., Edgar, J. C., & Canive, J. M. (2005). A parietal-frontal network studied by somatosensory oddball MEG responses, and its cross-modal consistency. *NeuroImage*, *28*(1), 99–114. https://doi.org/10.1016/j.neuroimage.2005.05.036

Hughes, L. E., Ghosh, B. C., & Rowe, J. B. (2013). Reorganisation of brain networks in frontotemporal dementia and progressive supranuclear palsy. *Neuroimage Clinical*, *2*, 459–468. https://doi.org/10.1016/j.nicl.2013.03.009

Hutchinson, J. B., & Turk-Browne, N. B. (2012). Memory-guided attention: Control from multiple memory systems. *Trends in Cognitive Sciences*, *16*(12), 576–579. https://doi.org/10.1016/j.tics.2012.10.003

Igelstrom, K. M., & Graziano, M. S. A. (2017). The inferior parietal lobule and temporoparietal junction: A network perspective.

Neuropsychologia, *105*, 70–83. https://doi.org/10.1016/j.neuropsychologia.2017.01.001

Iglesias, S., Mathys, C., Brodersen, K. H., Kasper, L., Piccirelli, M., den Ouden, H. E., & Stephan, K. E. (2013). Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. *Neuron*, *80*(2), 519–530. https://doi.org/10.1016/j.neuron.2013.09.009

Imada, T., Hari, R., Loveless, N., McEvoy, L., & Sams, M. (1993). Determinants of the auditory mismatch response. *Electroencephalography and Clinical Neurophysiology*, *87*(3), 144–153. https://doi.org/10.1016/0013-4694(93)90120-k

Jääskelainen, I. P., Ahveninen, J., Bonmassar, G., Dale, A. M., Ilmoniemi, R. J., Levanen, S., Lin, F. H., May, P., Melcher, J., Stufflebeam, S., Tiitinen, H., & Belliveau, J. W. (2004). Human posterior auditory cortex gates novel sounds to consciousness. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(17), 6809–6814. https://doi.org/10.1073/pnas.0303760101

Johnson, M. R., Mitchell, K. J., Raye, C. L., D'Esposito, M., & Johnson, M. K. (2007). A brief thought can modulate activity in extrastriate visual areas: Top-down effects of refreshing just-seen visual stimuli. *NeuroImage*, *37*(1), 290–299. https://doi.org/10.1016/j.neuroimage.2007.05.017

Kastner, S., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1998). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science*, *282*(5386), 108–111. https://doi.org/10.1126/science.282.5386.108

Kekoni, J., Hamalainen, H., Saarinen, M., Grohn, J., Reinikainen, K., Lehtokoski, A., & Naatanen, R. (1997). Rate effect and mismatch responses in the somatosensory system: ERP-recordings in humans. *Biological Psychology*, *46*(2), 125–142. https://doi.org/10.1016/s0301-0511(97)05249-6

Kellermann, T., Scholle, R., Schneider, F., & Habel, U. (2017). Decreasing predictability of visual motion enhances feed-forward processing in visual cortex when stimuli are behaviorally relevant. *Brain Structure & Function*, *222*(2), 849–866. https://doi.org/10.1007/s00429-016-1251-8

Kiat, J. E. (2018). Assessing cross-modal target transition effects with a visual-auditory oddball. *International Journal of Psychophysiology*, *129*, 58–66. https://doi.org/10.1016/j.ijpsycho.2018.04.010

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

Kim, H. (2014). Involvement of the dorsal and ventral attention networks in oddball stimulus processing: A meta-analysis. *Human Brain Mapping*, *35*(5), 2265–2284. https://doi.org/10.1002/hbm.22326

Kimura, M., Ohira, H., & Schröger, E. (2010). Localizing sensory and cognitive systems for pre-attentive visual deviance detection: An sLORETA analysis of the data of Kimura et al. (2009). *Neuroscience Letters*, *485*(3), 198–203. https://doi.org/10.1016/j.neulet.2010.09.011

Kimura, M., Schröger, E., & Czigler, I. (2011). Visual mismatch negativity and its importance in visual cognitive sciences. *Neuroreport*, *22*(14), 669–673. https://doi.org/10.1097/WNR.0b013e32834973ba

King, J. R., Gramfort, A., Schurger, A., Naccache, L., & Dehaene, S. (2014). Two distinct dynamic modes subtend the detection of unexpected sounds. *PLoS One*, *9*(1), e85791. https://doi.org/10.1371/journal.pone.0085791

Knight, R. T., & Scabini, D. (1998). Anatomic bases of event-related potentials and their relationship to novelty detection in humans. *Journal of Clinical Neurophysiology*, *15*(1), 3–13. https://doi.org/10.1097/00004691-199801000-00003

Langner, R., Kellermann, T., Boers, F., Sturm, W., Willmes, K., & Eickhoff, S. B. (2011). Modality-specific perceptual expectations selectively modulate baseline activity in auditory, somatosensory, and visual cortices. *Cerebral Cortex*, *21*(12), 2850–2862. https://doi.org/10.1093/cercor/bhr083

Lecaignard, F., Bertrand, O., Caclin, A., & Mattout, J. (2022). Neurocomputational underpinnings of expected surprise. *The Journal of Neuroscience*, *42*(3), 474–486. https://doi.org/10.1523/JNEUROSCI.0601-21.2021

Leong, Y. C., Radulescu, A., Daniel, R., DeWoskin, V., & Niv, Y. (2017). Dynamic interaction between reinforcement learning and attention in multidimensional environments. *Neuron*, *93*(2), 451–463. https://doi.org/10.1016/j.neuron.2016.12.040

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

Linden, D. E. (2005). The p300: Where in the brain is it produced and what does it tell us? *The Neuroscientist*, *11*(6), 563–576. https://doi.org/10.1177/1073858405280524

Lovero, K. L., Simmons, A. N., Aron, J. L., & Paulus, M. P. (2009). Anterior insular cortex anticipates impending stimulus significance. *NeuroImage*, *45*(3), 976–983. https://doi.org/10.1016/j.neuroimage.2008.12.070

Luo, D., Liu, J., Auksztulewicz, R., Wing Yip, T. K., Kanold, P. O., & Schnupp, J. W. (2023). Hierarchical deviant processing in auditory cortex of awake mice. *bioRxiv*. https://doi.org/10.1101/2023.01.18.524413

Macaluso, E. (2010). Orienting of spatial attention and the interplay between the senses. *Cortex*, *46*(3), 282–297. https://doi.org/10.1016/j.cortex.2009.05.010

Macdonald, M., & Campbell, K. (2011). Effects of a violation of an expected increase or decrease in intensity on detection of change within an auditory pattern. *Brain and Cognition*, *77*(3), 438–445. https://doi.org/10.1016/j.bandc.2011.08.014

Macey, P. M., Macey, K. E., Kumar, R., & Harper, R. M. (2004). A method for removal of global effects from fMRI time series. *NeuroImage*, *22*(1), 360–366. https://doi.org/10.1016/j.neuroimage.2003.12.042

Maheu, M., Dehaene, S., & Meyniel, F. (2019). Brain signatures of a multiscale process of sequence learning in humans. *eLife*, *8*, e41541. https://doi.org/10.7554/eLife.41541

May, P. J., & Tiitinen, H. (2010). Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology*, *47*(1), 66–122. https://doi.org/10.1111/j.1469-8986.2009.00856.x

Mesulam, M. M. (1998). From sensation to cognition. *Brain*, *121*(Pt 6), 1013–1052. https://doi.org/10.1093/brain/121.6.1013

Meyer, T., & Olson, C. R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(48), 19401–19406. https://doi.org/10.1073/pnas.1112895108

Meyer, T., Ramachandran, S., & Olson, C. R. (2014). Statistical learning of serial visual transitions by neurons in monkey inferotemporal cortex. *The Journal of Neuroscience*, *34*(28), 9332–9337. https://doi.org/10.1523/JNEUROSCI.1215-14.2014

Meyniel, F., Maheu, M., & Dehaene, S. (2016). Human inferences about sequences: A minimal transition probability model. *PLoS Computational Biology*, *12*(12), e1005260. https://doi.org/10.1371/journal.pcbi.1005260

Mittag, M., Takegata, R., & Winkler, I. (2016). Transitional probabilities are prioritized over stimulus/pattern probabilities in auditory deviance detection: Memory basis for predictive sound processing. *The Journal of Neuroscience*, *36*(37), 9572–9579. https://doi.org/10.1523/JNEUROSCI.1041-16.2016

Mohr, H. M., Linder, N. S., Linden, D. E., Kaiser, J., & Sireteanu, R. (2009). Orientation-specific adaptation to mentally generated lines in human visual cortex. *NeuroImage*, *47*(1), 384–391. https://doi.org/10.1016/j.neuroimage.2009.03.045

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

Mumford, D. (1992). On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biological Cybernetics*, *66*(3), 241–251. https://doi.org/10.1007/BF00198477

Näätänen, R. (1992). *Attention and brain function*. Psychology Press.

Näätänen, R., Astikainen, P., Ruusuvirta, T., & Huotilainen, M. (2010). Automatic auditory intelligence: An expression of the sensory-cognitive core of cognitive processes. *Brain Research Reviews*, *64*(1), 123–136. https://doi.org/10.1016/j.brainresrev.2010.03.001

Näätänen, R., & Gaillard, A. (1983). 5 The orienting reflex and the N2 deflection of the event-related potential (ERP). In *Advances in psychology* (Vol. 10, pp. 119–141). Elsevier.

Näätänen, R., Gaillard, A. W., & Mantysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, *42*(4), 313–329. https://doi.org/10.1016/0001-6918(78)90006-9

Näätänen, R., Jacobsen, T., & Winkler, I. (2005). Memory-based or afferent processes in mismatch negativity (MMN): A review of the evidence. *Psychophysiology*, *42*(1), 25–32. https://doi.org/10.1111/j.1469-8986.2005.00256.x

Näätänen, R., Kujala, T., & Light, G. (2019). *Mismatch negativity: A window to the brain*. Oxford University Press.

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

Näätänen, R., Paavilainen, P., Tiitinen, H., Jiang, D., & Alho, K. (1993). Attention and mismatch negativity. *Psychophysiology*, *30*(5), 436–450. https://doi.org/10.1111/j.1469-8986.1993.tb02067.x

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001). "Primitive intelligence" in the auditory cortex. *Trends in Neurosciences*, *24*(5), 283–288. https://doi.org/10.1016/s0166-2236(00)01790-2

Naeije, G., Vaulet, T., Wens, V., Marty, B., Goldman, S., & De Tiege, X. (2016). Multilevel cortical processing of somatosensory novelty: A magnetoencephalography study. *Frontiers in Human Neuroscience*, *10*, 259. https://doi.org/10.3389/fnhum.2016.00259

Naeije, G., Vaulet, T., Wens, V., Marty, B., Goldman, S., & De Tiege, X. (2018). Neural basis of early somatosensory change detection: A magnetoencephalography study. *Brain Topography*, *31*(2), 242–256. https://doi.org/10.1007/s10548-017-0591-x

Niedernhuber, M., Raimondo, F., Sitt, J. D., & Bekinschtein, T. A. (2022). Sensory target detection at local and global timescales reveals a hierarchy of supramodal dynamics in the human cortex. *The Journal of Neuroscience*, *42*(46), 8729–8741. https://doi.org/10.1523/JNEUROSCI.0658-22.2022

Niv, Y., Daniel, R., Geana, A., Gershman, S. J., Leong, Y. C., Radulescu, A., & Wilson, R. C. (2015). Reinforcement learning in multidimensional environments relies on attention mechanisms. *The Journal of Neuroscience*, *35*(21), 8145–8157. https://doi.org/10.1523/JNEUROSCI.2978-14.2015

Noppeney, U. (2021). Perceptual inference, learning, and attention in a multisensory world. *Annual Review of Neuroscience*, *44*, 449–473. https://doi.org/10.1146/annurev-neuro-100120-085519

Nordby, H., Roth, W. T., & Pfefferbaum, A. (1988). Event-related potentials to breaks in sequences of alternating pitches or interstimulus intervals. *Psychophysiology*, *25*(3), 262–268. https://doi.org/10.1111/j.1469-8986.1988.tb01239.x

O'Connor, A. R., Han, S., & Dobbins, I. G. (2010). The inferior parietal lobule and recognition memory: Expectancy violation or successful retrieval? *The Journal of Neuroscience*, *30*(8), 2924–2934. https://doi.org/10.1523/JNEUROSCI.4225-09.2010

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

Ostwald, D., Spitzer, B., Guggenmos, M., Schmidt, T. T., Kiebel, S. J., & Blankenburg, F. (2012). Evidence for neural encoding of Bayesian surprise in human somatosensation. *NeuroImage*, *62*(1), 177–188. https://doi.org/10.1016/j.neuroimage.2012.04.050

Paavilainen, P. (2013). The mismatch-negativity (MMN) component of the auditory event-related potential to violations of abstract regularities: A review. *International Journal of Psychophysiology*, *88*(2), 109–123. https://doi.org/10.1016/j.ijpsycho.2013.03.015

Parr, T., Kilner, J., & Friston, K. (2023). Functional asymmetry and the consequences of action: Comment on: Left and right temporal-parietal junctions (TPJs) as "match/mismatch" hedonic machines: A unifying account of TPJ function by Fabrizio Doricchi et al. *Physics of Life Reviews*, *44*, 145–147. https://doi.org/10.1016/j.plrev.2023.01.004

Parras, G. G., Casado-Roman, L., Schröger, E., & Malmierca, M. S. (2021). The posterior auditory field is the chief generator of prediction error signals in the auditory cortex. *NeuroImage*, *242*, 118446. https://doi.org/10.1016/j.neuroimage.2021.118446

Parras, G. G., Nieto-Diego, J., Carbajal, G. V., Valdes-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

Pazo-Alvarez, P., Cadaveira, F., & Amenedo, E. (2003). MMN in the visual modality: A review. *Biological Psychology*, *63*(3), 199–236. https://doi.org/10.1016/s0301-0511(03)00049-8

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*(5), 233–238. https://doi.org/10.1016/j.tics.2006.03.006

Phillips, H. N., Blenkmann, A., Hughes, L. E., Bekinschtein, T. A., & Rowe, J. B. (2015). Hierarchical organization of frontotemporal networks for the prediction of stimuli across multiple dimensions. *The Journal of Neuroscience*, *35*(25), 9255–9264. https://doi.org/10.1523/JNEUROSCI.5095-14.2015

Phillips, H. N., Blenkmann, A., Hughes, L. E., Kochen, S., Bekinschtein, T. A., Cam, C., & Rowe, J. B. (2016). Convergent evidence for hierarchical prediction networks from human electrocorticography and magnetoencephalography. *Cortex*, *82*, 192–205. https://doi.org/10.1016/j.cortex.2016.05.001

Picton, T. W. (1992). The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology*, *9*(4), 456–479. https://doi.org/10.1097/00004691-199210000-00002

Planton, S., & Dehaene, S. (2021). Cerebral representation of sequence patterns across multiple presentation formats. *Cortex*, *145*, 13–36. https://doi.org/10.1016/j.cortex.2021.09.003

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Prete, D. A., Heikoop, D., McGillivray, J. E., Reilly, J. P., & Trainor, L. J. (2022). The sound of silence: Predictive error responses to unexpected sound omission in adults. *The European Journal of Neuroscience*, *55*(8), 1972–1985. https://doi.org/10.1111/ejn.15660

Purves, D. E., Augustine, G. J., Fitzpatrick, D. E., Hall, W. C., LaMantia, A.-S. E., McNamara, J. O., & White, L. E. (2008). *Neuroscience*. 4th ed. Sunderland, MA: Sinauer Associates.

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

Rinne, T., Alho, K., Ilmoniemi, R. J., Virtanen, J., & Näätänen, R. (2000). Separate time behaviors of the temporal and frontal mismatch negativity

sources. *NeuroImage*, *12*(1), 14–19. https://doi.org/10.1006/nimg.2000.0591

Rinne, T., Degerman, A., & Alho, K. (2005). Superior temporal and inferior frontal cortices are activated by infrequent sound duration decrements: An fMRI study. *NeuroImage*, *26*(1), 66–72. https://doi.org/10.1016/j.neuroimage.2005.01.017

Ritter, W., Sussman, E., Deacon, D., Cowan, N., & Vaughan, H. G., Jr. (1999). Two cognitive systems simultaneously prepared for opposite events. *Psychophysiology*, *36*(6), 835–838. https://www.ncbi.nlm.nih.gov/pubmed/10554596

Rohe, T., & Noppeney, U. (2015). Cortical hierarchies perform Bayesian causal inference in multisensory perception. *PLoS Biology*, *13*(2), e1002073. https://doi.org/10.1371/journal.pbio.1002073

Sabri, M., Kareken, D. A., Dzemidzic, M., Lowe, M. J., & Melara, R. D. (2004). Neural correlates of auditory sensory memory and automatic change detection. *NeuroImage*, *21*(1), 69–74. https://doi.org/10.1016/j.neuroimage.2003.08.033

Saffran, J. R., & Thiessen, E. D. (2007). Domain-general learning capacities. In E. Hoff & M. Shatz (Eds.), *Blackwell handbook of language development* (pp. 68–86). Blackwell Publishing. https://doi.org/10.1002/9780470757833.ch4

Sams, M., Alho, K., & Näätänen, R. (1983). Sequential effects on the ERP in discriminating two stimuli. *Biological Psychology*, *17*(1), 41–58. https://doi.org/10.1016/0301-0511(83)90065-0

SanMiguel, I., Saupe, K., & Schröger, E. (2013). I know what is missing here: Electrophysiological prediction error signals elicited by omissions of predicted "what" but not "when". *Frontiers in Human Neuroscience*, *7*, 407. https://doi.org/10.3389/fnhum.2013.00407

Schröger, E., Bendixen, A., Denham, S. L., Mill, R. W., Bohm, T. M., & Winkler, I. (2014). Predictive regularity representations in violation detection and auditory stream segregation: From conceptual to computational models. *Brain Topography*, *27*(4), 565–577. https://doi.org/10.1007/s10548-013-0334-6

Schröger, E., Bendixen, A., Trujillo-Barreto, N. J., & Roeber, U. (2007). Processing of abstract rule violations in audition. *PLoS One*, *2*(11), e1131. https://doi.org/10.1371/journal.pone.0001131

Schröger, E., Marzecova, A., & SanMiguel, I. (2015). Attention and prediction in human audition: A lesson from cognitive psychophysiology. *The European Journal of Neuroscience*, *41*(5), 641–664. https://doi.org/10.1111/ejn.12816

Seghier, M. L. (2013). The angular gyrus: Multiple functions and multiple subdivisions. *The Neuroscientist*, *19*(1), 43–61. https://doi.org/10.1177/1073858412440596

Sereno, M. I., & Huang, R. S. (2014). Multisensory maps in parietal cortex. *Current Opinion in Neurobiology*, *24*(1), 39–46. https://doi.org/10.1016/j.conb.2013.08.014

Seth, A. K., Suzuki, K., & Critchley, H. D. (2011). An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology*, *2*, 395. https://doi.org/10.3389/fpsyg.2011.00395

Shalgi, S., & Deouell, L. Y. (2007). Direct evidence for differential roles of temporal and frontal components of auditory change detection. *Neuropsychologia*, *45*(8), 1878–1888. https://doi.org/10.1016/j.neuropsychologia.2006.11.023

Shen, G., Smyk, N. J., Meltzoff, A. N., & Marshall, P. J. (2018a). Neuropsychology of human body parts: Exploring categorical boundaries of tactile perception using somatosensory mismatch responses. *Journal of Cognitive Neuroscience*, *30*(12), 1858–1869. https://doi.org/10.1162/jocn_a_01313

Shen, G., Smyk, N. J., Meltzoff, A. N., & Marshall, P. J. (2018b). Using somatosensory mismatch responses as a window into somatotopic processing of tactile stimulation. *Psychophysiology*, *55*(5), e13030. https://doi.org/10.1111/psyp.13030

Sherman, M. T., Seth, A. K., & Kanai, R. (2016). Predictions shape confidence in right inferior frontal gyrus. *The Journal of Neuroscience*, *36*(40), 10323–10336. https://doi.org/10.1523/JNEUROSCI.1092-16.2016

Shinozaki, N., Yabe, H., Sutoh, T., Hiruma, T., & Kaneko, S. (1998). Somatosensory automatic responses to deviant stimuli. *Brain Research. Cognitive Brain Research*, *7*(2), 165–171. https://doi.org/10.1016/s0926-6410(98)00020-2

Shirazibeheshti, A., Cooke, J., Chennu, S., Adapa, R., Menon, D. K., Hojjatoleslami, S. A., Witon, A., Li, L., Bekinschtein, T., & Bowman, H. (2018). Placing meta-stable states of consciousness within the predictive coding hierarchy: The deceleration of the accelerated prediction error. *Consciousness and Cognition*, *63*, 123–142. https://doi.org/10.1016/j.concog.2018.06.010

Spackman, L. A., Towell, A., & Boyd, S. G. (2010). Somatosensory discrimination: An intracranial event-related potential study of children with refractory epilepsy. *Brain Research*, *1310*, 68–76. https://doi.org/10.1016/j.brainres.2009.10.072

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, *38*(4), 387–401. https://doi.org/10.1016/0013-4694(75)90263-1

Stefanics, G., Heinzle, J., Horvath, 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

Stefanics, G., Kremlacek, J., & Czigler, I. (2014). Visual mismatch negativity: A predictive coding view. *Frontiers in Human Neuroscience*, *8*, 666. https://doi.org/10.3389/fnhum.2014.00666

Stefanics, G., Stephan, K. E., & Heinzle, J. (2019). Feature-specific prediction errors for visual mismatch. *NeuroImage*, *196*, 142–151. https://doi.org/10.1016/j.neuroimage.2019.04.020

Suda, Y., Tada, M., Matsuo, T., Kawasaki, K., Saigusa, T., Ishida, M., Mitsui, T., Kumano, H., Kirihara, K., Suzuki, T., Matsumoto, K., Hasegawa, I., Kasai, K., & Uka, T. (2022). Prediction-related frontal-temporal network for omission mismatch activity in the macaque monkey. *Frontiers in Psychiatry*, *13*, 557954. https://doi.org/10.3389/fpsyt.2022.557954

Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nature Reviews. Neuroscience*, *15*(11), 745–756. https://doi.org/10.1038/nrn3838

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13*(9), 403–409. https://doi.org/10.1016/j.tics.2009.06.003

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*, *11*(9), 1004–1006. https://doi.org/10.1038/nn.2163

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, *150*(3700), 1187–1188. https://doi.org/10.1126/science.150.3700.1187

Todorovic, A., & de Lange, F. P. (2012). Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. *The Journal of Neuroscience*, *32*(39), 13389–13395. https://doi.org/10.1523/JNEUROSCI.2227-12.2012

Tomasi, D., & Volkow, N. D. (2011). Association between functional connectivity hubs and brain networks. *Cerebral Cortex*, *21*(9), 2003–2013. https://doi.org/10.1093/cercor/bhq268

Trejo, L. J., Ryan-Jones, D. L., & Kramer, A. F. (1995). Attentional modulation of the mismatch negativity elicited by frequency differences between binaurally presented tone bursts. *Psychophysiology*, *32*(4), 319–328. https://doi.org/10.1111/j.1469-8986.1995.tb01214.x

Tse, C. Y., Rinne, T., Ng, K. K., & Penney, T. B. (2013). The functional role of the frontal cortex in pre-attentive auditory change detection. *NeuroImage*, *83*, 870–879. https://doi.org/10.1016/j.neuroimage.2013.07.037

Tse, C. Y., Shum, Y. H., Xiao, X. Z., & Wang, Y. (2021). Fronto-occipital mismatch responses in pre-attentive detection of visual changes: Implication on a generic brain network underlying mismatch negativity (MMN). *NeuroImage*, *244*, 118633. https://doi.org/10.1016/j.neuroimage.2021.118633

Tse, C. Y., Tien, K. R., & Penney, T. B. (2006). Event-related optical imaging reveals the temporal dynamics of right temporal and frontal cortex activation in pre-attentive change detection. *NeuroImage*, *29*(1), 314–320. https://doi.org/10.1016/j.neuroimage.2005.07.013

Uhrig, L., Dehaene, S., & Jarraya, B. (2014). A hierarchy of responses to auditory regularities in the macaque brain. *The Journal of Neuroscience*, *34*(4), 1127–1132. https://doi.org/10.1523/JNEUROSCI.3165-13.2014

Urakawa, T., Inui, K., Yamashiro, K., & Kakigi, R. (2010). Cortical dynamics of the visual change detection process. *Psychophysiology*, *47*(5), 905–912. https://doi.org/10.1111/j.1469-8986.2010.00987.x

Vossel, S., Geng, J. J., & Fink, G. R. (2014). Dorsal and ventral attention systems: Distinct neural circuits but collaborative roles. *The Neuroscientist*, *20*(2), 150–159. https://doi.org/10.1177/1073858413494269

Wacongne, C., Changeux, J. P., & Dehaene, S. (2012). A neuronal model of predictive coding accounting for the mismatch negativity. *The Journal of Neuroscience*, *32*(11), 3665–3678. https://doi.org/10.1523/JNEUROSCI.5003-11.2012

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

Wang, L., Amalric, M., Fang, W., Jiang, X., Pallier, C., Figueira, S., Sigman, M., & Dehaene, S. (2019). Representation of spatial sequences using nested rules in human prefrontal cortex. *NeuroImage*, *186*, 245–255. https://doi.org/10.1016/j.neuroimage.2018.10.061

Wang, L., Uhrig, L., Jarraya, B., & Dehaene, S. (2015). Representation of numerical and sequential patterns in macaque and human brains. *Current Biology*, *25*(15), 1966–1974. https://doi.org/10.1016/j.cub.2015.06.035

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

Wetzel, N., & Schröger, E. (2014). On the development of auditory distraction: A review. *PsyCh Journal*, *3*(1), 72–91. https://doi.org/10.1002/pchj.49

Winkler, I., & Czigler, I. (2012). Evidence from auditory and visual event-related potential (ERP) studies of deviance detection (MMN and vMMN) linking predictive coding theories and perceptual object representations. *International Journal of Psychophysiology*, *83*(2), 132–143. https://doi.org/10.1016/j.ijpsycho.2011.10.001

Winkler, I., Denham, S. L., & Nelken, I. (2009). Modeling the auditory scene: Predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences*, *13*(12), 532–540. https://doi.org/10.1016/j.tics.2009.09.003

Woldorff, M. G., Hackley, S. A., & Hillyard, S. A. (1991). The effects of channel-selective attention on the mismatch negativity wave elicited by deviant tones. *Psychophysiology*, *28*(1), 30–42. https://doi.org/10.1111/j.1469-8986.1991.tb03384.x

Yabe, H., Tervaniemi, M., Reinikainen, K., & Näätänen, R. (1997). Temporal window of integration revealed by MMN to sound omission. *Neuroreport*, *8*(8), 1971–1974. https://doi.org/10.1097/00001756-199705260-00035

Yamaguchi, S., & Knight, R. T. (1991). Age effects on the P300 to novel somatosensory stimuli. *Electroencephalography and Clinical Neurophysiology*, *78*(4), 297–301. https://doi.org/10.1016/0013-4694(91)90184-6

Yamaguchi, S., & Knight, R. T. (1992). Effects of temporal-parietal lesions on the somatosensory P3 to lower limb stimulation. *Electroencephalography and Clinical Neurophysiology*, *84*(2), 139–148. https://doi.org/10.1016/0168-5597(92)90018-7

Yaron, A., Hershenhoren, I., & Nelken, I. (2012). Sensitivity to complex statistical regularities in rat auditory cortex. *Neuron*, *76*(3), 603–615. https://doi.org/10.1016/j.neuron.2012.08.025

Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zollei, L., Polimeni, J. R., Fischl, B., Liu, H., & Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(3), 1125–1165. https://doi.org/10.1152/jn.00338.2011

Yucel, G., McCarthy, G., & Belger, A. (2007). fMRI reveals that involuntary visual deviance processing is resource limited. *NeuroImage*, *34*(3), 1245–1252. https://doi.org/10.1016/j.neuroimage.2006.08.050

Yucel, G., Petty, C., McCarthy, G., & Belger, A. (2005). Graded visual attention modulates brain responses evoked by task-irrelevant auditory pitch changes. *Journal of Cognitive Neuroscience*, *17*(12), 1819–1828. https://doi.org/10.1162/089892905775008698

Zachlod, D., Ruttgers, B., Bludau, S., Mohlberg, H., Langner, R., Zilles, K., & Amunts, K. (2020). Four new cytoarchitectonic areas surrounding the primary and early auditory cortex in human brains. *Cortex*, *128*, 1–21. https://doi.org/10.1016/j.cortex.2020.02.021

Zhang, Q., Luo, C., Ngetich, R., Zhang, J., Jin, Z., & Li, L. (2022). Visual selective attention P300 source in frontal-parietal lobe: ERP and fMRI study. *Brain Topography*, *35*(5–6), 636–650. https://doi.org/10.1007/s10548-022-00916-x

Zhao, C., Valentini, E., & Hu, L. (2015). Functional features of crossmodal mismatch responses. *Experimental Brain Research*, *233*(2), 617–629. https://doi.org/10.1007/s00221-014-4141-4

Zvyagintsev, M., Zweerings, J., Sarkheil, P., Bergert, S., Baqapuri, H., Neuner, I., Gaebler, A. J., & Mathiak, K. (2020). Auditory mismatch processing: Role of paradigm and stimulus characteristics as detected by fMRI. *Biological Psychology*, *154*, 107887. https://doi.org/10.1016/j.biopsycho.2020.107887

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Grundei, M., Schmidt, T. T., & Blankenburg, F. (2023). A multimodal cortical network of sensory expectation violation revealed by fMRI. *Human Brain Mapping*, *44*(17), 5871–5891. https://doi.org/10.1002/hbm.26482

Author Contributions

Declaration pursuant to Sec. 7 (3), fourth sentence, of the Doctoral Study Regulations regarding my own share of the submitted scientific or scholarly work that has been published or is intended for publication within the scope of my publication-based work

I. Last name, first name: Grundei, Miro
Institute: Department of Education and Psychology, Freie Universität Berlin
Doctoral study subject: Psychology
Title: Mismatch responses: Probing probabilistic inference in the brain

II. Numbered listing of works submitted (title, authors, where and when published and/or submitted):

1. Gijsen*, S, Grundei*, M., Lange, R. T., Ostwald, D. & Blankenburg, F. (2021). Neural surprise in somatosensory Bayesian learning. PLoS computational biology, 17(2), e1008068.
2. Grundei, M., Schröder, P., Gijsen, S. & Blankenburg, F. (2023). EEG mismatch responses in a multimodal roving stimulus paradigm provide evidence for probabilistic inference across audition, somatosensation and vision. Human Brain Mapping.
3. Grundei*, M., Schmidt* T. T., & Blankenburg, F. (2023). A multimodal cortical network of sensory expectation violation revealed by fMRI. Human Brain Mapping, 44(17), 5871-5891.

*shared first authorship

III. Explanation of own share of these works:

Regarding II.1: Study conceptualization and design (substantial), methodology (substantial), programming of task (majority), data collection (substantial), data analysis (substantial), discussion of results (substantial), writing/revising the manuscript (substantial).

Regarding II.2: Study conceptualization and design (vast majority), methodology (vast majority), programming of task (vast majority), data collection (all), data analysis (all), discussion of results (substantial), writing/revising the manuscript (vast majority).

Regarding II.3: Study conceptualization and design (majority), methodology (majority), programming of task (vast majority), data collection (substantial), data analysis (majority), discussion of results (substantial), writing/revising the manuscript (majority).

IV. Names, addresses, and e-mail addresses for the relevant co-authors:

Regarding II.1: Sam Gijsen (1,2), sam.gijsen@fu-berlin.de
Robert T. Lange (3,4), Robert.t.lange@tu-berlin.de
Dirk Ostwald (5), dirk.ostwald@ovgu.de
Felix Blankenburg (1), felix.blankenburg@fu-berlin.de

Regarding II.2: Pia Schröder (1), pia.schroeder@fu-berlin.de
Sam Gijsen (1,2), see above
Felix Blankenburg (1), see above

Regarding II.3: Timo T. Schmidt (1), timo.t.schmidt@fu-berlin.de
Felix Blankenburg (1), see above

(1) Neurocomputation and Neuroimaging Unit, Freie Universität Berlin, 14195 Berlin, Germany
(2) Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, 10117 Berlin, Germany
(3) Berlin Institute of Technology, Berlin, Germany
(4) Einstein Center for Neurosciences, Berlin, Germany
(5) Institute of Psychology, Otto von Guericke Universität Magdeburg, 39106 Magdeburg, Germany

Date, doctoral candidate signature: ……...

I confirm the declaration made by Miro Grundei under III.:

Name: Sam Gijsen Signature: ……..

Name: Robert T. Lange Signature: ……..

Name: Dirk Ostwald Signature: ……..

Name: Felix Blankenburg Signature: ……..

Name: Pia Schröder Signature: ……..

Name: Timo T. Schmidt Signature: ……..

Eidesstattliche Erklärung

Hiermit versichere ich,

- dass ich die vorliegende Arbeit eigenständig und ohne unerlaubte Hilfe verfasst habe,
- dass Ideen und Gedanken aus Arbeiten anderer entsprechend gekennzeichnet wurden,
- dass ich mich nicht bereits anderwärtig um einen Doktorgrad beworben habe und keinen Doktorgrad in dem Promotionsfach Psychologie besitze, sowie
- dass ich die zugrundeliegende Promotionsordnung vom 08.08.2016 anerkenne.

Berlin, 14. April 2023

Miro Grundei